HOW TO USE CENTERING PROBABILITIES OF GALAXY CLUSTERS FOR COSMOLOGICAL INVESTIGATIONS

EDUARDO ROZO[1], MATTHEW BECKER

Draft version May 28, 2013

ABSTRACT

Abstract goes here.

Subject headings: cosmology: lensing

1. INTRODUCTION

Stuff goes here.

2. CLUSTER ABUNDANCES

To estimate the richness function $N(\lambda)$ one traditionally bins galaxy clusters in richness bins $[\lambda_{min}, \lambda_{max}]$. However, if all clusters have several possible centers, and each center has its own associated richness, how is one to decide which bin does any given cluster belong to? Here, we demonstrate how to use centering probabilities to resolve this question.

2.1. *Estimating the Richness Function in the Presence of Observational Errors: Mean and Variance*

We wish to estimate the number of galaxy clusters in a richness bin characterized by the binning function ψ_a, such that

$$\psi_a(\lambda) = \begin{cases} 1 & \lambda \in [\lambda^a_{min}, \lambda^a_{max}] \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

The richness λ is the true cluster richness, evaluated at the correct (but unknown) cluster center. The number of galaxy clusters in bin a, denoted N_a, can be written simply as

$$N_a = \sum_\alpha \psi^\alpha_a \tag{2}$$

where the index α indexes the galaxy clusters, and ψ^α_a is the value of $\psi_a(\lambda_\alpha)$.

Assuming the richness bins are non-overlapping, any given cluster can belong to one and only one bin, so ψ^α_a is a variable that is either zero or one with probability q^α_a. The reason we denote this probability q rather than p will be made apparent later. The expectation value of N_a is

$$\langle N_a \rangle = \sum_\alpha q^\alpha_a. \tag{3}$$

The covariance matrix can also be easily computed,

$$\mathrm{Cov}(N_a, N_b) = \sum_{\alpha\beta} \mathrm{Cov}(\psi^\alpha_a, \psi^\beta_b). \tag{4}$$

Since clusters α and β are independent,

$$\mathrm{Cov}(\psi^\alpha_a, \psi^\beta_b) = \delta_{\alpha,\beta}\mathrm{Cov}(\psi^\alpha_a, \psi^\alpha_b). \tag{5}$$

To compute $\mathrm{Cov}(\psi^\alpha_a, \psi^\alpha_b)$, we note that a cluster can only belong to one bin, so $\psi^\alpha_a \psi^\beta_b = \delta_{a,b}\psi^\alpha_a$. Thus,

$$\langle \psi^\alpha_a \psi^\alpha_b \rangle = \delta_{a,b} \langle \psi^\alpha_a \rangle = \delta_{a,b} q^\alpha_a \tag{6}$$

and therefore

$$\mathrm{Cov}(\psi^\alpha_a, \psi^\alpha_b) = \delta_{a,b} q^\alpha_a - q^\alpha_a q^\alpha_b. \tag{7}$$

Putting it all together,

$$\mathrm{Cov}(N_a, N_b) = \delta_{a,b} \sum_\alpha q^\alpha_a - \sum_\alpha q^\alpha_a q^\alpha_b \tag{8}$$

$$= \delta_{a,b} \langle N_a \rangle - \sum_\alpha q^\alpha_a q^\alpha_b. \tag{9}$$

In the limit of deterministic binning, q^α_a is either zero or one, and $q^\alpha_a q^\alpha_b = \delta_{a,b} q^\alpha_a$. Consequently, in this limit $\mathrm{Cov}(N_a, N_b) = 0$, and there is no error in the observationally estimated abundance function, as expected. Note too that the variance $\mathrm{Var}(N_a)$ due to random binning is always less than Poisson, so we generically expect this source of error to be sub-dominant in cluster abundance studies.

2.2. *Relation to Miscentering*

In the above section, we showed how given the binning probability q^α_a, one can estimate the abundance function of galaxy clusters in the presence of scatter, along with the corresponding observational error. For our purposes, miscentering is simply a form of observational scatter that must be incorporated into the calculation of q^α_a.

Let then λ_α be the true richness of cluster α about its cluster center, and ψ_a be the binning function for bin a. For simplicity, we will assume that there are only two possible centers for cluster α, with probabilities p_1 and p_2, such that $p_1 + p_2 = 1$. The labels are chosen so center 1 is the most likely center, i.e. $p_1 > p_2$. The binning probability is given simply by

$$q^\alpha_a = \langle \psi^\alpha_a \rangle = \int d\lambda_\alpha\, P(\lambda_\alpha | \lambda_1, \lambda_2) \psi_a(\lambda_\alpha) \tag{10}$$

where the probability $P(\lambda_\alpha|\lambda_1, \lambda_2)$ is the posterior of the richness of galaxy cluster α. Our job then is to recover an expression for this posterior.

