Chapter 1

Formalism

With the goal of calibrating the $\lambda{-}$M relationship presented in Ry12, I observed a sample of galaxy clusters using CARMA/SZA. As noted in Chap. **??**, and discussed in detail in Chap. **??**, the SZ signal of a galaxy cluster is expected to serve as a cluster mass proxy with a low amount of intrinsic scatter. In order to fully determine the joint scaling properties the SZ signal and λ in this observed sample, I use a Markov chain Monte Carlo analysis to determine the posterior likelihood of the model parameters given the observations. The constraints from our data come from both the SZ signal and galaxy richness measurements, as well as the fact that the sample represents the 28 richest galaxy clusters in the SDSS DR7 survey volume. In particular, the volume-complete nature of the sample provides leverage on the mass function needed to constrain the scaling relation parameters.

1.1 Halo Mass Function and Preliminaries

The dark matter halo is the fundamental theoretical unit of cosmological structure formation. The halo mass function represents the number of objects in a matter density field that will undergo non-linear collapse and form distinct, gravitationally bound objects. Analytic expressions for the halo mass function have a long history (Press & Schechter, 1974). Modern treatments typically provide fitting formulas calibrated with N-body simulations (Sheth & Tormen, 1999; Jenkins et al., 2001; Bhattacharya et al., 2011). I follow Tinker et al. (2008), using their fits, and define the mass function as

$$\frac{dn}{d\log M} = f(\sigma)\frac{\bar{\rho}_m}{M}\frac{d\log\sigma^{-1}}{d\log M}, \tag{1.1}$$

where $n(M)$ is the number density of halos of a mass, M. The parameter σ^2 is the variance of the linear matter power spectrum weighted by a window function for collapse (a spherical top hat in this case). The function $f(\sigma)$ tracks the first crossings across the barrier between collapsed and non-collapsed regions. The average matter density of the Universe is $\bar{\rho}_m = \Omega_m(z)\rho_{\rm crit}$. We use the fitting functions provided by Eisenstein & Hu (1999) to compute the transfer function used to calculate the matter power spectrum from a featureless primordial power spectrum.

The differential cluster count with respect to mass is

$$\frac{dN}{d\log M} = \frac{dn}{d\log M} V(\Omega, z_{min}, z_{max}). \tag{1.2}$$

The cosmological volume, V, is determined by integrating

$$V(\Omega, z_{min}, z_{max}) = \int_{\Omega} d\Omega \int_{z_{\rm min}}^{z_{\rm max}} dz \frac{c}{H_0} \frac{(1+z)^2 D_A^2(z)}{E(z)} \tag{1.3}$$

over a solid angle Ω and redshift range $z_{\rm min}$ to $z_{\rm max}$. The function $D_A^2(z)$ is the square of the angular diameter distance, c is the speed of light, and $E(z)H_0$ is the value of the Hubble constant at redshift z. In practice, I use the analytic expression for the volume provided by Carroll et al. (1992) as it improves the speed of computation.

In order to relate the observed cluster quantities to the dark matter halo mass function, we must specify a differential cluster count with respect to the observable. The probability of any particular halo from a volume-complete sample having a mass M_i is proportional to dN/dM evaluated at M_i. If the halo has an observable quantity q, and the probability density of q given M is $P(\log q|M)$, then the differential number of clusters with respect to q is the mass function weighted by $P(\log q|M)$ marginalized over $\log M$,

$$\frac{dN}{d\log q} = \int_{\log M} d\log M \frac{dN}{d\log M} P(\log q|M). \tag{1.4}$$

We model $P(\log q|M)$ as a normal distribution and perform the marginalization over a range of masses encompassing the mass range of interest, typically 10^{14} M$_\odot$ to 2×10^{16} M$_\odot$, chosen so that our results are not sensitive to small changes in the integration limits.

