Mass bias basics

In galaxy cluster analysis, the mass bias is defined as the difference between a cluster's true mass $M_{\rm true}$ and its observed (i.e. measured) mass $M_{\rm obs}$, written as $\Delta M = M_{\rm obs} - M_{\rm true}$. Sometimes this is written as a ratio: $\mathcal{C} = M_{\rm true}/M_{\rm obs} = 1 - \Delta M/M_{\rm obs}$.

When measuring galaxy cluster masses, a mass bias is caused by systematics, sometimes in combination with each other. Specifically, the mass bias is caused by systematics that are unaccounted for which affect observations of the cluster. Let's take stacked weak gravitational lensing as an example. For stacked lensing, the mass M we are measuring is the average mass of the stack.

Weak lensing

The weak lensing profile for a stack with mean mass M is defined as

$$\Delta\Sigma(R|M) = \bar{\Sigma}(< R|M) - \Sigma(R|M)\,, \tag{1}$$

where $\Delta\Sigma$ is the differential surface mass density at radius R, $\bar{\Sigma}(< R)$ is the average surface mass density below R, and $\Sigma(R)$ is the surface mass density at R. The surface mass density is given as the line-of-sight integral of the overdensity of the halo a distance R from its center:

$$\Sigma(R|M) = \int_{-\infty}^{\infty} {\rm d}r_z\ \Delta\rho\left(\sqrt{R^2+r_z^2}|M\right) = 2\bar{\rho}\int_0^{\infty} {\rm d}r_z\ \xi_{\rm hm}\left(\sqrt{R^2+r_z^2}|M\right)\,, \tag{2}$$

where $\Delta\rho(r|M) = \rho(r|M) - \bar{\rho} = \bar{\rho}\xi_{\rm hm}$ is the difference between the halo density and the background density ($\rho_{\rm crit}$ or ρ_m, depending on halo definition) and $\xi_{\rm hm}$ is the halo-matter correlation function. By measuring the shape of galaxies behind galaxy clusters, we can get an estimate of the lensing profile $\widetilde{\Delta\Sigma}$. In other words, $\widetilde{\Delta\Sigma}$ would be our "data", and the parameters that go into $\Delta\Sigma$ are our "model".

In measuring the mass of a cluster stack, we want to measure the posterior $P(\vec{\theta}|\widetilde{\Delta\Sigma})$, or the distribution of our model parameters $\vec{\theta}$ given our data. Using Bayes' theorem we can write this in terms of the likelihood of our data and a prior

$$P(\vec{\theta}|\widetilde{\Delta\Sigma}) = \frac{P(\widetilde{\Delta\Sigma}|\vec{\theta})P(\vec{\theta})}{P(\widetilde{\Delta\Sigma})}\,, \tag{3}$$

where we can neglect the prior on the data, since it is just a normalization, and $M \in \vec{\theta}$. A mass bias occurs when computing $\Delta\Sigma$ from the model $\vec{\theta}$ results in a biased estimate of the lensing signal. In other words, $\Delta\Sigma(M+\Delta M) \approx \widetilde{\Delta\Sigma}(M)$. Therefore, when we evaluate the posterior of the parameters, the marginalized posterior on M yields an expectation value of $M + \Delta M$. If not corrected for, this mass bias propagates into cosmological parameters (see, e.g., the Planck cluster cosmology papers).

One thing to note is that the mass bias depends on the likelihood of the data $P(\widetilde{\Delta\Sigma}|\vec{\theta})$. Oftentimes, the log of this likelihood looks like

$$\ln P(\widetilde{\Delta\Sigma}|\vec{\theta}) = -\frac{1}{2}\left[\ln|C| + \sum_{i,j}(\widetilde{\Delta\Sigma} - \Delta\Sigma(\vec{\theta}))_i C^{-1}_{i,j}(\widetilde{\Delta\Sigma} - \Delta\Sigma(\vec{\theta}))_j\right] \tag{4}$$

where C is the covariance matrix, and $\Delta\Sigma(\vec{\theta})$ is our model for the lensing profile computed from our parameters. The indices i, j run over radial bins R. Therefore, not only is our mass bias going to depend on $\Delta\Sigma(\vec{\theta})$, but also on our covariance matrix C.

Measuring mass bias

Using simulations, we can estimate the mass bias and correct for it. An N-body simulation evolves the positions of dark matter particles from high redshift to present day. Those particles interact gravitationally, and can collapse to form bound dark matter halos, which are known to host galaxy clusters. So, in a simulation, we have two sets of data, positions of dark matter particles $\vec{r}_{\rm dm}$ and halo catalogs that consist of positions and masses of the halos $\{\vec{r}_{\rm h}, M\}_i, i \in [1, ..., N_{\rm h}]$.

Galaxy clusters are identified not by their mass, but by some observable quantity, such as richness λ. Assuming we have a decent model for $P(\lambda|M)$, we can draw from this distribution and assign each halo a richness λ_i. Now, by grouping halos in bins of observable, we can make simulated stacks that have the same properties as the real data.