Let then λ_T be the true richness of the galaxy cluster, and λ_F be the richness about the false center. The richnesses λ_1 and λ_2 take the form

$$\lambda_1 = z_1\lambda_T + z_2\lambda_F + \Delta_1 \tag{11}$$

$$\lambda_2 = z_2\lambda_T + z_1\lambda_F + \Delta_2 \tag{12}$$

where Δ_1 and Δ_2 are the observational noise about cluster centers 1 and 2, and z_1 is a random variable that governs whether the cluster is properly centered or not. z_1 is always either 0 (miscentered cluster) or 1 (properly centered cluster), and its expectation value is the centering probability p_1, $p_1 = \langle z_1 \rangle$. The variable $z_2 = 1 - z_1$. Defining $P_\Delta(\Delta_1, \Delta_2)$ as the probability distribution characterizing the observational errors, we have

$$P(\lambda_1, \lambda_2 | \lambda_T, \lambda_F) = p_1 P_\Delta(\lambda_1 - \lambda_T, \lambda_2 - \lambda_F) + p_2 P_\Delta(\lambda_1 - \lambda_F, \lambda_2 - \lambda_T). \tag{13}$$

[1] Einstein Fellow

The probability distribution $P(\lambda_T, \lambda_F|\lambda_1, \lambda_2)$ can then be recovered by application of Bayes's theorem,

$$P(\lambda_T, \lambda_F) = P(\lambda_1, \lambda_2|\lambda_T, \lambda_F)\frac{P_0(\lambda_T, \lambda_F)}{P_0(\lambda_1, \lambda_2)} \tag{14}$$

The denominator $P_0(\lambda_1, \lambda_2)$ is simply a normalization constant K. As for our prior, given cosmological parameters and scaling relations, we can write a physical model for $P_0(\lambda_T)$. For λ_F, we assume $\lambda_F = \lambda_T + \delta$, where δ is a random offset with probability P_δ. We have then

$$P_0(\lambda_T, \lambda_F) = P_0(\lambda_F|\lambda_T)P_0(\lambda_T) = P_\delta(\lambda_F - \lambda_T)P_0(\lambda_T), \tag{15}$$

which is simply proportional to $P_0(\lambda_T)$ if we assume a constant prior for P_δ. Putting everything together, we arrive at

$$\begin{aligned} P(\lambda_T, \lambda_F) = {} & p_1 K^{-1} P_0(\lambda_T) P_\Delta(\lambda_1 - \lambda_T, \lambda_2 - \lambda_F) \\ & + p_2 K^{-1} P_0(\lambda_T) P_\Delta(\lambda_1 - \lambda_F, \lambda_2 - \lambda_T) \end{aligned} \tag{16}$$

The probability $P(\lambda_T|\lambda_1, \lambda_2)$ is obtained by marginalizing over λ_F. Assuming that the error distributions are Gaussian, we can readily integrate over λ_F to arrive at

$$P(\lambda_T|\lambda_1, \lambda_2) = K^{-1}P_0(\lambda_T)\sum p_i G(\lambda_T|\lambda_i, \sigma_i) \tag{17}$$

where $G(\lambda_T|\lambda_i, \sigma_i)$ is a Gaussian distribution of mean λ_i and standard deviation σ_i.

We can now readily integrate this posterior distribution over the binning function to arrive at the binning probability q_a^α,

$$q_a^\alpha = \sum_i p_i \left\langle \psi_a(\lambda_\alpha) \right\rangle_i \tag{18}$$

where

$$\left\langle \psi_a(\lambda_\alpha) \right\rangle_i = K^{-1} \int d\lambda_\alpha \ P_0(\lambda_\alpha) G(\lambda_\alpha|\lambda_i, \sigma_i) \psi_a(\lambda_\alpha). \tag{19}$$

The prior $P_0(\lambda_\alpha)$ can depend on model parameters $\mathbf{\Omega}$ characterizing the abundance function of galaxy clusters, and the parameters themselves can be unknown.

2.3. *Connecting Theory to Data in the Presence of Miscentering*

In section 2.1, we saw how to compute the expectation value and variance of the abundance function $\mathbf{N}_{obs}$. The distribution $P_{obs}(\mathbf{N}_{obs}|\mathcal{D}, \mathbf{\Omega})$ characterizes the observational errors in the abundance function $\mathbf{N}$ of the Universe. Note that since both the mean and covariance matrix of $\mathbf{N}_{obs}$ depend on the prior $P_0(\lambda|\mathbf{\Omega})$, the distribution $P_{obs}(\mathbf{N}|\mathcal{D}, \mathbf{\Omega})$ depends not just on the data $\mathcal{D}$ but also on the parameters of interest $\mathbf{\Omega}$.

Our theoretical model assumes the abundance function $\mathbf{N}_{th}$ is a random realization of a probability distribution $P_{th}(\mathbf{N}_{th}|\mathbf{\Omega})$. The variance for P_{th} is the standard Poisson and sample variance. Since these two sources of variance are independent of the sources of variance for the observed abundance function $\mathbf{N}_{obs}$, the joint distribution for $\mathbf{N}_{th}$ and $\mathbf{N}_{obs}$ is

$$P(\mathbf{N}_{th}, \mathbf{N}_{obs}|\mathcal{D}, \mathbf{\Omega}) = P_{th}(\mathbf{N}_{th}|\mathbf{\Omega})P_{obs}(\mathbf{N}_{obs}|\mathcal{D}, \mathbf{\Omega}). \tag{20}$$

Of course, in reality, we must enforce the condition $\mathbf{N}_{obs} = \mathbf{N}_{true}$, so the joint probability takes the form

$$P(\mathbf{N}_{th}, \mathbf{N}_{obs}|\mathcal{D}, \mathbf{\Omega}) \propto P_{th}(\mathbf{N}_{th}|\mathbf{\Omega})P_{obs}(\mathbf{N}_{obs}|\mathcal{D}, \mathbf{\Omega})\delta(\mathbf{N}_{obs} - \mathbf{N}_{th}). \tag{21}$$