1.2 General Likelihood Derivation

We order a sample of N_{cl} halos, where each cluster i has a RedMaPPer richness value $\lambda = \lambda_i$, and a SZ signal $Y_{SZ} = Y_i$, such that λ_1 is richness of the richest cluster. The second richest clusters has $\lambda = \lambda_2$ down to $\lambda = \lambda_{N_{cl}}$ for the least rich cluster. The data vector, $\mathbf{d}$, for N_{cl} clusters is

$$\mathbf{d} = \left\{(\lambda_1, Y_1), (\lambda_2, Y_2), ..., (\lambda_{N_{cl}}, Y_{N_{cl}})\right\}. \tag{1.5}$$

Due to the intrinsic scatter relative to mass, Y_1 will not necessarily be the largest value of the Y_{SZ} parameter for the sample, and there is no similar order for Y_{SZ}.

Given $\mathbf{d}$, the likelihood function consists of two parts.

1. A cluster with richness λ_i and SZ signal Y_i must exist in the survey volume.

2. There can be no clusters richer than λ_i and less rich than λ_{i-1}. If $i = 1$, there can be no halos richer than λ_1.

The probability that a halo with richness λ_i exists is proportional to the total differential halo count in the survey volume relative to the observable λ evaluated at λ_i; $P(\lambda_i) \propto dN/d\lambda|_{\lambda_i}$. For the second component of the likelihood, we treat the likelihood that there are no halos with richness values between the adjacent, ordered richness values of the halos in the sample as a Poisson distribution. For a halo i the probability that there are zero halos with richness greater than λ_i and less than λ_{i-1} is equal to the Possion probability with a mean of zero,

$$P\left(N\left(\lambda_{i-1} > \lambda \geq \lambda_i\right) = 0\right) = \exp\left(-\langle N\left(\lambda_{i-1} > \lambda \geq \lambda_i\right)\rangle\right). \tag{1.6}$$

The expression $\langle N\left(\lambda_{i-1} > \lambda \geq \lambda_i\right)\rangle$ in Eqn. 1.6 is the expectation value for the number of halos with $\lambda_{i-1} > \lambda \geq \lambda_i$ and is defined by

$$\langle N\left(\lambda_{i-1} > \lambda \geq \lambda_i\right)\rangle = \int_{\log\lambda_i}^{\log\lambda_{i-1}} d\log\lambda \frac{dN}{d\log\lambda}. \tag{1.7}$$

The probability that λ_i is the i^{th} largest richness measurement in the sample is proportional to the product of the probabilities of each of the criteria,

$$P\left(\lambda_i \mid \lambda_{i-1}\right) \propto \left.\frac{dN}{d\lambda}\right|_{\lambda_i} \times \exp\left(-\langle N(\lambda_{i-1} > \lambda \geq \lambda_i)\rangle\right). \tag{1.8}$$

Descending the ordered list $\mathbf{d}$, the probability that λ_1 is the richest halo is

$$P(\lambda_1) \propto \left.\frac{dN}{d\lambda}\right|_{\lambda_1} \times \exp\left(-\langle N(\lambda \geq \lambda_1)\rangle\right), \tag{1.9}$$

because there are no halos richer than λ_1. The probability that λ_1 and λ_2 are the first and second richest halos is

$$P(\lambda_1, \lambda_2) = P(\lambda_1) \times P\left(\lambda_2 \mid \lambda_1\right). \tag{1.10}$$

The expression for $P\left(\lambda_2 \mid \lambda_1\right)$ is evaluated in a similar way to $P(\lambda_1)$. There must be a cluster with richness λ_2 and no clusters of richness between λ_2 and λ_1. That is

$$P\left(\lambda_2 \mid \lambda_1\right) \propto \left.\frac{dN}{d\lambda}\right|_{\lambda_2} \times \exp\left(-\langle N(\lambda_2 < \lambda < \lambda_1)\rangle\right). \tag{1.11}$$

