Bayesian Nonparametric Methods Applied to Velocity Distribution of Clusters of Galaxies

Vitor Medeiros Sampaio

Instituto Nacional de Pesquisas Espaciais

2019

Summary

Dynamical Equilibrium
Nurture vs Nature

- Equilibrium for gravitational-bound objects such as clusters of galaxies depends not only on the intrinsic properties, but also on the environment in which the object is embed within ;
- An usual descriptor for the equilibrium state is density, however recent findings show the velocity distribution as an equally plausible way to describe it ;
- In addition, the phase space projected in the line-of-sight also plays a major role. Recent simulations shows that galaxies in different processes inside a cluster of galaxies tend to lie within a specific region in the PPS ;

Dynamical Equilibrium
Nurture vs Nature

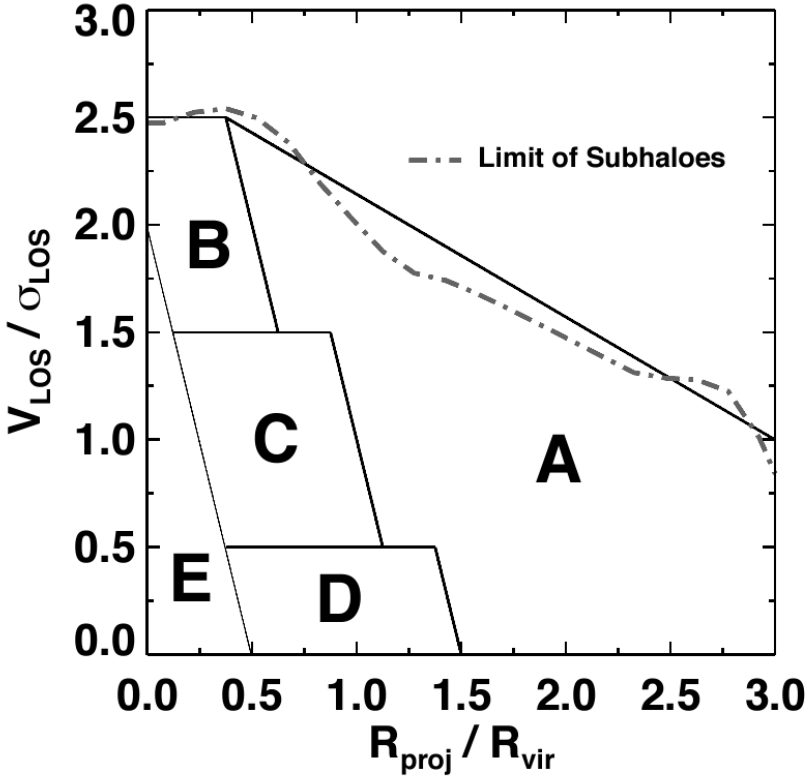
- Equilibrium for gravitational-bound objects such as clusters of galaxies depends not only on the intrinsic properties, but also on the environment in which the object is embed within ;
- An usual descriptor for the equilibrium state is density, however recent findings show the velocity distribution as an equally plausible way to describe it ;
- In addition, the phase space projected in the line-of-sight also plays a major role. Recent simulations shows that galaxies in different processes inside a cluster of galaxies tend to lie within a specific region in the PPS ;

Dynamical Equilibrium
Nurture vs Nature

- Equilibrium for gravitational-bound objects such as clusters of galaxies depends not only on the intrinsic properties, but also on the environment in which the object is embed within ;
- An usual descriptor for the equilibrium state is density, however recent findings show the velocity distribution as an equally plausible way to describe it ;
- In addition, the phase space projected in the line-of-sight also plays a major role. Recent simulations shows that galaxies in different processes inside a cluster of galaxies tend to lie within a specific region in the PPS ;

Dynamical Equilibrium
Nurture vs Nature

- **A** :Region mainly containing *Interlopers* ;
- **B** :Region dominated by *Recent Infallers* ;
- **C** :Region with galaxies mixed from **B** and **D** ;
- **D** :Region dominated by *Intermediate infallers* and *Backsplash* galaxies ;
- **E** :Area containing mainly *Ancient Infallers* ;



Objectives

- Clusters which the velocity distribution follows a Gaussian Distribution have different characteristics from those which does not, based on this there are two main objectives :
- Study the Projected Phase Space for Clusters previously classified according to their Gaussianity of the velocity distribution.
- Propose a new method to determine when the velocity distribution follows a Gaussian Distribution.

