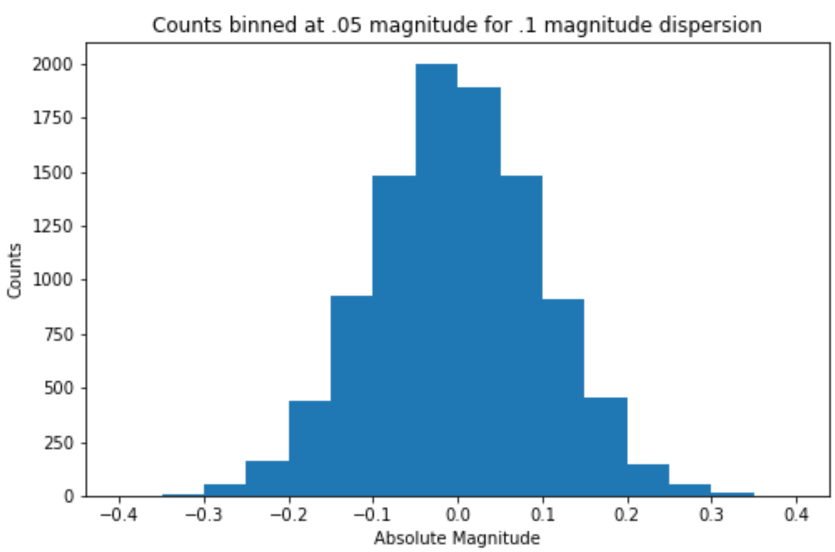
Nikhil Patten
1 March, 2022
Dr. Pierce
ASTR5465

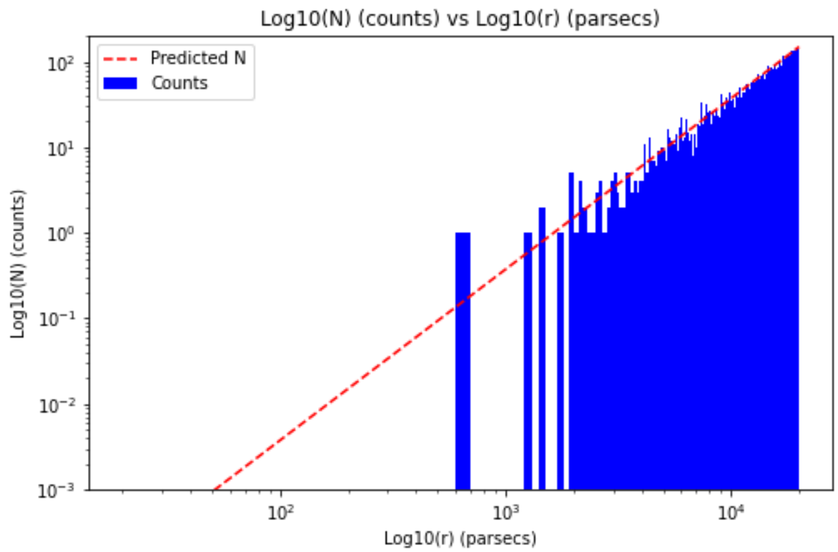
Sample selection can introduce bias into your sample. One such selection bias is called Malmquist bias where one has a population distributed in both absolute magnitude (luminosity) and in distance and then selects a subsample by apparent magnitude. The resulting sample of stars or galaxies will be biased as a function of distance. Let's simulate the effects of Malmquist bias in Pyton (or whatever you prefer). This is a long assignment so be sure to start early.