Note that

$$\begin{aligned} P(\lambda_1, \lambda_2) &= P(\lambda_1) \times P\left(\lambda_2 \mid \lambda_1\right), \\ &\propto \left.\frac{dN}{d\lambda}\right|_{\lambda_1} \left.\frac{dN}{d\lambda}\right|_{\lambda_2} \exp\left(-\langle N(\lambda > \lambda_1)\rangle\right) \exp\left(-\langle N(\lambda_2 < \lambda < \lambda_1)\rangle\right), \\ &\propto \left.\frac{dN}{d\lambda}\right|_{\lambda_1} \left.\frac{dN}{d\lambda}\right|_{\lambda_2} \exp\left(-\langle N(\lambda > \lambda_2)\rangle\right). \end{aligned} \tag{1.12}$$

This expression can be generalized to

$$P(\lambda_1, ..., \lambda_{N_{cl}}) \propto \exp\left(-\left\langle N(\lambda > \lambda_{N_{cl}})\right\rangle\right) \times \prod_{i=1}^{N_{cl}} \left.\frac{dN}{d\lambda}\right|_{\lambda_i}. \tag{1.13}$$

The cluster must also exist with the correct value of Y_i. Only the $dN/d\lambda$ portion of the likelihood is modified. The full expression is

$$P(\mathbf{d}) \propto \exp\left(-\left\langle < N(\lambda > \lambda_{N_{cl}})\right\rangle\right) \times \prod_{i=1}^{N_{cl}} \left.\frac{dN}{d\lambda dY}\right|_{(\lambda_i, Y_i)}. \tag{1.14}$$

We model the observables as log-normal distributions around the mass, so integrate over $d\log M$. Then,

$$\left.\frac{dN}{d\log\lambda d\log Y}\right|_{(\lambda_i, Y_i)} = \int d\log M \frac{dN}{d\log M} P(\log\lambda_i, \log Y_i|M), \tag{1.15}$$

and

$$\begin{aligned} \left\langle N(\lambda > \lambda_{N_{cl}})\right\rangle &= \int_{\log\lambda_{N_{cl}}}^{\infty} d\log\lambda \frac{dN}{d\log\lambda} \\ &= \int_{\log\lambda_{N_{cl}}}^{\infty} d\log\lambda \int_{\log M} d\log M \frac{dN}{d\log M} P\left(\log\lambda \mid M\right). \end{aligned} \tag{1.16}$$

In a fixed cosmology, mass calibration involves determining the parameters of $P(\log\lambda_i, \log Y_i|M)$ and $P\left(\log\lambda \mid M\right)$ through a fit to data.

1.3 Likelihood Estimator with Measurement Noise

The probability derivation in Sec. 1.2 is valid in the limit where observational errors on the cluster Y_i and λ_i are small. The errors on the λ_i are typically one or two percent (Rykoff et al., 2012), but the observational uncertainty on Y_i can be larger. We model the observed value of Y for a given cluster, Y_i, as a probability distribution around an unknown true value, Y^t, and assume that the errors are uncorrelated. In this case, the probability of measuring Y_i given Y^t, $P\left(Y_i \,|\, Y^t\right)$, serves as a smoothing window on the likelihood distribution,

$$P(\lambda, Y_i) = \int_{Y^t} dY^t P\left(Y_i \,|\, Y^t\right) P(\lambda, Y^t). \tag{1.17}$$

$P(\lambda, Y^t)$ is the joint differential cluster count relative to both λ and Y, similar to Eqn. 1.14. We assume a Gaussian functional form for $P\left(Y_i \,|\, Y^t\right)$, except in the case when SZ signal is not detected. The case of missing SZ information is treated in the next section. The Gaussian $P\left(Y_i \,|\, Y^t\right)$ is centered on Y_i, with variance of Y_i measured from the data, $\sigma^2_{Y_i}$. In general,

$$P\left(Y_i \,|\, Y^t\right) = \frac{1}{\sqrt{2\pi\sigma^2_{Y_i}}} \exp\left(-\frac{(Y_i - Y^t)^2}{2\sigma^2_{Y_i}}\right). \tag{1.18}$$