Objectives

- Clusters which the velocity distribution follows a Gaussian Distribution have different characteristics from those which does not, based on this there are two main objectives :
- Study the Projected Phase Space for Clusters previously classified according to their Gaussianity of the velocity distribution.
- Propose a new method to determine when the velocity distribution follows a Gaussian Distribution.

Objectives

- Clusters which the velocity distribution follows a Gaussian Distribution have different characteristics from those which does not, based on this there are two main objectives :
- Study the Projected Phase Space for Clusters previously classified according to their Gaussianity of the velocity distribution.
- Propose a new method to determine when the velocity distribution follows a Gaussian Distribution.

Determining the Gaussianity of the Velocity Distribution

A Bayesian Nonparametric Approach

- While in Parametric processes the number of parameters are constraint, in a Nonparametric approach there is no need to specify the number of parameters, as they are added automatically by the method according to the data's need.
- The Bayesian method takes into account the whole distribution for the parameter, while frequentists usually only take the highest probability value. In a linear regression, instead of a single line, the result would be two distinct probability distributions for the slope and intercept coefficients.

Determining the Gaussianity of the Velocity Distribution

A Bayesian Nonparametric Approach

- While in Parametric processes the number of parameters are constraint, in a Nonparametric approach there is no need to specify the number of parameters, as they are added automatically by the method according to the data's need.
- The Bayesian method takes into account the whole distribution for the parameter, while frequentists usually only take the highest probability value. In a linear regression, instead of a single line, the result would be two distinct probability distributions for the slope and intercept coefficients.

BNP Implementation

Dirichlet Process and Markov Chain Monte Carlo

- Build a fully nonparametric model is not straightforward. A lot of complex mathematics is used and the usual method is the Dirichlet Process.
- Also, to get an Bayesian result, it is necessary to determine what is called the *Likelihood*. This is calculated using Markov Chain Monte Carlo (MCMC).

BNP Implementation

Dirichlet Process and Markov Chain Monte Carlo

- Build a fully nonparametric model is not straightforward. A lot of complex mathematics is used and the usual method is the Dirichlet Process.
- Also, to get an Bayesian result, it is necessary to determine what is called the *Likelihood*. This is calculated using Markov Chain Monte Carlo (MCMC).

Markov Chain Monte Carlo

Sampling to Estimate the Desired Function

- Using a MCMC method, the parameter space is explored in order to get an estimate for the desired distribution.
- If dealing with a Gaussian,for example, it will randomly select values for the mean and standard deviation, sample from this hypothetical Gaussian and compare with the observed data.

