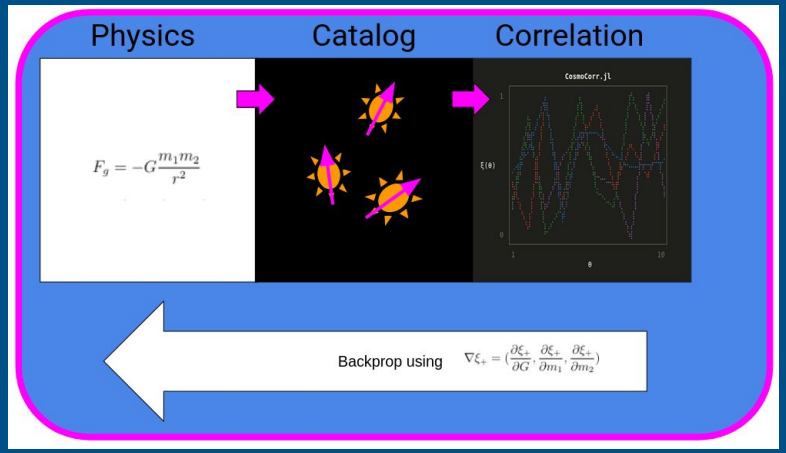
12.05.24

On Soft Clustering for Correlation Estimators:

Model Uncertainty, Differentiability, and Surrogates

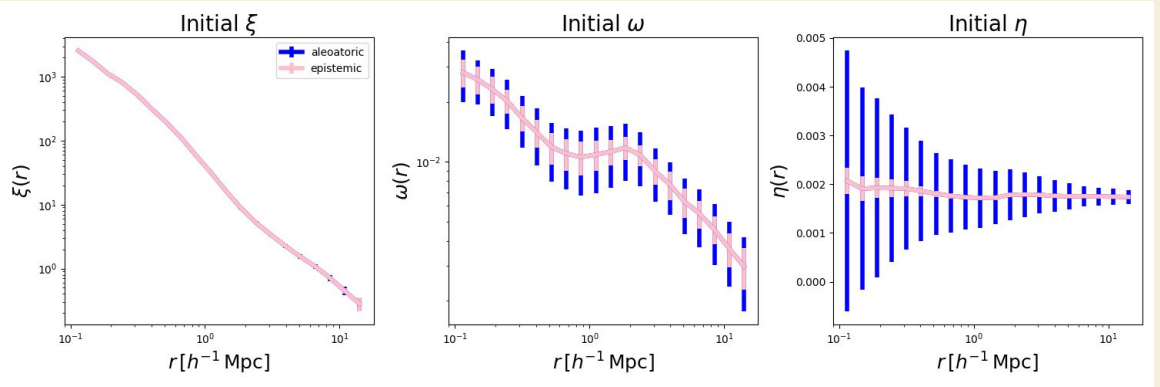


Edward Berman

1. **Correlation Function Overview**
2. Model Uncertainty
3. Differentiability
4. Surrogates

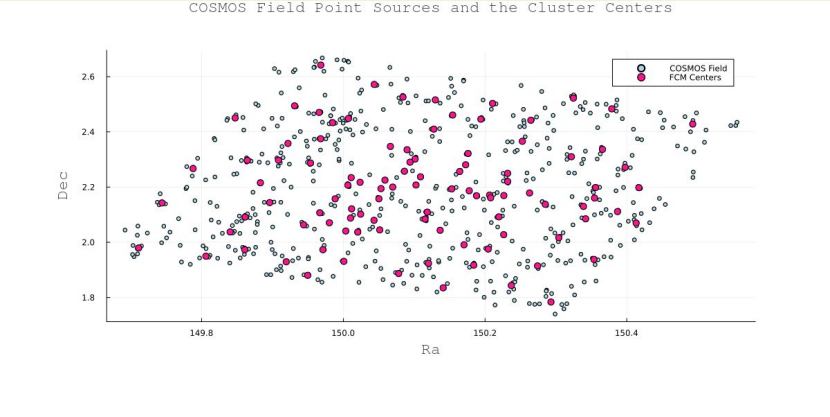
Correlation Functions

Expected value of the product between two quantities as a function of how far apart they are. Used all over cosmology.