Given a prior $P_0(\mathbf{\Omega}, \mathcal{D})$, the probability $P(\mathbf{N}_{th}, \mathbf{N}_{obs}, \mathbf{\Omega}, \mathcal{D})$ is

$$P(\mathbf{N}_{th}, \mathbf{N}_{obs}, \mathbf{\Omega}, \mathcal{D}) = P(\mathbf{N}_{th}, \mathbf{N}_{obs}|\mathcal{D}, \mathbf{\Omega})P_0(\mathbf{\Omega}, \mathcal{D}). \tag{22}$$

Adopting a separable prior for $\mathbf{\Omega}$ and $\mathcal{D}$, we have then

$$\begin{aligned} P(\mathbf{N}_{th}, \mathbf{N}_{obs}, \mathbf{\Omega}|\mathcal{D}) &= \frac{P(\mathbf{N}_{th}, \mathbf{N}_{obs}, \mathbf{\Omega}, \mathcal{D})}{P_0(\mathcal{D})} && (23) \\ &= P_0(\mathbf{\Omega})P(\mathbf{N}_{th}, \mathbf{N}_{obs}|\mathcal{D}, \mathbf{\Omega}). && (24) \end{aligned}$$

The posterior in the parameters of interest is obtained by marginalizing over $\mathbf{N}_{th}$ and $\mathbf{N}_{obs}$,

$$P(\mathbf{\Omega}|\mathcal{D}) = P_0(\mathbf{\Omega}) \int d\mathbf{N}_{th} d\mathbf{N}_{obs} \ P(\mathbf{N}_{th}, \mathbf{N}_{obs}|\mathbf{\Omega}, \mathcal{D}). \tag{25}$$

The first of these integrals is trivial thanks to the δ function requiring that $\mathbf{N}_{obs} = \mathbf{N}_{true}$. Letting $\mathbf{N}$ be the remaining integration variable, we arrive at

$$P(\mathbf{\Omega}|\mathcal{D}) = P_0(\mathbf{\Omega}) \int d\mathbf{N} \ P_{th}(\mathbf{N}|\mathbf{\Omega})P_{obs}(\mathbf{N}|\mathcal{D}, \mathbf{\Omega}). \tag{26}$$

When both P_{obs} and P_{th} are Gaussian distribution, the integral is a convolution of two Gaussian, which is itself a Gaussian. The posterior reduces to

$$P(\mathbf{\Omega}|\mathcal{D}) \propto P_0(\mathbf{\Omega}) \exp\left(-\frac{1}{2}\mathbf{\Delta} \cdot \mathbf{C}_{tot}^{-1} \cdot \mathbf{\Delta}\right) \tag{27}$$

where

$$\begin{aligned} \mathbf{\Delta} &= \left\langle \mathbf{N}|\mathcal{D}, \mathbf{\Omega} \right\rangle_{obs} - \left\langle \mathbf{N}|\mathbf{\Omega} \right\rangle_{th} && (28) \\ \mathbf{C}_{tot} &= \mathbf{C}_{th} + \mathbf{C}_{obs} && (29) \end{aligned}$$

Here, $\langle \cdot \rangle_{obs}$ and $\langle \cdot \rangle_{th}$ correspond to expectation values with respect to the probabilities P_{obs} and P_{th} respectively.

3. STACKING OF CENTERING-DEPENDENT CLUSTER OBSERVABLES

We now turn to the problem of estimating a cluster observable X which depends on the assigned cluster center. In general, we do not know where the true cluster center lies, but rather have a variety of possible centers $i = 1, \ldots, \nu$ about which we can measure the quantity X. The problem we now consider is how one can estimate the quantity X about the true cluster center given the values $X_1, \ldots, X_\nu$. We restrict ourselves to the case $\nu = 2$, i.e. each cluster has at most two likely cluster centers.

3.1. *The Basic Idea*

Let X be a cluster observable — e.g. weak lensing tangential shear — that depends on the assigned cluster center, and let X_T be the value of X about the true cluster center. Likewise, X_F is defined as the value of X about the second, incorrect center. We also define X_1 and X_2 as the value of X measured about the most likely center, and the second most likely center. Finally, we define p_1 as the probability that the most likely center is the correct center, and p_2 as the probability that the second most likely center is correct. By assumption, $p_1 + p_2 = 1$. Given a large ensemble of said clusters, the expectation value for X_1 and X_2 are simply

$$\begin{aligned} \left\langle X_1|\mathcal{G} \right\rangle &= p_1 X_T + p_2 X_F && (30) \\ \left\langle X_2|\mathcal{G} \right\rangle &= p_2 X_T + p_1 X_F. && (31) \end{aligned}$$

where we have explicitly denoted the fact that the expectation values of X_1 and X_2 depend on the galaxy data at hand. We can readily invert these relations to arrive at

$$X_T = \frac{1}{p_1 - p_2}\left[p_1 \left\langle X_1|\mathcal{G} \right\rangle - p_2 \left\langle X_2|\mathcal{G} \right\rangle\right] \tag{32}$$

$$X_F = \frac{1}{p_1 - p_2}\left[-p_2 \left\langle X_1|\mathcal{G} \right\rangle + p_1 \left\langle X_2|\mathcal{G} \right\rangle\right]. \tag{33}$$

We can use equation 32 to define an estimator $\hat{X}_T$ for the value of X about the properly centered cluster,

$$\hat{X}_T = \frac{1}{p_1 - p_2}\left[p_1 X_1 - p_2 X_2\right]. \tag{34}$$

Upon taking expectation values, equation 32 guarantees that $\hat{X}_T$ is indeed an unbiased estimator for X_T. In other words, even though we do not know what the correct center is, we can still estimate the value of X about the correct cluster center, so long as we know the centering probabilities p_1 and p_2.