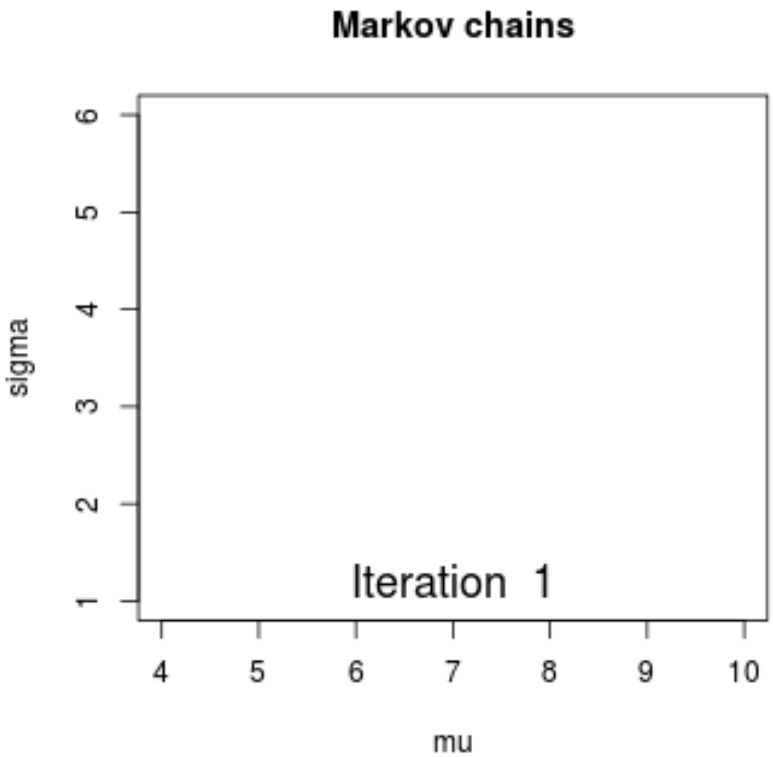
Markov Chain Monte Carlo

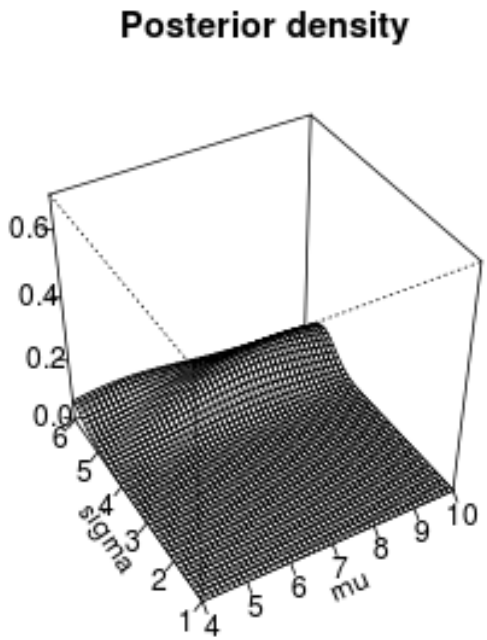
Sampling to Estimate the Desired Function

- Using a MCMC method, the parameter space is explored in order to get an estimate for the desired distribution.
- If dealing with a Gaussian,for example, it will randomly select values for the mean and standard deviation, sample from this hypothetical Gaussian and compare with the observed data.

Markov Chain Monte Carlo

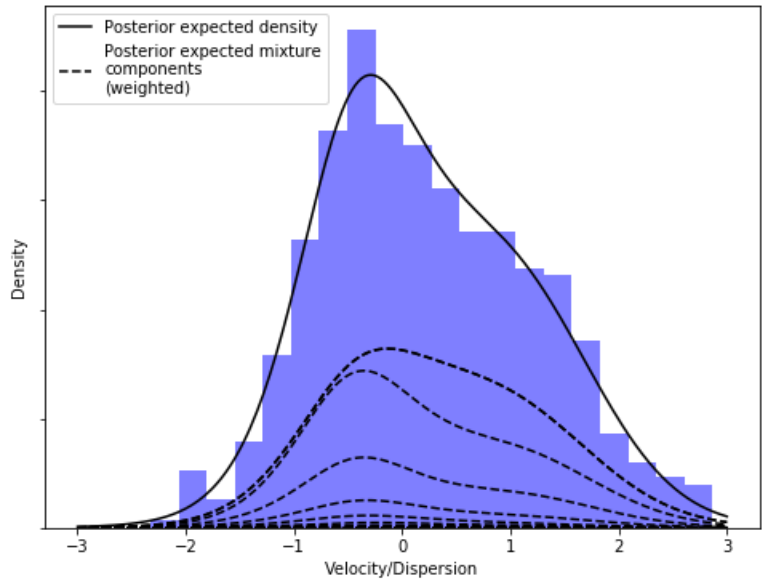
Sampling to Estimate the Desired Function

