

Likelihood free inference of the halo model parameters with simulations of galaxy mock catalogs

ABSTRACT

The Halo Occupation Distribution model is a powerful prescription of how galaxies populate dark matter halos that has had much success in reproducing large scale galaxy clustering and understanding galaxy evolution within the context of the halo model. Constraints on the parameters of the HOD model have so far been inferred using a likelihood function that assumes, first a functional—often times Gaussian—form of the likelihood, and Second, an analytical model for the galaxy clustering statistics. We use Approximate Bayesian Computation in order to infer the HOD parameters by assuming a prior over the parameters and a distance function between statistical summaries of the data and statistical summaries of the simulations. Using a mock galaxy catalog created with known HOD parameters, we obtain precise constraints on the HOD parameters that reproduces the summary statistics of the mock galaxy catalog accurately. Since ABC does not require computing the likelihood, we include observables beyond galaxy clustering such as the galaxy number density and the galaxy group multiplicity function. We demonstrate that adding these measurements to our set of observables leads to stronger constraints on the HOD parameters, specifically the ones that govern the population of the satellite galaxies.

1. Introduction

2. Method

2.1. Approximate Bayesian Computation

ABC is a likelihood free method for inference of a set of parameters that are required for the simulation of a physical system, through a forward model, a distance function, and a prior probability distribution over the parameters. Let us denote the the data by \mathcal{D} . We assume that there exist a forward model \mathcal{M} that simulates the data $\bar{\mathcal{D}}$ for a given set of parameters θ

$$\bar{\mathcal{D}}_{\theta} = \mathcal{M}(\theta) \tag{1}$$

Furthermore, ABC requires adopting a distance function between the data \mathcal{D} and the

simulation $\bar{\mathcal{D}}_\theta$. The distance function, in general, can be a vector $\mathbf{d}(\mathcal{D}, \bar{\mathcal{D}}_\theta)$

$$\mathbf{d}(\mathcal{D}, \bar{\mathcal{D}}_\theta) = [d_{1,\theta}, \dots, d_{M,\theta}], \quad (2)$$

$$d_{i,\theta} = d_i(\mathcal{S}_i(\mathcal{D}), \mathcal{S}_i(\bar{\mathcal{D}}_\theta)) \text{ for } i = \{1, \dots, M\}, \quad (3)$$

where $d_i(\cdot)$ is a scalar function that defines a distance between i -th summary statistics of the data $\mathcal{S}_i(\mathcal{D})$ and the same summary statistics of the simulation $\mathcal{S}_i(\bar{\mathcal{D}}_\theta)$, and M is the total number of statistical summaries used in the analysis.

We draw a large number of proposals from the prior distribution over the parameter space, and reject the samples for which the distance between the data and the simulation corresponding to that sample is more than a certain threshold ϵ . This requirement is satisfied *if and only if* every component of the distance vector is less than its corresponding threshold. Note that, each summary statistics has a separate threshold. This procedure is done iteratively until, we are left with a certain number of samples that satisfy the distance criteria. The final ensemble approximates the posterior pdf over the parameters of the forward model of the data. This sampling method is called *rejected sampling*.

Rejected sampling is expensive. Instead we use population monte carlo, adaptive importance sampling,

2.2. Halo Model

Halo Occupation distribution is a powerful tool for description of galaxy/halo relation.

...

2.3. Summary statistics

Talk about 2pcf, group multiplicity function, and number density ...

3. Forward modeling of the galaxy population with HALOTOOLS

Hello World! Talk about why it is better to simulate the mock and compute the summary statistics of the mock instead of using analytical formulas...

Talk about all the assumptions that go into analytic formulas...

Talk about why we do not want to write down likelihood function for gmf, 2pcf, ...

...

4. Mock data

Hello World!

Plots showing 2pcf, gmf, ... of the mock data. How we compute the covariance matrix from mocks... Talk about error bars. Talk about poisson errors in group populations ...

5. Results

Show the posteriors... Show evolution of the posterior errorbars with iteration. Show epsilon. Convince that we are converging ... Draw samples from the posterior, populate the mock catalogs with those samples, and compute the summary statistics with them and demonstrate that they match the data...

6. Discussion

talk about cosmological inference and why this is good first step for likelihood free inference in large scale ...

talk about why it is hard to do cosmology ...

talk about halo model ...

talk about HOD and why it is too simple to be the right model ...

talk about assembly bias perhaps ... and how we can use the same machinery for detection of assembly bias ...

talk about writing down a likelihood for gmf, tpcf, ... , and the gaussianity assumption
....

talk about our results ...

talk about future directions ...

REFERENCES