1. **Start by using a Gaussian random number generator to generate a set of stars (∼5-10,000) with a particular average absolute magnitude (i.e., <M> = 0) and with a set of intrinsic dispersions (sigma$_M$) of 0.1, 0.3 0.5, 1.0 magnitude. Assume a uniform special density distribution in 3-d and assign a true distance (and distance modulus, m-M) to each star. Hint: you can simply populate discrete shells (r, r+dr) and assign your stars as dN = 4πr^{2}dr, where r is the midpoint distance of the bin.**

2. **Make a histogram showing the number distribution in M and a second graph showing log(N) vs. log(r), i.e., distance. Do this for just one choice of sigma$_M$. Show that the second graph is consistent with a uniform special distribution by over plotting the above function.**

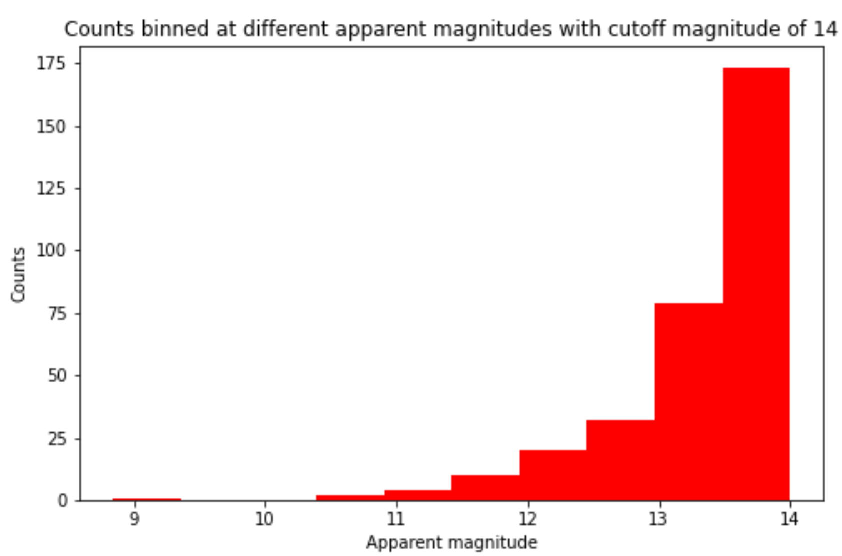


Above is the histogram of the number of stars generated from the Gaussian distribution binned by magnitude with an average absolute magnitude of 0.

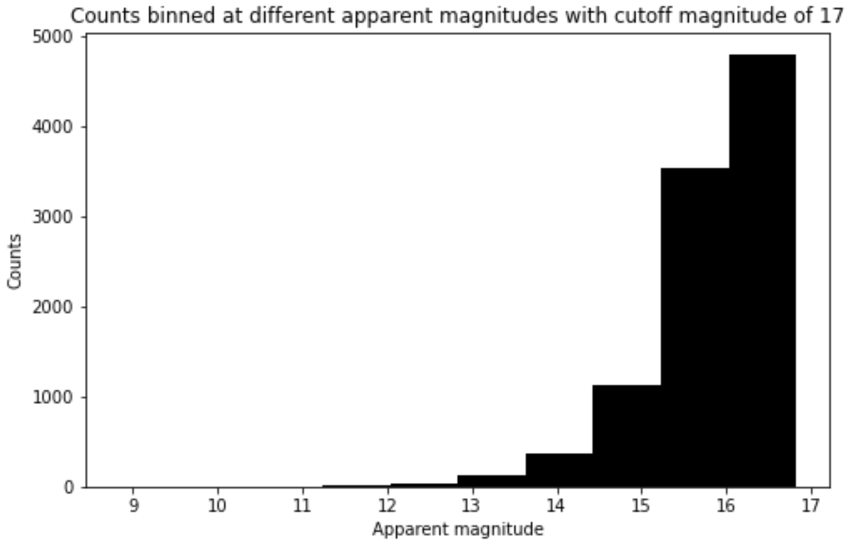


Plotted is the log of the star counts as a function of the log of the radius overplotted with the model of a spherical distribution of stars. As we can see, the generated stars fit the spherical distribution model.