In practice, we find the above estimator to be numerically unstable, which can easily be seen by the fact that the prefactor $1/(p_1-p_2)$ diverges as $p_1 \rightarrow p_2$. Consequently, in practice, we use a somewhat more elaborate estimator as detailed below. The basic idea behind said estimator, however, is no different from this quick sketch.

3.2. *Ensemble Averaging or Stacking*

Consider a galaxy cluster α with two possible centers, and centering probability p_1^α and p_2^α for the most likely center and second most likely center respectively. Letting X_T and X_F be the value of an observable X about the correct (true) cluster center and the incorrect (false) center of the cluster, then the expectation value of X about centers 1 and 2 is given by equations 30 and 31. We can also readily estimate the covariance matrix of X_1 and X_2 in this model. Specifically, since a fraction p_1 of the clusters has the most likely center equal to the true cluster center, one has that

$$\left\langle X_1^2|\mathcal{G}\right\rangle = p_1(X_T^2+\Delta X_1)^2 + p_2(X_F^2+\Delta X_1)^2 \tag{35}$$
$$\left\langle X_2^2|\mathcal{G}\right\rangle = p_2(X_T^2+\Delta X_2)^2 + p_1(X_F^2+\Delta X_2)^2 \tag{36}$$
$$\langle X_1 X_2|\mathcal{G}\rangle = X_T X_F + r\Delta X_1 \Delta X_2 \tag{37}$$

Here, ΔX_1 is the measurement error of X_1, and ΔX_2 is the measurement error of X_2. X_1 and X_2 are allowed to be correlated, with a correlation coefficient r. In conjunction with equations 30 and 31, these equations can be used to solve for $\mathrm{Var}(X_1)$, $\mathrm{Var}(X_2)$, and $\mathrm{Cov}(X_1,X_2)$. We find

$$\mathrm{Var}(X_1|\mathcal{G}) = \Delta X_1^2 + p_1 p_2 (X_T - X_F)^2 \tag{38}$$
$$\mathrm{Var}(X_2|\mathcal{G}) = \Delta X_2^2 + p_1 p_2 (X_T - X_F)^2 \tag{39}$$
$$\mathrm{Cov}(X_1,X_2|\mathcal{G}) = r\Delta X_1 \Delta X_2 - p_1 p_2 (X_T - X_F)^2. \tag{40}$$

The dependence of these variance terms on $\mathcal{G}$ is through the observational errors ΔX_1 and ΔX_2, as well as the centering probabilities p_1 and p_2.

Consider then a cluster ensemble, where X_T^α and X_F^α are the values of X about the true and false cluster centers for cluster α. We assume X_T^α and X_F^α are drawn from a probability distribution $P(X_T,X_F|\mathbf{\Omega})$, where $\mathbf{\Omega}$ are model parameters. We adopt a Gaussian probability distribution, so the appropriate model parameters are the model means X_T and X_F, as well as the associated variance σ_T^2 and σ_F^2, and the correlation coefficient r_{TF}. Further, for each cluster α, we assume that the probability $P(X_1^\alpha,X_2^\alpha|X_T^\alpha,X_F^\alpha;\mathcal{G})$ is a Gaussian distribution with mean and variance as detailed above. Consequently, the probability distribution $P(X_1^\alpha,X_2^\alpha|\mathbf{\Omega})$ is

$$P(X_1^\alpha,X_2^\alpha|\mathbf{\Omega}) = \int dX_T^\alpha dX_F^\alpha \; P(X_1^\alpha,X_2^\alpha|X_T^\alpha,X_F^\alpha) P(X_T^\alpha,X_F^\alpha|\mathbf{\Omega}). \tag{41}$$

This is a simple convolution of Gaussians, so $P(X_1^\alpha,X_2^\alpha|\mathbf{\Omega})$ is itself a Gaussian distribution. The mean is again given by equations 30 and 31, but with X_T and X_F now interpreted as above, while the full covariance matrix is

$$\mathrm{Var}(X_1|\mathcal{G}) = \sigma_T^2 + \Delta X_1^2 + p_1 p_2 (X_T - X_F)^2 \tag{42}$$
$$\mathrm{Var}(X_2|\mathcal{G}) = \sigma_F^2 + \Delta X_2^2 + p_1 p_2 (X_T - X_F)^2 \tag{43}$$
$$\mathrm{Cov}(X_1,X_2|\mathcal{G}) = r_{TF}\sigma_T\sigma_F + r\Delta X_1 \Delta X_2 - p_1 p_2 (X_T - X_F)^2 \tag{44}$$

We are now in a position to write the full likelihood for X_T and X_F given a cluster ensemble,

$$\mathcal{L}(X_T,X_F) \propto \prod_\alpha P(X_1^\alpha,X_2^\alpha|\mathbf{\Omega}) \tag{45}$$

or

$$\ln\mathcal{L} = -\frac{1}{2}\sum_\alpha \left(\mathbf{X}_\alpha - \langle\mathbf{X}\rangle\right)\mathbf{C}_\alpha^{-1}\left(\mathbf{X}_\alpha - \langle\mathbf{X}\rangle\right). \tag{46}$$

The mean $\langle X_1^\alpha\rangle$ and $\langle X_2^\alpha\rangle$ are still given by equations 30 and 31, while the covariance matrix $\mathrm{Cov}(X_1,X_2)$ is given as above.