Correlation Functions

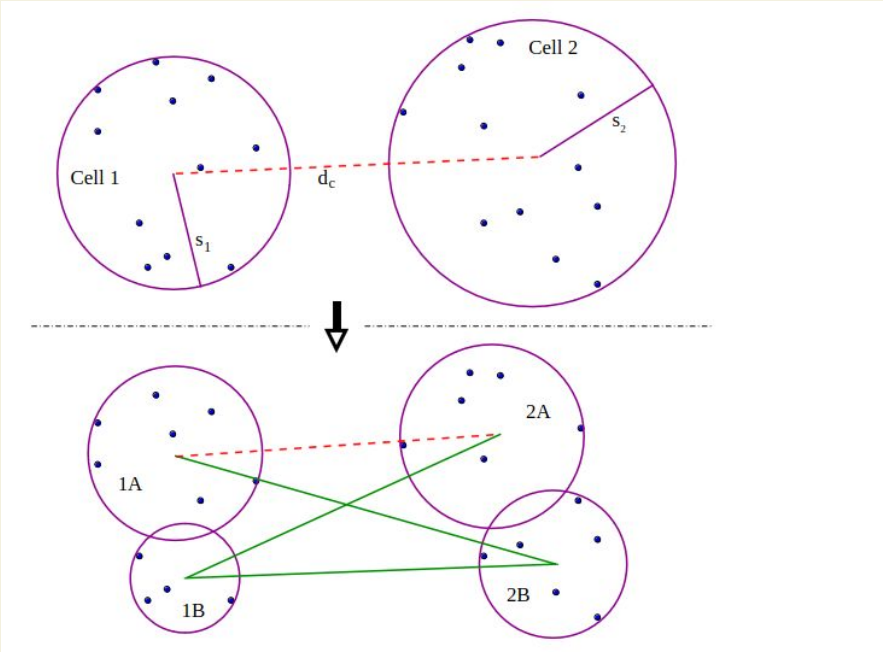
- The number of distance calculations grows quadratically with the number of objects (SLOW!!!)
- Thus, we cluster :D



1. Correlation Function Overview
2. **Model Uncertainty**
3. Differentiability
4. Surrogates

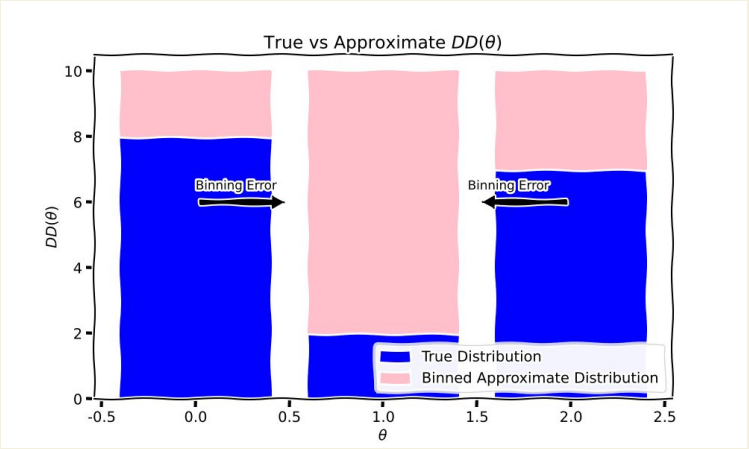
Model Uncertainty

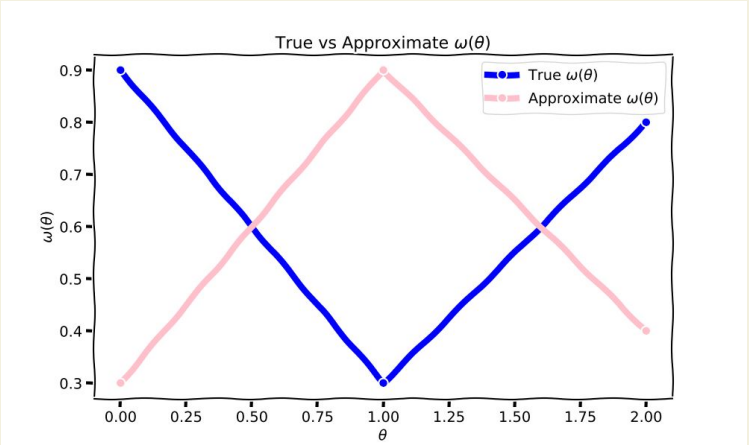
- For a significant number of objects, the number of distances that are overestimated is close to the number that are underestimated
- We can even restrict the error to one distance bin [Jarvis 2004]
- But what if we don't have that many objects?