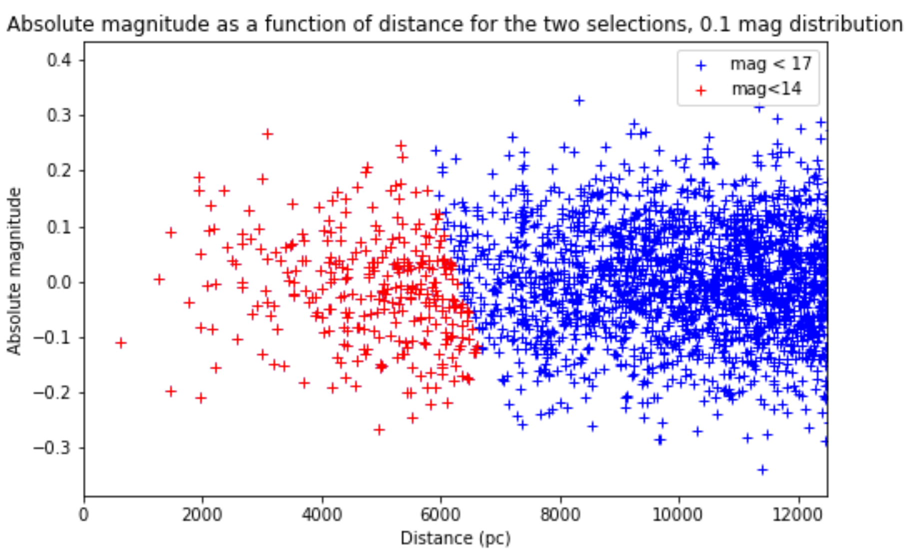
3. **Next compute the estimated distance moduli for each star assuming the known <M>, but not their actual individual values, for the different values of sigma$_M$. Apply a couple of reasonable cuts in apparent magnitude using their actual individual <M> and actual distances to limit the sample. Here you are limiting your sample in apparent magnitude since telescope time is scarce. Make histograms of two examples of your apparent magnitude cuts (one with a brighter cut and one with a fainter cut) and make a graph of M (actual) vs. Distance (actual) to show the effects of your sample selection criteria in apparent magnitude. Repeat this for a different value of sigma$_M$. Discuss what you see in the context of a distribution in M and distance being cut in apparent magnitude (m).**



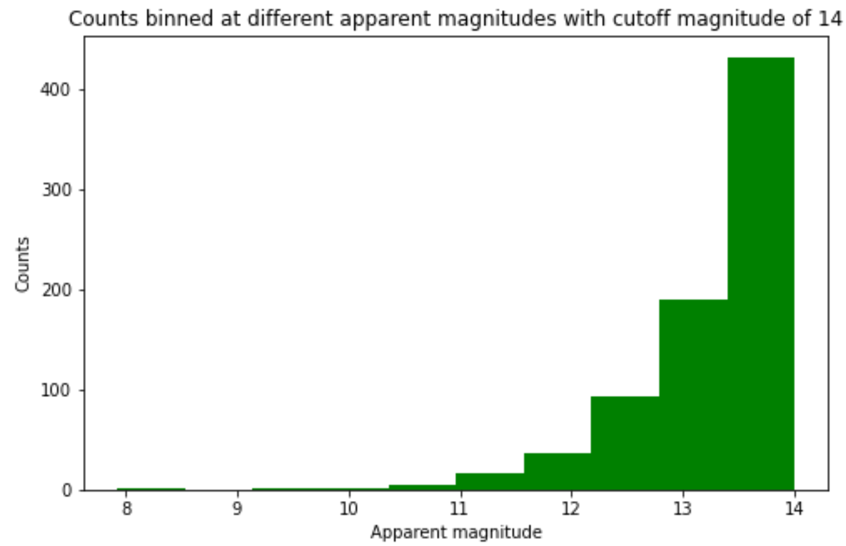
Above is a histogram of the star counts binned at different magnitudes but limited to apparent magnitude of 14. As we would expect, there are very few bright stars and the counts increase as we go up in apparent magnitude.