In principle, we can use the above likelihood to estimate the model parameters $\mathbf{\Omega} = \{X_T,X_F,\sigma_T^2,\sigma_F^2,r_{TF}\}$. In practice, however, the estimates for the intrinsic scatters σ_T and σ_F will depend on the detailed knowledge of observational errors. In the limit of low S/N per cluster that is typical of cluster stacking, this is an undesirable assumption. That is, we do not wish to use this data to estimate the intrinsic scatter parameters σ_T, σ_F, and r_{TF}: we wish to focus exclusively on the means. Consequently, we *assume* values for these parameters *a priori* for the purposes of defining the estimators $\hat{X}_T$ and $\hat{X}_F$. Errors in the a priori variance estimates result then in sub-optimal but unbiased X_T and X_F estimators. Since the stacked data typically has high S/N, the minor hit in S/N due to the sub-optimal weighting because of our a priori choice of weights is an acceptable tradeoff. Consequently, from this point on, we assume that the full covariance matrix $\mathrm{Cov}(X_1,X_2)$ is known *for the purposes of defining an estimator $\hat{X}_T$ and $\hat{X}_F$.*

Assuming the covariance matrix $\mathbf{C}$ is known, we can now set $\partial\ln\mathcal{L}/\partial X_T = 0$ and $\partial\ln\mathcal{L}/\partial X_F = 0$. We find that the estimators for X_T and X_F are obtained by inverting the following matrix equation

$$\begin{pmatrix}\hat{X}_T\\ \hat{X}_F\end{pmatrix} = \begin{pmatrix} A & B\\ B & D\end{pmatrix}\begin{pmatrix}Y_T\\ Y_F\end{pmatrix} \tag{47}$$

where

$$Y_T = \sum_\alpha\sum_{ij}(C_\alpha)_{ij}^{-1}X_i^\alpha p_j^\alpha \tag{48}$$

$$Y_F = \sum_\alpha\sum_{ij}(C_\alpha)_{ij}^{-1}X_i^\alpha(1-p_j)^\alpha \tag{49}$$

$$A = \sum_\alpha\sum_{ij}(C_\alpha)_{ij}^{-1}p_i^\alpha p_j^\alpha \tag{50}$$

$$B = \sum_\alpha\sum_{ij}(C_\alpha)_{ij}^{-1}(1-p_i^\alpha)p_j^\alpha \tag{51}$$

$$D = \sum_\alpha\sum_{ij}(C_\alpha)_{ij}^{-1}(1-p_i^\alpha)(1-p_j^\alpha). \tag{52}$$

In the above expressions, the first sum is over all clusters α, while the second sum is over i and j ranging from 1 to 2, that is, the sum is over the two candidate centers for each cluster. At a qualitative level, the important thing to note is that Y_T and Y_F are inverse variance weighted sums of X_1 and X_2, and that these weighted averages are related by a linear equation to X_T and X_F.

Our final expressions for $\hat{X}_T$ and $\hat{X}_F$ are

$$\hat{X}_T = \frac{1}{AD-B^2}\left[DY_T - BY_F\right] \tag{53}$$

$$\hat{X}_F = \frac{1}{AD-B^2}\left[-BY_T + AY_F\right]. \tag{54}$$

Relating these expressions back to our sketch in section 3.1, we note that rather than inverting the relation between (X_1^α, X_2^α) and (X_T^α, X_F^α) on a cluster by cluster basis, and then averaging $\hat{X}_T^\alpha$ over all clusters, our stacking procedure takes a weighted sum of X_1 and X_2 for each cluster to directly produce a single $\hat{X}_T$ estimate for the full cluster ensemble. This results in an algorithm that is much more numerically stable, as the pre-factor $1/(p_1 - p_2)$ for $\hat{X}_T^\alpha$ never arises. Roughly speaking, our assumption that all the (X_T^α, X_F^α) values are drawn from a distribution $P(X_T^\alpha|X_F^\alpha|X_T, X_F)$ regularizes the inversion for the cases where $p_1 \approx p_2$.

3.3. *Limiting Solution and Expectation Values*

It is worth considering how our estimators above behave in the limit of known cluster centering, i.e. $p_1^\alpha = 1$ for all clusters α. We further assume the measurements errors for X_1^α and X_2^α are uncorrelated, so that C_α is diagonal. In this case, $A = \sum_\alpha \mathrm{Var}(X_1^\alpha)^{-1}$, $B = 0$, and $D = \sum_\alpha \mathrm{Var}(X_2^\alpha)$. Similarly, $Y_T = \sum_\alpha \mathrm{Var}(X_1^\alpha)^{-1} X_1^\alpha$ and $Y_F = \sum_\alpha \mathrm{Var}(X_2^\alpha) X_2^\alpha$. Putting it all together, our final expressions for $\hat{X}_T$ in this limit becomes

$$\hat{X}_T = \frac{Y_T}{A} = \frac{\sum w_\alpha X_1^\alpha}{\sum w_\alpha} \tag{55}$$

where

$$w_\alpha^{-1} = \mathrm{Var}(X_1^\alpha). \tag{56}$$

This is simply the standard inverse-variance weighted estimator that is traditionally used in stacking (e.g. **?**).