With our dark matter particle positions and halo catalog, one thing we can measure is the halo-matter correlation function $\xi_{\rm hm}$. This is found by pair counting, and can be computed using the following estimator:

$$\xi(r) = \frac{D_h D_{dm}(r)}{RR(r)} - 1\,. \tag{5}$$

In this equation, $D_h D_{dm}(r)$ is the number of halo-particle pairs at a distance $r = ||\vec{r}_{\rm h} - \vec{r}_{\rm dm}||$. $RR(r)$ is the number of pairs of random points separated by r. This equation assumes that the number densities have all been scaled correctly, but these are details we don't need to worry about. Pair counting is usually done by a specialized code, such as Corrfunc or TreeCorr. The details of the implementations of those codes is beyond the scope of this summary, and we can assume that the $\xi_{\rm hm}$ we get out is correct.

At this point, we have halo-matter correlation functions for stacks of halos binned by some quantity in the simulation. We can convert this into a simulated $\hat{\Delta\Sigma}$ profile following Equation 1 and Equation 2. That is, we integrate the correlation function along the line-of-sight to compute $\Sigma(R)$, from which we can compute $\hat{\Delta\Sigma}$. There are some details about how to extrapolate $\xi_{\rm hm}(r)$ to very small scales, but this is too minor for this discussion. For reference, in this package, the halo profile snaps to an NFW profile at those small scales, even though this profile isn't totally correct.

Finally, we plug our simulated profile $\hat{\Delta\Sigma}$ into Equation 4 and measure the halo mass. Since we know the masses of the halos in the simulations, we know $M_{\rm true}$. The mass we recover is $M_{\rm obs}$, meaning we have an estimate of the mass bias $\mathcal{C}$. More specifically, we have a full posterior for $\mathcal{C}$, which includes a mean and variance.

Systematics

I mentioned earlier that systematics are the cause of the mass bias. Even though we can compute a mass bias and correct for it, we should always try to minimize the bias by incorporating systematics into our simulated profile. Here is a list of systematics that can be applied in this package and how they affect the likelihood:

- Multiplicative bias of the form $A \times \Delta\Sigma$. A multiplicative bias A is almost completely degenerate with mass. The degeneracy is only slightly broken since the profile shape has a weak mass dependence. This dependence is not likely to be detected in DES or most of LSST, meaning for now a multiplicative bias is completely degenerate with mass. This kind of a bias can occur from shape and photo-z biases.

- Boost factors $\mathcal{B}(R)\Delta\Sigma$. This is also known as contamination of the source galaxy population, from galaxies either in the foreground or from within the cluster. This has the effect of re-weighting the terms in Equation 4. In this package, the functional form and free parameters of the boost factor profile can be specified or left out entirely.

- Miscentering $(1 - f_{\rm mis})\Delta\Sigma + f_{\rm mis}\Delta\Sigma_{\rm mis}$, where $f_{\rm mis}$ is the fraction of clusters in a stack that are miscentered. Miscentering is caused by the misidentification of cluster centers, which suppresses the lensing profile on small scales. This is a linear effect on the lensing profile, meaning it also has the effect of re-weighting terms in the likelihood. As was the case for boost factors, the miscentering parameters and form of the miscentering distribution can be specified or left out.

- Model bias. This refers to the fact that the model we use in Equation 4 $\Delta\Sigma(\vec{\theta})$ is not a perfect model. For instance, halo profiles are not NFW. The halo bias $b(M)$ at large scales might not be accurately modeled. Etc. This systematic is always present in this package, and can only be mitigated by finding better lensing models. This can be accomplished through simulation based models (emulators) and/or modeling additional effects on the lensing profile (triaxiality, projection effects, subhalos, intrinsic alignments, etc.).

- Covariance matrix. A poorly estimated covariance matrix can cause a mass bias. For instance, the measured lensing profile $\widetilde{\Delta\Sigma}$ will experience some scatter. Let's suppose the lensing profile has two radial bins that are uncorrelated, one of which was scattered high and the other was not scattered much at all. If the covariance weights the scattered bin, we will over estimate the mass of the stack. This is somewhat of a contrived example, but the reasoning applies to regions of the lensing profile that are better/worse modeled. Inner regions, for instance, are more strongly affected by baryons, while large scales are not. If we weigh them the same through the covariance matrix, we will obtain a mass bias. For this reason, any mass bias we calculate is completely dependent on what covariance matrix we choose. This means the mass bias is analysis specific, and a bias computed by one group in one experiment cannot be arbitrarily applied to another, unless their covariances are estimated in the same way.

Modeling the mass bias

Once a mass bias $\mathcal{C}$ is computed for each simulated stack, we can model the mass bias. This is a necessary step, because we can have more or less data in our simulation compared to real data, and we want to leverage the simulation as much as possible in order to apply it to the mass measurements on real data. The alternative would be to apply the calibration for the ith simulated stack to the mass measurement of the ith real stack, so $M_i = \mathcal{C}_i M_{\rm obs,i}$.

A sensible model for the mass bias is a power law in observable and redshift, as well as some intrinsic scatter.

$$\langle \mathcal{C} | \lambda, z \rangle = \mathcal{C}_0 \left(\frac{\lambda}{\lambda_0} \right)^{\alpha} \left(\frac{1+z}{1+z_0} \right)^{\beta} \tag{6}$$

$$\mathrm{Var}(\mathcal{C} | \lambda, z) = \sigma_0^2 \tag{7}$$

Here, the model parameters are the amplitude of the mass bias $\mathcal{C}_0$, the powers of the observable and redshift scaling α and β, and the intrinsic scatter σ_0.
More to come...