

BAYESIAN INFERENCE OF THE MASS-RICHNESS RELATION

SEAN MCLAUGHLIN

Kavli Institute for Particle Astrophysics and Cosmology & Physics Department, Stanford University, Stanford, CA 94305, USA
 October 30, 2015

ABSTRACT

In this project I attempt to infer the mass-richness relation for galaxy clusters. I begin by building a model and establishing functional relationships between parameters. I then generate some mock data and compare it to data from cosmological simulations. Finally, I use MCMC to attempt to infer the global parameters of the relation.

1. INTRODUCTION

Richness (λ) is a measure of how many galaxies are in a galaxy cluster. This value is (like all things) related to the mass of the cluster by a power law.

$$\lambda = a M^b + \sigma$$

Where σ is some insofar undefined intrinsic scatter. This model allows us to estimate the probability of a galaxy halo having a given mass given some richness. Using a collection of observed richnesses, we wish to infer the parameters of this model $\{a, b, \sigma\}$ in a known cosmology. This would be trivial to do if the masses of each halo were known along with their richnesses; however, dark matter halos can not be directly observed.

One answer is to make use of the cluster mass function. This function defines the number density of halos of a given mass as a function of redshift and cosmological parameters. By treating the mass function as an unnormalized PDF of mass given redshift and cosmology, it will be possible to marginalize over the uncertainty in mass and place bounds on the richness relation.

This paper is structured as follows. In [section 2](#) I describe the probability model and assumptions, and in [section 3](#), I show generated mock data and simulation data used in comparison. I then describe the practical implementation of this probabilistic framework and the results of that inference in [section 4](#). After some discussion in [section 5](#), I present my conclusions in [section 6](#).

2. MODEL

The model has the PGM shown in [Figure 2](#). Implicit in this model are a few assumptions. I've assumed that the redshift of each individual cluster is perfectly known, not an observed quantity. I've also assumed some cosmology Ω . We could unfix it and choose to infer both the relation parameters and the cosmology, or fix the parameters to focus just on the cosmology. Those options will be considered in future work. I've also assumed at this point that all richnesses have been observed perfectly. The only scatter on the richness, then, is intrinsic. This PGM decomposes to the distribution:

$$P(a, b, \sigma | \Omega, \{z\}, \{\lambda\}) = \int \prod_{k=1}^N P(\lambda_k | M_k, a, b, \sigma) P(M_k | z_k, \Omega) P(a, b, \sigma) dM_k \quad (1)$$

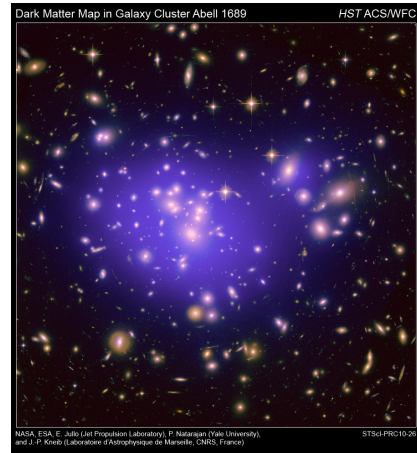


Figure 1. Image of a galaxy cluster, with dark matter overlaid in purple. Richness is the count of galaxies in a cluster like this, and I am attempting to measure the dark matter mass from that count. Credit to HST and NASA.

Where it is clear to see the most challenging portion of the inference will be the integration over all the masses M_k . I will follow with explanations for the choices of priors and likelihoods.

2.1. Richness Likelihood

I'll assume that the likelihood for the richness $P(\lambda_k | M_k, a, b, \sigma)$ is lognormally distributed. More explicitly, the log likelihood is given by:

$$\log \mathcal{L}(\lambda_k | M_k, a, b, \sigma) = \frac{-(b \log(a M_k) - \log(\lambda_k))^2}{2 \sigma^2} - \log(\sigma \lambda_k) - \log(\sqrt{2\pi}) \quad (2)$$

Where the $\log \sqrt{2\pi}$ will be dropped in implementation, as the constant will not effect MCMC.