Further, we can demonstrate that our estimators $\hat{X}_T$ and $\hat{X}_F$ are unbiased. Consider first the expectation value of Y_T. Y_T is itself a linear combination of X_1^α and X_2^α,

$$Y_T = \sum_\alpha \sum_j M_{ij}^\alpha p_j^\alpha X_1^\alpha + \sum_\alpha \sum_{ij} M_{ij}^\alpha p_j^\alpha X_2^\alpha. \tag{57}$$

where we defined $\mathbf{M} = \mathbf{C}^{-1}$ for each cluster α. Upon taking expectation values, we use equations 30 and 31 to express the expectation values of X_1^α and X_2^α in terms of X_T and X_F. We find

$$\begin{aligned}\langle Y_T \rangle &= X_T \sum_\alpha \sum_j (M_{1j}^\alpha p_1^\alpha p_j^\alpha + M_{2j}^\alpha p_j^\alpha p_2^\alpha) \\ &\quad + X_F \sum_\alpha \sum_j (M_{1j}^\alpha p_2^\alpha p_j^\alpha + M_{2j}^\alpha p_j^\alpha p_1^\alpha) \qquad (58) \\ &= AX_T + BX_F. \qquad (59)\end{aligned}$$

For the last equality, we use the fact that $p_1^\alpha = (1 - p_2^\alpha)$. We can perform a similar calculation for Y_F, and we arrive at

$$\langle Y_T \rangle = AX_T + BX_F \tag{60}$$

$$\langle Y_F \rangle = BX_T + DX_F. \tag{61}$$

We can now insert these expectation values into the expectation value for $\hat{X}_T$ and $\hat{X}_F$. We find

$$\left\langle \hat{X}_T \right\rangle = X_T \tag{62}$$

$$\left\langle \hat{X}_F \right\rangle = X_F, \tag{63}$$

proving the estimators are unbiased. It is worth emphasizing that in this derivation we did not need to assume that $\mathbf{C}$ was the true covariance matrix of X_1^α and X_2^α: any covariance matrix $\mathbf{C}$ will result in unbiased estimators, supporting our earlier statement that a miss-estimate of the covariance matrix $\mathbf{C}$ will result in sub-optimal but unbiased estimators for X_T and X_F.

3.4. *Cluster Stacking with Probabilistic Binning*

Equation 72 is, in some sense, deceptively simple. That is, given an ensemble of galaxy clusters, equation 72 allows us to estimate X_T for the cluster ensemble. However, if one wishes to stack cluster in a richness bin, how is one to "bin" clusters if the richness is not uniquely defined because of probabilistic centering? Our naive expectation is that the expressions from equations 72 and 73 need to be further weighed by the probability that a cluster is in the richness bin of interest, and one then extends the sums over all galaxy clusters.

For instance, consider the quantity Y_T defined in the previous section,

$$Y_T = \sum_\alpha \sum_{ij} (C_\alpha)_{ij}^{-1} X_i^\alpha p_j^\alpha. \tag{64}$$

where the sum is meant to be over all clusters in a richness bin a. The natural generalization of this quantity in the presence of probabilistic binning is

$$Y_T^a = \sum_\alpha q_a^\alpha \sum_{ij} (C_\alpha)_{ij}^1 X_i^\alpha p_j^\alpha. \tag{65}$$

In this case, the sum now extends over *all* galaxy clusters, and the summand has been multiplied by q_a^α, the probability that cluster α is in bin a. In other words, each cluster contributes to each bin in proportion to its bin-membership probability. In the case of deterministic binning $q_a^\alpha = 1$ if the cluster is bin a, and $q_a^\alpha = 0$ otherwise, so the whole sum collapses to our original expression. Just as we can generalize Y_T in this way, so too can we generalize Y_F, A, B, C, and D, thereby arriving at the final $\hat{X}_T$ and $\hat{X}_F$ estimators.

These "natural generalizations" for the $\hat{X}_T$ and $\hat{X}_F$ estimators can be formally derived. Specifically, let X_T^a and X_F^a be the model parameters that characterize the means of X_T and X_F for each richness bin a. The probability that cluster α has richness (X_1^α, X_2^α) with cluster α in bin a — a condition characterized by the binning state variable $\vec{\psi}^\alpha = \{\psi_1^\alpha, \ldots, \psi_\mu^\alpha\}$ where μ is the number of richness bins — is

$$P_\alpha(X_1^\alpha, X_2^\alpha, \vec{\psi}^\alpha) = \sum_a P(X_1^\alpha, X_2^\alpha, \vec{\psi}^\alpha | \alpha \in a) P(\alpha \in a) \tag{66}$$

and

$$P(X_1^\alpha, X_2^\alpha, \vec{\psi}^\alpha | \alpha \in a) = P(X_1^\alpha, X_2^\alpha | \vec{\psi}^\alpha, \alpha \in a) P(\vec{\psi}^\alpha | \alpha \in a) \tag{67}$$