Model Uncertainty

- For a significant number of objects, the number of distances that are overestimated is close to the number that are underestimated
- We can even restrict the error to one distance bin [Jarvis 2004]
- But what if we don't have that many objects?



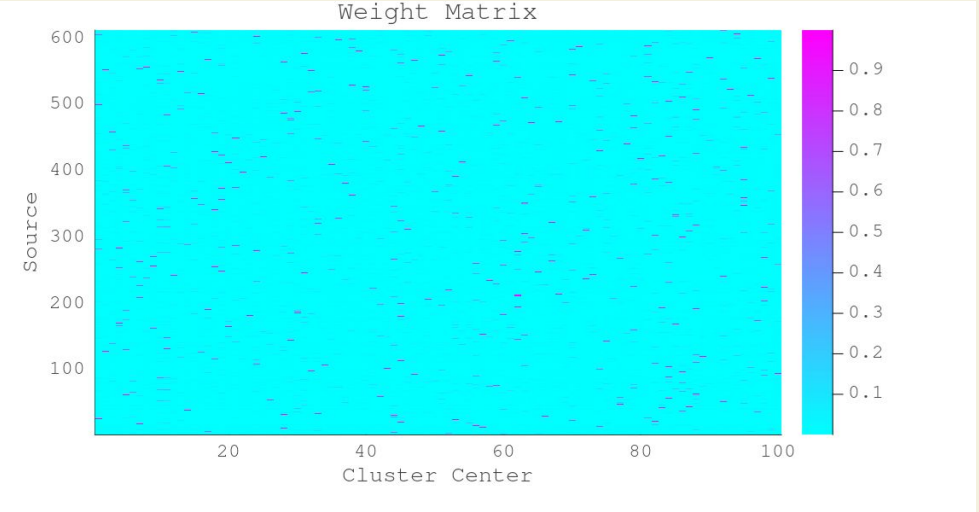


Model Uncertainty

- How can we quantify the uncertainty caused by the clustering?
- Can we reconcile clustering uncertainties from data uncertainties?

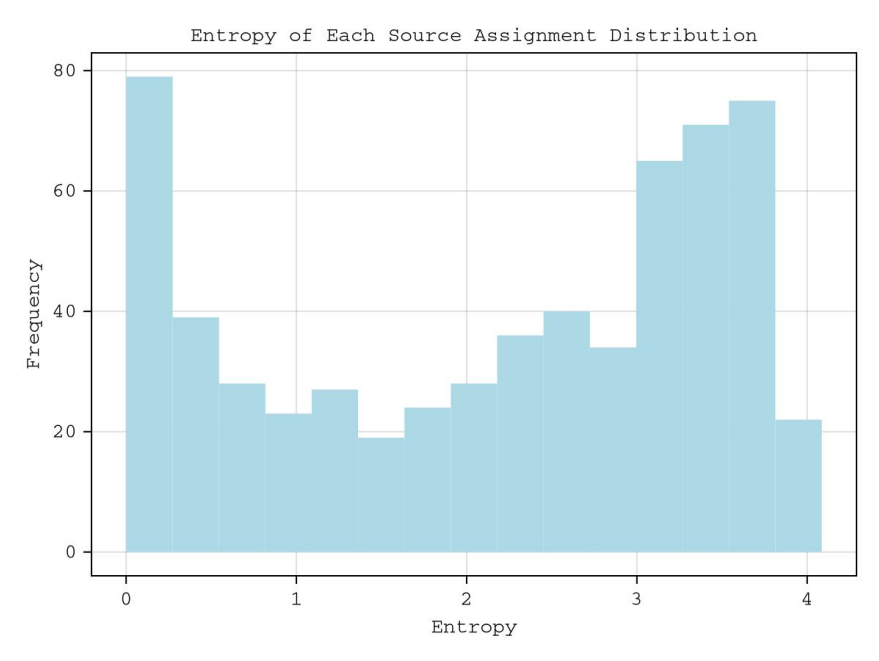
Model Uncertainty

- How can we quantify the uncertainty caused by the clustering? **Make cluster assignments probabilistic (fuzzy-c-means). Repeat inference!**
- Can we reconcile clustering uncertainties from data uncertainties? **Yes! Combine the inference error bar with a bootstrapping or jackknife technique.**