2.2. Mass Likelihood

As stated above, the likelihood for an individual cluster mass $P(M_k | z_k, \Omega)$ is given by a normalized mass function. This turns out to be non-trivial in practice, and worth some discussion. The standard mass function is the Tinker function, which has the form:

$$n_{Tinker}(M, z) = \frac{dn}{dM} = \frac{\bar{\rho}_m(z)}{M} \frac{d \log \sigma^{-1}(M, z)}{dM} f(\sigma, z)$$

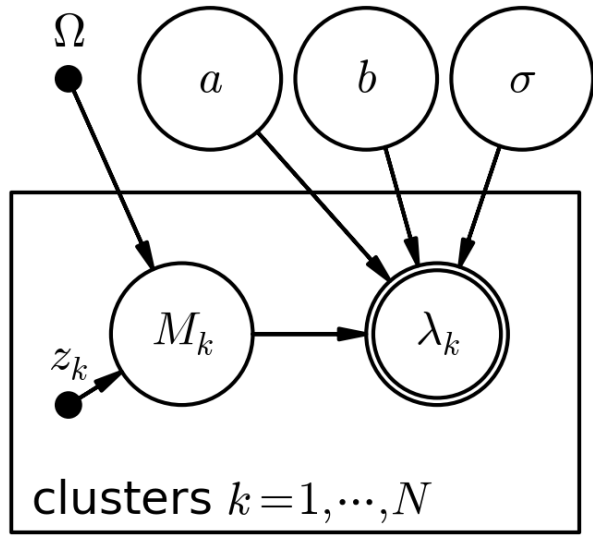


Figure 2. Probabilistic graphical model.

Where $\bar{\rho}_m$ is the mean comoving matter density and σ^2 is the variance of the linearly evolved density field. $f(\sigma, z)$ is an analytic fitting function, generally of the form:

$$f(\sigma, z) = A \left[\left(\frac{\sigma}{b} \right)^{-a} + 1 \right] e^{-c/\sigma^2}$$

Where $\{A, a, b, c\}$ are allowed to vary as a power law of $(1+z)^\alpha$. This is all of course incredibly complex, both to code up and run computationally. Luckily et al. Evrard et al. (2014) showed the following approximation:

$$n_i = A \exp - \sum_{j=1}^i \beta_j \mu^j / j!$$

Where $\mu \equiv M/M_p$ for a parameter M_p which along with the remaining parameters are computed explicitly. The trouble is, in this paper those parameters are only computed for two redshifts, $z = 0.23$ and 1.5 . For simplicity in this case, I've interpolated between the given values. I'm mostly interested in redshifts between 0 and 1, which will be stable with this approximation. I will have to find a better approach in future work.

So, explicitly, the (unnormalized) log likelihood is:

$$\log \mathcal{L}(M|z, \Omega) = - \sum_{j=1}^i \frac{\beta_j(z)}{j!} \left(\frac{M}{M_p(z)} \right)^j + \log A(z)$$

For the MCMC, I did not normalize the distribution. For mock sampling normalization is required, so a trapezoidal approximate integrator was used.

2.3. Priors

I've assumed each of the parameters among $\{a, b, \sigma\}$ is independent. I sought to construct non-informative priors.

For a , there are few prior restrictions except that it must be greater than 0 since there can be no negative

richness. Since in the log-log plot of richness v. mass, $\log a$ is the intercept of the "line", so I considered a prior in $\log a$ similar to one for an intercept in a linear fit. However, it was found that will actually bias a to exponentially large values. Better instead may be an exponential in $\log(a)$ which would be approximately uniform in a . This has one hyperparameter to tune, χ .

$$\log \mathcal{L}(\log a) = \log \chi - \chi \log a$$

For b , about the same is known. Physically, it must be greater than 0. On a log-log plot, b is the slope of the line of best fit. A distribution that favors all angles of that line equally will be a good choice. So, I've defined

$$b \equiv \tan(\theta)$$

And insisted $\theta \sim U(0, \frac{\pi}{2})$ such that so long as $\arctan b > 0$:

$$\log \mathcal{L}(\theta) = \log \mathcal{L}(\arctan(b)) = \log\left(\frac{2}{\pi}\right)$$

Lastly, is the prior for σ for which several options were considered. σ needs to be invariant to scaling, so the Jeffery's prior $\mathcal{L} \propto 1/\sigma$ is a good choice. However, this has the disadvantage of being an improper prior. Another choice is the inverse gamma distribution. It is the conjugate prior for the log-Normal distribution. It has 2 hyperparameters α and β , which is possibly too many. In this work I chose the inverse gamma distribution, rather than choose an improper prior. In future work I'd like to make use of it's conjugate prior properties.

$$\log \mathcal{L}(\sigma) = -\log \Gamma(\sigma|\alpha, \beta)$$

3. DATA

The data for this analysis was taken from the Buzard simulation, v1.0. The redMaPPer cluster finder was run against this simulation and produced a catalog of clusters. Figure 3 shows the richness and redshifts generated by redMaPPer. Unfortunately, connecting the clusters observed by redMaPPer with their true masses is a non-trivial. I was not able to do so at the time of this work. There are additional features to the data worth discussing. RedMaPPer truncates it's observed richnesses at 20. Truncation is a very challenging issue to deal with statistically, and I will visit it in future work. Fortunately, the redshifts are distributed within the range my mass approximation is accurate.