Now, if cluster α is in bin a, then $P(\vec{\psi}^\alpha | \alpha \in a) = 0$ unless $\psi_a^\alpha = 1$, in which case $P(\vec{\psi}^\alpha | \alpha \in a) = 1$. Consequently, $P(\vec{\psi}^\alpha | \alpha \in a) = \psi_a^\alpha$. This, together with the the fact that the binning probability $q_a^\alpha = P(\alpha \in a)$ implies that the full probability distribution $P_\alpha(X_1^\alpha, X_2^\alpha, \vec{\psi}^\alpha)$ can be written as

$$P_\alpha(X_1^\alpha, X_2^\alpha, \vec{\psi}^\alpha) = \sum_a \psi_a^\alpha q_a^\alpha P(X_1^\alpha, X_2^\alpha | X_T^a, X_F^a). \tag{68}$$

For convenience, we can rewrite this as

$$P_\alpha(X_1^\alpha, X_2^\alpha, \vec{\psi}^\alpha) = \exp\left\{ \sum_a \psi_a^\alpha p_a^\alpha \ln P(X_1^\alpha, X_2^\alpha | X_T^a, X_F^a) \right\}, \tag{69}$$

where the equality holds because ψ_a^α is either zero or one. The probability distribution for the full ensemble is obtained by multiplying over all galaxy clusters. The corresponding log-likelihood is

$$\mathcal{L}(X_T^\alpha, X_F^\alpha) = \sum_\alpha \sum_a \left[\psi_a^\alpha \ln(q_a^\alpha) + \psi_a^\alpha \ln P(X_1^\alpha, X_2^\alpha | X_T^a, X_F^a)\right]. \tag{70}$$

To obtain the maximum likelihood estimators, we adopt an expectation–maximization algorithm. Briefly, one assumes an arbitrary initial guess X_T^a and X_F^a. With this initial guess, one computes the expectation of $\mathcal{L}$ over the hidden data ψ_a^α, i.e. the binning membership of each galaxy cluster. Having computed the expectation value $\langle\mathcal{L}\rangle$ over the random variables ψ_a^α, one maximizes $\langle\mathcal{L}\rangle$ with respect to X_T^a and X_F^a to arrive at estimators $\hat{X}_T^a$ and $\hat{X}_F^b$. One can now re-evaluate the expectation value of $\mathcal{L}$ over the hidden data variables ψ_a^α, and then use $\langle\mathcal{L}\rangle$ to define a new set of estimators. The procedure can then be iterated until convergence is achieved. We are fortunate in the fact that, because the hidden variables ψ_a^α are in fact independent of X_T^α, convergence is achieved after only one step.

The expectation value of the log-likelihood over the hidden data variables is

$$\langle\mathcal{L}\rangle = \sum_\alpha \sum_a q_a^\alpha \ln P(X_1^\alpha, X_2^\alpha | X_T^a, X_F^a) \tag{71}$$

where we have dropped constant terms that do not depend on X_T^a or X_F^a. Using our Gaussian distribution for $P(X_1^\alpha, X_2^\alpha | X_T^a, X_F^a)$, we find that the maximum likelihood estimators for X_T^a and X_F^a are given by the naive generalizations detailed in the previous section. That is, we arrive at

$$\hat{X}_T^a = \frac{1}{AD - B^2}\left[DY_T - BY_F\right] \tag{72}$$

$$\hat{X}_F^a = \frac{1}{AD - B^2}\left[-BY_T + AY_F\right]. \tag{73}$$

$$Y_T = \sum_\alpha q_a^\alpha \sum_{ij} (C_\alpha)^{-1}_{ij} X_i^\alpha p_j^\alpha \tag{74}$$

$$Y_F = \sum_\alpha q_a^\alpha \sum_{ij} (C_\alpha)^{-1}_{ij} X_i^\alpha (1-p_j)^\alpha \tag{75}$$

$$A = \sum_\alpha q_a^\alpha \sum_{ij} (C_\alpha)^{-1}_{ij} p_i^\alpha p_j^\alpha \tag{76}$$

$$B = \sum_\alpha q_a^\alpha \sum_{ij} (C_\alpha)^{-1}_{ij} (1-p_i^\alpha) p_j^\alpha \tag{77}$$

$$D = \sum_\alpha q_a^\alpha \sum_{ij} (C_\alpha)^{-1}_{ij} (1-p_i^\alpha)(1-p_j^\alpha). \tag{78}$$

4. MONTE CARLO TESTS

We wish to test the above formalism in a controlled environment. Specifically, we use a simple theoretical model to generate artificial data, and use this artificial data to recover the underlying parameters of the model using the formalism laid out above. As will be apparent, our generative model is exceedingly simple, and is *not* meant to represent the complexities of real data. Quite the contrary, our tests in this section our only intended to test whether the formalism laid out above can be used to correctly recover the underlying parameters of the generative model.

Abundance is a modified Schecter, X will be WL mass, centering probabilities based on galaxy luminosities.

5. SAMPLE STUDY: CENTER-CORRECTED WEAK LENSING PROFILES OF redMaPPer CLUSTERS

Show plots and stuff.

6. AN EMPIRICAL TEST OF THE redMaPPer CENTERING PROBABILITIES WITH WEAK GRAVITATIONAL LENSING

The above formalism is exact: given the centering probabilities p_i^α, one may always recover the true $\Delta\Sigma$ profile about a cluster's true center, even if that cluster center is not known. However, if the probabilities p_i^α are miss-estimated, then the recovered weak lensing profile will be biased. Fortunately, one may define a null-test that must be satisfied when the estimated probabilities are correct. Namely, if the centering probabilities are correct, the profiles $\Delta\Sigma_1$ and $\Delta\Sigma_2$ are related to each other in a fully deterministic fashion.