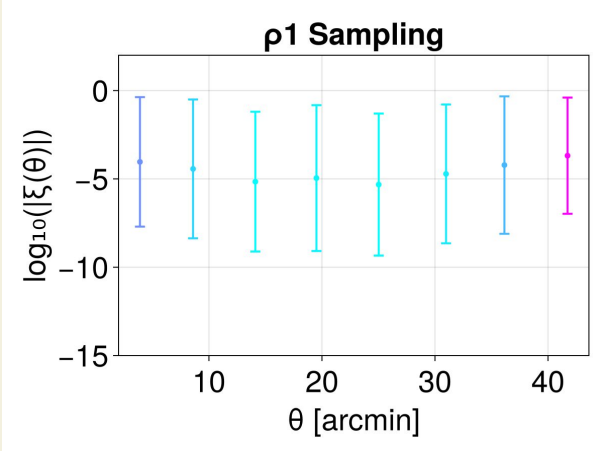
Model Uncertainty

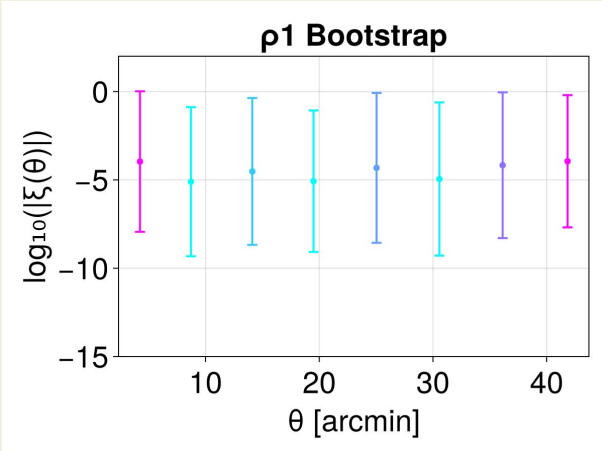
- How can we quantify the uncertainty caused by the clustering? **Make cluster assignments probabilistic (fuzzy-c-means). Repeat inference!**
- Can we reconcile clustering uncertainties from data uncertainties? **Yes! Combine the inference error bar with a bootstrapping or jackknife technique.**



Model Uncertainty

- We find that for data sparse science cases, the error bars are comparable!!