Extending this analysis to a sample of N_{cl} clusters involves taking the product of each individual convolution. Modifying Eqn. 1.14, the likelihood is:

$$P(\mathbf{d}) \propto \exp\left(-\langle N(\lambda > \lambda_\nu)\rangle\right) \times \prod_{i=1}^{N_{cl}} \int_{Y^t} d\log Y^t P\left(Y_i \,|\, Y^t\right) \left.\frac{dN}{d\log\lambda d\log Y}\right|_{(\lambda_i, Y^t)}. \tag{1.19}$$

1.4 Missing Data

The likelihood derived thus far assumes the existence of Y measurements for every cluster in the sample. As discussed in Chap. **??**, 5 of the 28 clusters in the maxBCG-SZ sample do not have Y measurements. As the derivation of the likelihood depends on the volume-complete, λ-selected nature of the sample, it is impossible to simply ignore the clusters with missing Y measurements. Therefore it is necessary to modify the likelihood to take this missing Y information into account. As we parameterize the Y observations as a probability distribution, we marginalize over all possible Y_i.

$$\int_{Y_i} \frac{dN}{d\lambda dY}\bigg|_{(\lambda_i,Y_i)} dY_i = \frac{dN}{d\lambda}\bigg|_{\lambda_i}. \tag{1.20}$$

Marginalizing over all possible observed values of Y_i is the most conservative choice to make for the undetected clusters. It is possible to model the non-detections as upper limits, which can be included with little modification to the formalism. In practice, however, the calculation for an upper limit on Y_{SZ} in these observations is not trivial. Due to the spatial filtering of the interferometer (see Chap. **??**) and the potentially large offsets between the SZ centroid and the BCG at which λ is measured (see Chap. **??**), any upper limit requires a number of complicating assumptions. Any upper limit information will only improve the constraints. In theory, this treatment of missing data allows us to extend our sample to a much larger number of clusters, given that the maxBCG catalog contains over 13,000 entries. In practice, the computational effort needed to evaluate the likelihood scales with the number of clusters, so there is a break-even point. Simulations of the experiment indicated that the improvements in the constraints on the $\lambda-$M relation from adding another 25 clusters were minimal, so this work consists of only the 28 clusters which have SZ observations.

1.5 Implementation

The $\lambda-$M and $Y-$M relationships are encapsulated in the probability distributions that a halo will have observables λ and Y given mass M. We assume that these probability distributions are lognormal, which is a reasonable approximation (Cunha, 2009). The mean observable-mass relationships are defined by normalizations, A_{obs}, and slopes, B_{obs}, around a pivot point M_*. The observables each have an intrinsic logarithmic scatter about the mean relation at fixed log mass, $\sigma_{\lambda|M}$ and $\sigma_{Y|M}$ which are correlated with coefficient ρ. The covariance matrix is

$$\mathbf{C} = \begin{pmatrix} \sigma^2_{\lambda|M} & \rho\sigma_{\lambda|M}\sigma_{Y|M} \\ \rho\sigma_{\lambda|M}\sigma_{Y|M} & \sigma^2_{Y|M} \end{pmatrix} \tag{1.21}$$

For a given mass M, the mean relations are

$$\langle \log Y|M\rangle = A_Y + B_Y \log \frac{M}{M_*} \tag{1.22}$$

$$\langle \log \lambda|M\rangle = A_\lambda + B_\lambda \log \frac{M}{M_*}. \tag{1.23}$$

We arrange the observable relations as a vector $\mathbf{x} = (\log \lambda_i - \log \lambda_M, \log Y^t - \log Y_M)$, and the probability $P\left(\log \lambda, \log Y \mid M\right)$ is a bivariate Gaussian,

$$P\left(\log \lambda, \log Y \mid M\right) = \frac{1}{2\pi|\mathbf{C}|^{1/2}} \exp\left(-\frac{1}{2}\mathbf{x}^T\mathbf{C}^{-1}\mathbf{x}\right). \tag{1.24}$$

The values of $\log \lambda_M$ and $\log Y_M$ are the values predicted by the mean observable mass relations for λ and Y_{SZ}.