Let $\mathbf{x}_c$ be the centering offset of a galaxy cluster. The surface density contrast profile $\Delta\Sigma(R|\mathbf{x}_c)$ about this offset cluster center can be related to the properly centered surface density profile $\Sigma(x)$ via

$$\Delta\Sigma(R|\mathbf{x}_c) = \left\langle\bar{\Sigma}(R|\mathbf{x}_c)\right\rangle - \bar{\Sigma}(R|\mathbf{x}_c) \tag{79}$$

where

$$\bar{\Sigma}(R|\mathbf{x}_c) = \frac{1}{2\pi R}\int d^2\mathbf{x}\ \Sigma(x)\delta(|\mathbf{x}-\mathbf{x}_c|-R) \tag{80}$$

$$\left\langle\bar{\Sigma}(R|\mathbf{x}_c)\right\rangle = \frac{1}{\pi R^2}\int_0^R (2\pi R' dR')\ \bar{\Sigma}(R'|\mathbf{x}_c). \tag{81}$$

Using the δ function, one may carry out the integrals over angle to arrive an integral relation between the profile $\Sigma(x)$ and $\Delta\Sigma(R|\mathbf{x}_c)$,

$$\Delta\Sigma(R|\mathbf{x}_c) = \int_{-\infty}^{\infty} d\ln x\ \Sigma(x)\left[\tilde{f}(x,x_c,R) - f(x,x_c,R)\right] \tag{82}$$

where

$$f = \frac{1}{\pi}\frac{x\left[x^2+x_c^2-2xx_c\alpha h(\alpha)\right]^{1/2}}{Rx_c(1-\alpha^2)^{1/2}} \tag{83}$$

$$\tilde{f} = \frac{2}{R^2}\int_0^R dR'R'\, f(x,x_c,R') \tag{84}$$

and

$$\alpha = \frac{R^2-x^2-x_c^2}{2xx_c}, \tag{85}$$

and $h(\alpha) = 1$ for $\alpha \in [-1,1]$, and $h(\alpha) = 0$ otherwise. As we will see later, we will need to evaluate these expressions when $x_c = 0$, in which case the functions f and $\tilde{f}$ reduce to

$$f(x,0,R) = R\delta(x-R) \tag{86}$$

$$\tilde{f}(x,0,R) = 2\left(\frac{x}{R}\right)^2\theta(R-x) \tag{87}$$

where $\theta(R-x) = 1$ if $R > x$ and $\theta(R-x) = 0$ otherwise.

The important thing about equation 82 is that it's an integral equation. If we discretize Σ and $\Delta\Sigma$ into logarithmic radial bins — hence the change to $d\ln x$ in the measure in equation 82 — the integral becomes a sum,

$$\Delta\Sigma_a(\mathbf{x}_c) = \sum_i (\Delta\ln x) M_{ai}(x,x_c)\Sigma_i \tag{88}$$

where the matrix $M_{ai}(x,x_c)$ is obtained by evaluating the kernel in equation 82 at the appropriate radius. In matrix form, this reduces to $\mathbf{\Delta\Sigma} = \mathbf{M}(x_c)\mathbf{\Sigma}$ where $\mathbf{M}$ is a matrix, and $\mathbf{\Sigma}$ and $\mathbf{\Delta\Sigma}$ are vectors. Inverting this matrix equation, we find

$$\mathbf{\Sigma} = \mathbf{M}^{-1}(x_c)\mathbf{\Delta\Sigma}(x_c) \tag{89}$$

where we have made it clear that the $\mathbf{\Delta\Sigma}$ is the profile observed at the displaced location x_c. Let now $\mathbf{\Delta\Sigma}_0$ be the density contrast profile obtained when the cluster is correctly centered. On has then

$$\mathbf{\Delta\Sigma}(x_c) = \mathbf{M}(x_c)\mathbf{M}^{-1}(x_c=0)\mathbf{\Delta\Sigma}_0 \tag{90}$$

Because the displacement vector $\mathbf{x}_c$ between the first and second most likely galaxies is always known, the above equation relates the density profiles $\Delta\Sigma_1$ and $\Delta\Sigma_2$ estimated in the previous section. That is one has

$$\mathbf{\Delta\Sigma}_2 = \mathbf{M}(x_{12})\mathbf{M}^{-1}(x_c=0)\mathbf{\Delta\Sigma}_1 \tag{91}$$

where x_{12} is the separation between galaxies 1 and 2.

Because the profiles $\Delta\Sigma_1$ and $\Delta\Sigma_2$ are themselves linear combinations of the profiles $\Delta\tilde{\Sigma}_1$ and $\Delta\tilde{\Sigma}_2$, equation 91 is a zero-free parameter relation that relates two direct observables. Consequently, if equation 91 is not satisfied in the data, one has evidence that the centering probabilities being estimated are incorrect. There are, however, some caveats that must be kept in mind. First, the true relation is in fact an integral relation, so some error must necessarily be introduced by the discretization process. One can, however, characterize this error by adopting realizations of model profiles known to satisfy the integral relation, and then evaluating the error associated with the discretization process. Second, the integral relation necessarily extends over regions where there may not be data. The simplest solution here is to remove one more edge-bin, and compare the integral relations with and without the last bin. Evidently, one should not expect to hold the integral relation to hold for those bins in which the addition or subtraction of an edge bin has a significant impact on the result. Besides these two caveats, however, this should be a powerful test of the lensing probabilities.