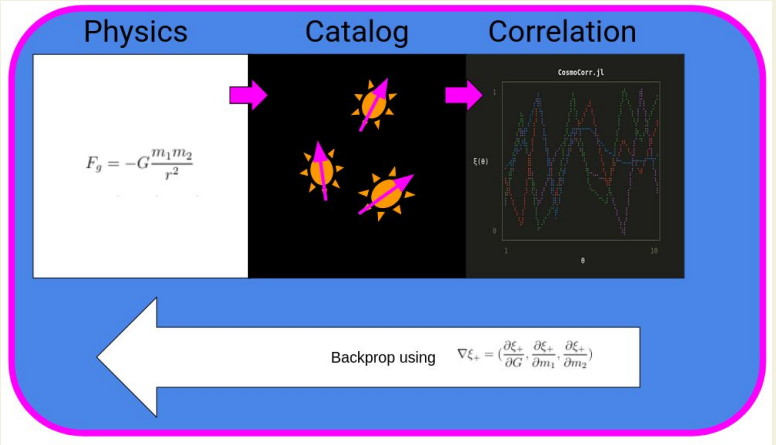




1. Correlation Function Overview
2. Model Uncertainty
3. **Differentiability**
4. Surrogates

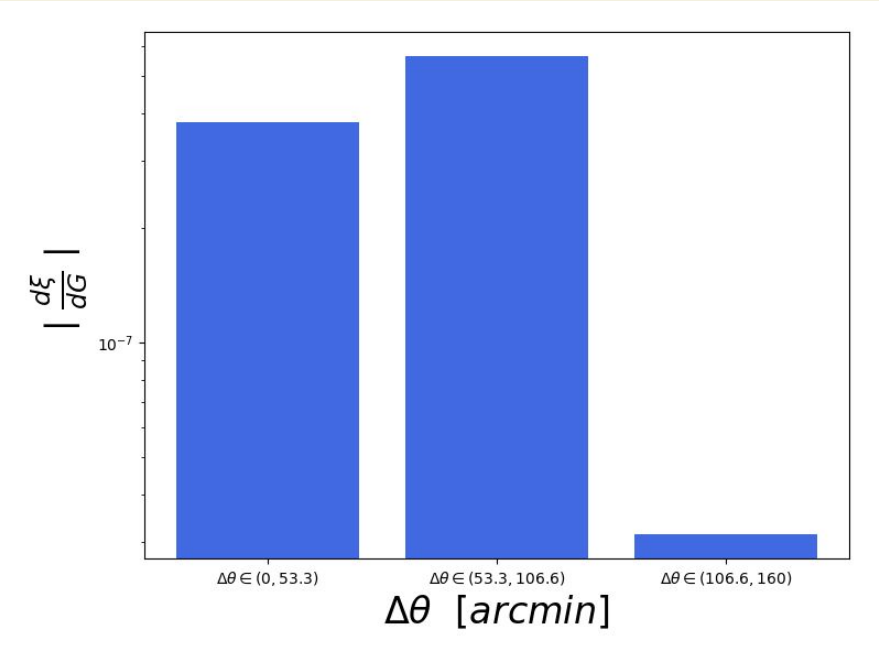
Differentiability

- **I LOVE SIMULATIONS :D**
- I want to go from forward model, to galaxy catalog, to correlation function
- How do correlations depend on parameters in the forward model
- *We want gradient based optimization*



Differentiability

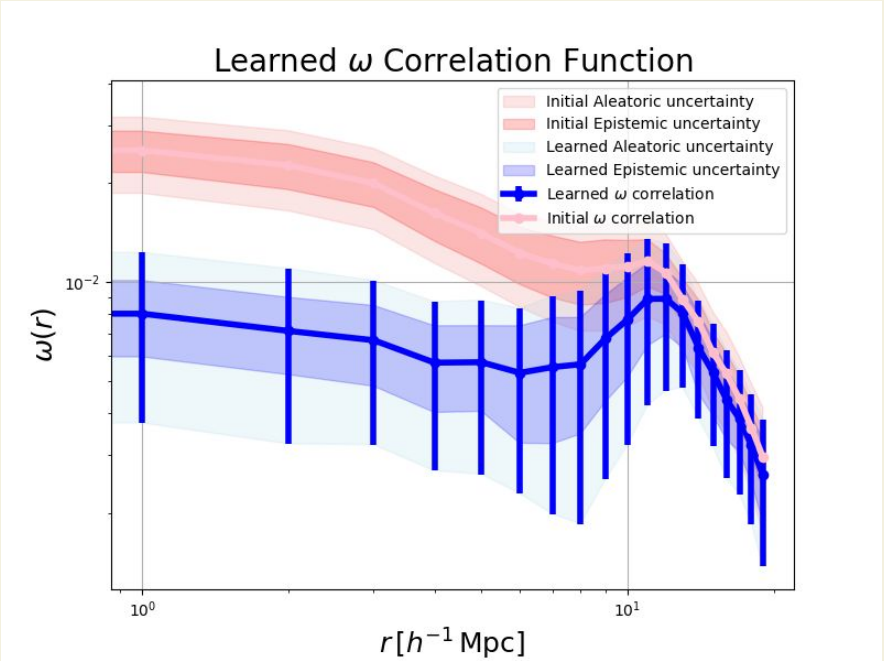
- Our algorithm for model uncertainty relied on sampling (not differentiable)
- We can amend this with three different approaches:
  - Gumbel-max reparameterization
  - No assignments (weighted averaging)
  - Skip Gradients



1. Correlation Function Overview
2. Model Uncertainty
3. Differentiability
4. **Surrogates**

Surrogates

- Some differentiation techniques are already approximate
- They can also be slow to compute
- Can we learn the relationship between astrophysical model parameters and correlations directly and use that for our differentiable solution? **Yes :D**



Conclusions. contact, and thanks =]

- Correlation Functions are Kewl
- For more, see: https://github.com/EdwardBerman/cosmo-corr
- [Berman et al in prep.] [AAS winter session 2025]
- http://ebrmn.space/
- f(berman, ed, northeastern) where f(x,y,z) = x.y@z.edu
- Thanks for list'nin'