For these seven parameters, $\theta = \{A_\lambda, B_\lambda, \sigma_\lambda, A_Y, B_Y, \sigma_Y, \rho\}$ we evaluate the likelihood

$$\log \mathcal{L}(\theta) = \log P(\mathbf{d}|\theta) \tag{1.25}$$

using a Markov chain Monte Carlo analysis using a Metropolis-Hastings algorithm (Hastings, 1970). The full probability model is determined using Eqn. 1.19. Putting everything together, the likelihood, in the case where the cluster has both a λ and Y_{SZ} measurment, is implemented as follows:

$$\mathcal{L}(\theta) \propto \exp\left(-\int_{\ln\lambda_{N_{cl}}}^{\infty} d\log\lambda \int_M d\log M \frac{dN}{d\log M}\frac{1}{\sqrt{2\pi\sigma_{\lambda|M}}}\exp\left(-\frac{(\log\lambda-\log\lambda_M)^2}{2\sigma^2_{\lambda|M}}\right)\right)\times\ldots$$

$$\prod_{i=1}^{N_{cl}}\int_{Y^t} dY^t \frac{\exp\left(-\frac{(Y_i-Y^t)^2}{2\sigma^2_{Y_i}}\right)}{\lambda_i\sqrt{2\pi\sigma^2_{Y_i}}}\int_M d\log M\frac{dN}{d\log M}\frac{\exp\left(-\frac{1}{2}\mathbf{x}^T\mathbf{C}^{-1}\mathbf{x}\right)}{2\pi|\mathbf{C}|^{1/2}}. \quad (1.26)$$

The set of 7 parameters in the observable-mass relations represents a minimal set. For larger analyses, these parameters can be more complicated. In particular, the maxBCG-SZ sample in this work spans a narrow redshift range, $z \in [0.2, 0.3]$. For larger redshift ranges, the observable-mass relationships can be modeled as redshift-dependent quantities. For this work, however, we make use of the minimal model defined above. For the analysis of this sample, we assume a fixed cosmology, although the framework is general enough so that cosmological parameters can be included in θ. Priors are implemented in any of three different ways. There are hard limits. For example, the covariance between the intrinsic scatters of the mass-observable relationship is required to be $\rho \in [-1, 1]$. In the case of the scatters, a flat prior can be placed on the intrinsic *variance* around the mass-observable relationship. Finally, any parameter may have a Gaussian prior, with a defined central value and width. Unless otherwise noted, parameters fit using this implementation use a non-informative uniform prior on the parameter.

1.6 Determing the Integration Limit for Y_{SZ}

References

Bhattacharya, S., Heitmann, K., White, M., et al. 2011, ApJ, 732, 122

Carroll, S. M., Press, W. H., & Turner, E. L. 1992, ARA&A, 30, 499

Cunha, C. 2009, Phys. Rev. D, 79, 063009

Eisenstein, D. J., & Hu, W. 1999, ApJ, 511, 5

Hastings, W. K. 1970, Biometrika, 57, 97

Jenkins, A., Frenk, C. S., White, S. D. M., et al. 2001, MNRAS, 321, 372

Press, W. H., & Schechter, P. 1974, ApJ, 187, 425

Rykoff, E. S., Koester, B. P., Rozo, E., et al. 2012, ApJ, 746, 178

Sheth, R. K., & Tormen, G. 1999, MNRAS, 308, 119

Tinker, J., Kravtsov, A. V., Klypin, A., et al. 2008, ApJ, 688, 709