I attempted to create mock data to match these data. Their distribution is shown in Figure 4. I chose hyperparameters $\chi, \alpha, \beta = 1$. The distribution of redshifts was difficult to reproduce, but the produced richness has a comparable distribution to the "true" data. The mocks were produced with the parameters shown in Table 1. In the next section I attempt to recover these parameters with MCMC.

4. REGRESSION ON MOCK DATA AND RESULTS

An inference was run with the likelihoods shown in Section 2 on the mock data generated in Section 3 using MCMC. The implementation chosen was the emcee sampler Foreman-Mackey et al. (2013). The sampler was first run with 1792 walkers and 100 steps. That number

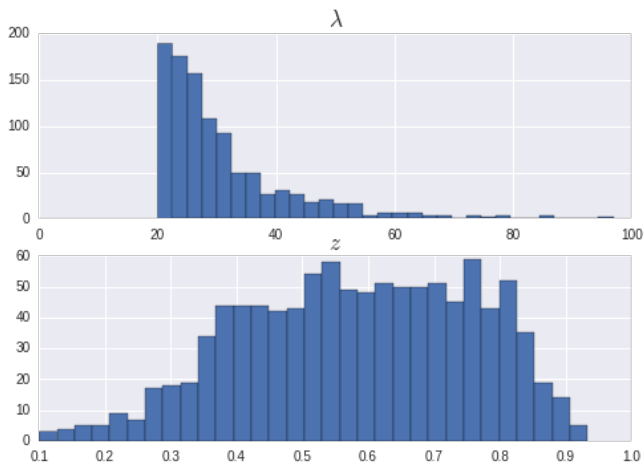


Figure 3. Histograms of richness and redshift for the data used in this analysis.

Table 1
Parameter values used in mock data production.

Parameter	True Value	Recovered Value 1	Recovered Value 2
a	1e-11	1.001e-11	0.950e-11
b	0.89	0.878	0.169
σ	0.10	2.055	1.339

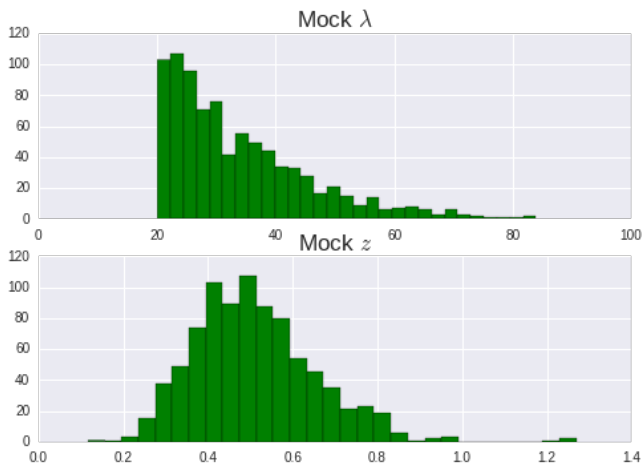


Figure 4. Histograms of mock distributions for richness and redshift used in this analysis.

of walkers was chosen to be twice the number of dimensions which is the minimum number for the algorithm. A corner plot of the samples is shown in Figure 5, and the MAP values are shown in Table 1.

It was run again with a new set of walkers and steps (1000, 1000). In order for this to be computationally tractable, the size of the data had to be reduced. 100 random data points were drawn from the mock data. The recovered values are also shown in Table 1 and the corner plot is Figure 6.

5. DISCUSSION

It is clear from the corner plots in Figures 5 and 6 that there is a problem with this sampling distribution. In both sampling cases, σ was unable to be recovered