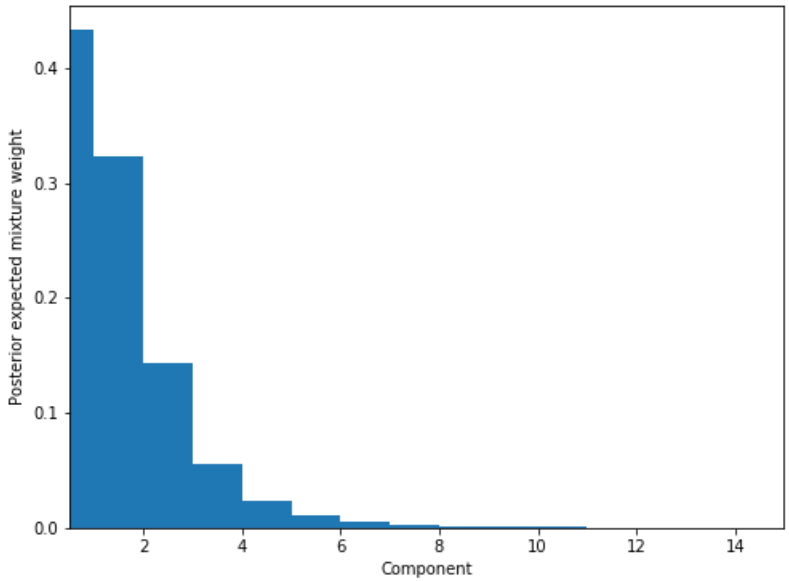




Some Preliminary Results
The Estimated Density Function

- The posterior and weights found for the Cluster Yang 005 are the following :

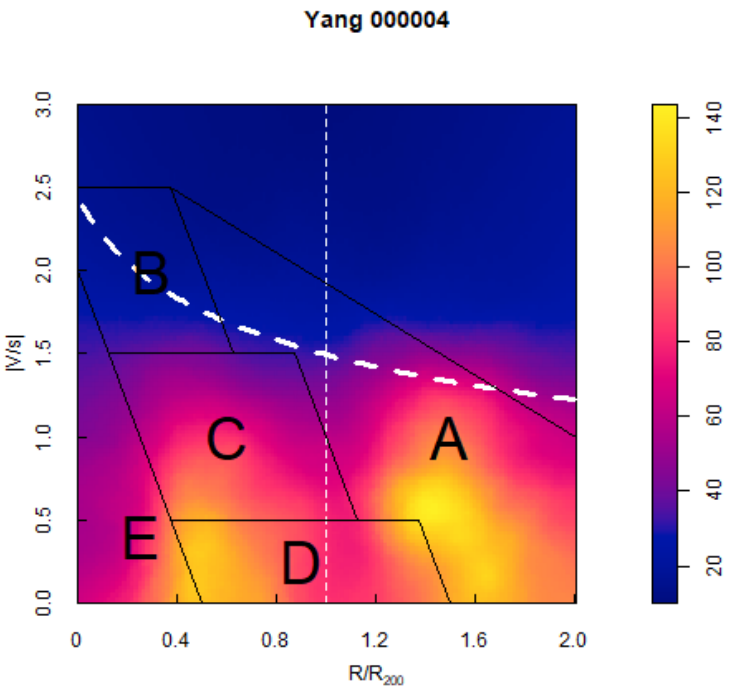


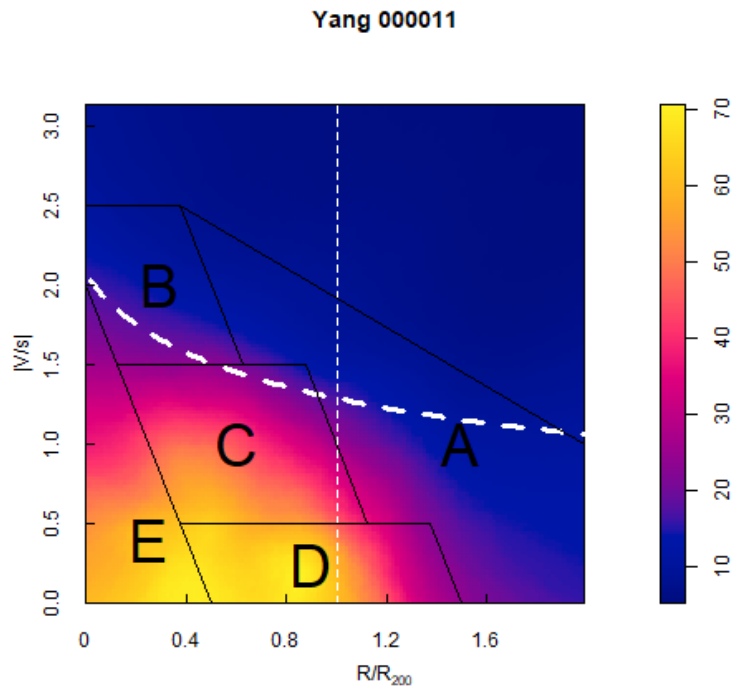


Projected Phase Space

Differences Between Gaussian and Non-Gaussian Clusters

- An analysis considering the phase space can lead to further evidence in this difference between clusters with Gaussian Distribution and those with NonGaussian Distribution.





Let's get to work!

Thanks!