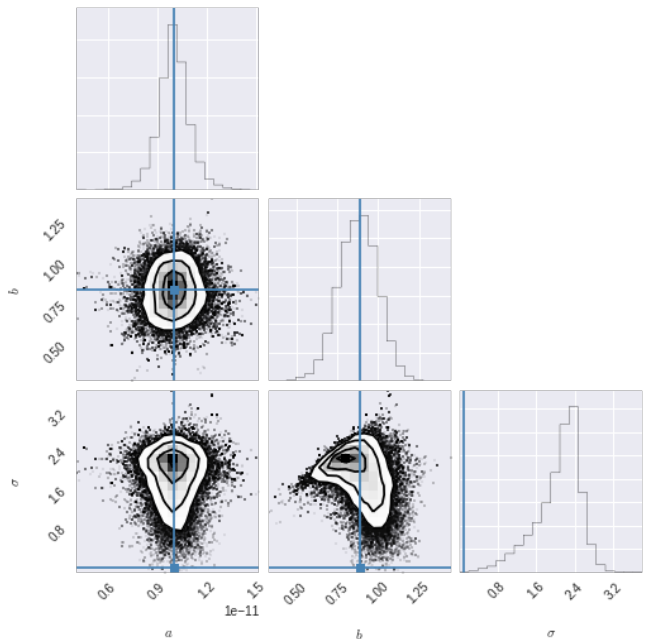


Figure 5. A corner plot of the samples for the first run of the MCMC sampler on mock data, with 1792 walkers and 100 steps.

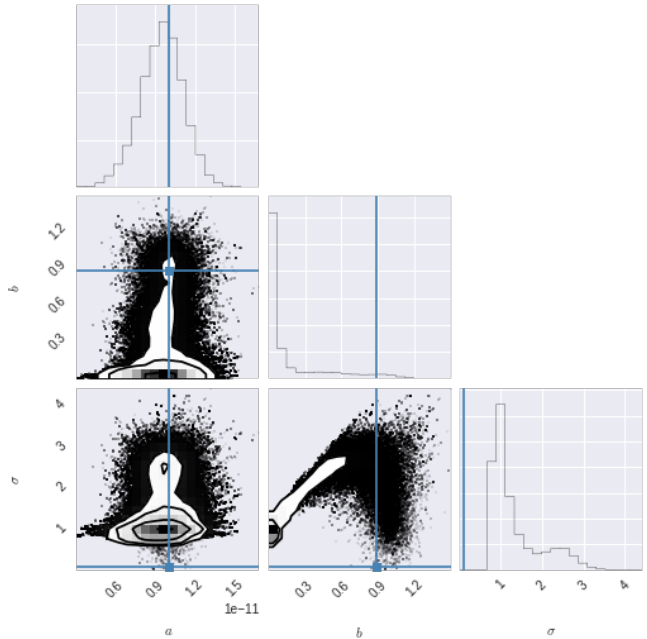


Figure 6. A corner plot of the samples for the second run of the MCMC sampler on mock data, with 1000 walkers and 1000 steps.

successfully at all. It is possible that the noise model is not a good one. The richness may not be lognormally distributed in truth. There is also the possibility of separating the richness likelihood into two separate ones: a "true" likelihood and an "observed" one. This would account for measurement errors separate from intrinsic scatter. However, it is not clear that there is such a notion as a "true" richness, and it will have to be carefully considered.

Possibly more concerning is the differences between the two chains. The first chain was run with a large number

of walkers and appears to have converged. Some of the distributions have unusual structure. However, with an even larger chain the sampler completely fails. It is not clear why the plot of b v. σ has such an odd shape and takes such a strange path to a poor value. This behavior was repeated when the sampler was run against the true richness data from the Buzzard simulation, so it is unlikely a problem with the mock data in particular. It is possible the mass function does not contain enough information about the masses for this inference. It is also possible that a more careful treatment of the truncated data is required.

6. CONCLUSIONS

Conclusions at this point are difficult to draw for this work. Recovery of the mass-richness relation is clearly a complex problem, and more care and thought is required in order to do it correctly. There is a great deal of future work to consider for this project to attempt to solve problems of non-convergent sampling.

1. The validity of the mass function approximation used in this work should be inspected closely.
2. Each portion of the likelihood should be checked to ensure it is not causing any bias.

3. The prospect of separating richness errors into measurement and intrinsic errors will be explored.
4. An explicit connection between richness data from redMaPPer and the halo's true values will be made.
5. The truncation of observed richnesses will be more carefully treated.
6. The MCMC sampler will be moved to a computing system with enough memory to handle a large chain over all data points.

All of this being said, there are some good conclusions which can be drawn from the analysis at this point. This model still shows promise as a procedure for insuring the mass-richness relation. Also, although the sampler was restricted by its size, running the sampler is surprisingly quick. Calculating a chain for 2000 walkers over 100 steps for about 2000 parameters takes only about 5 minutes on a 4 core processor. This suggests this implementation will scale well, and will not be crippled by relaxing a few assumptions. I am optimistic that this will prove to be a successful way to infer the mass-richness relation.

REFERENCES

- Evrard, A. E., Arnault, P., Huterer, D., & Farahi, A. 2014, *Monthly Notices of the Royal Astronomical Society*, 441, 3562
- Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. 2013, *Publications of the Astronomical Society of the Pacific*, 125, pp. 306