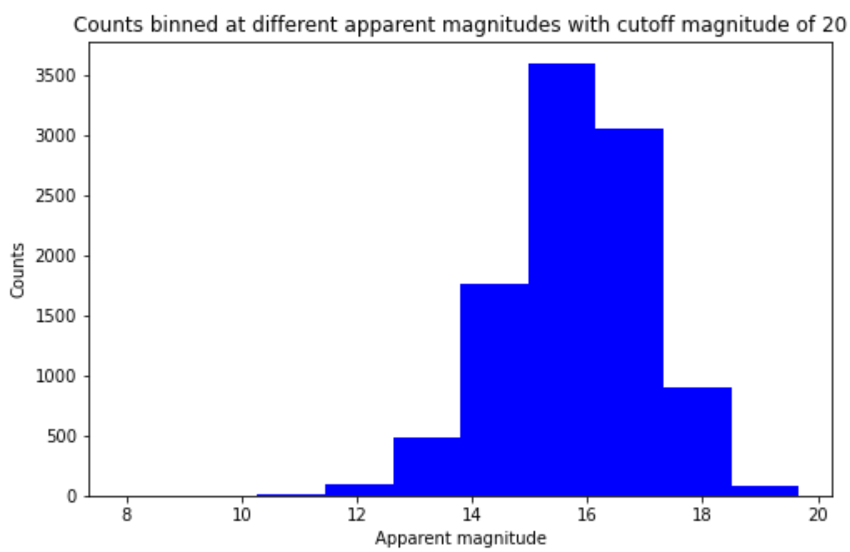
Here is a histogram of the star counts at different magnitudes with no limit on apparent magnitude. Comparing this histogram to the one above reveals that by limiting the cut-off magnitude to 14, we are excluding the vast majority of stars in the sample.



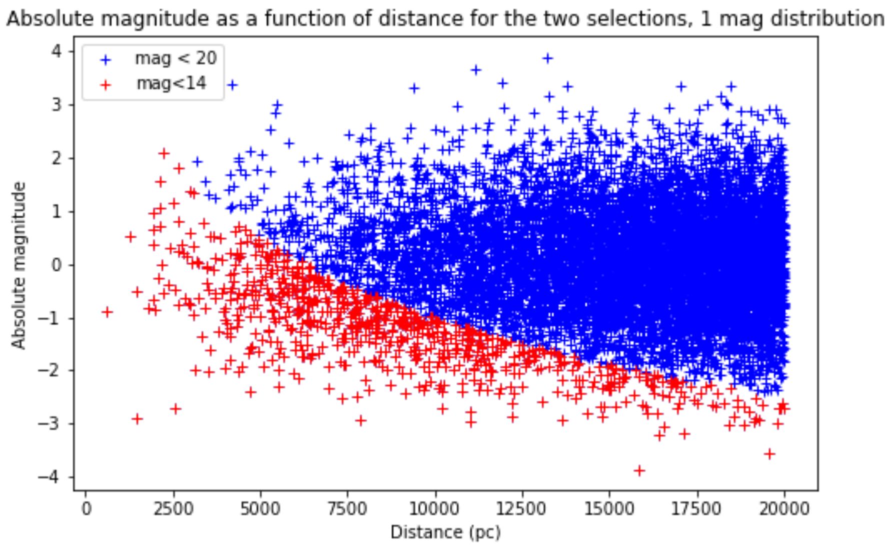
When M is limited to cut-off apparent magnitude of 14, we see a slight selection effect. For stars that are further away, we are preferentially selecting stars that are intrinsically brighter. We can see this effect on the plot of absolute magnitude versus distance as the stars with apparent magnitude less than 14 curves slightly at larger distances selecting only the brighter stars.



Repeating the above process for the sample of stars with larger variations in Absolute magnitude. We can see the same effect of the counts of stars increases with apparent magnitude.

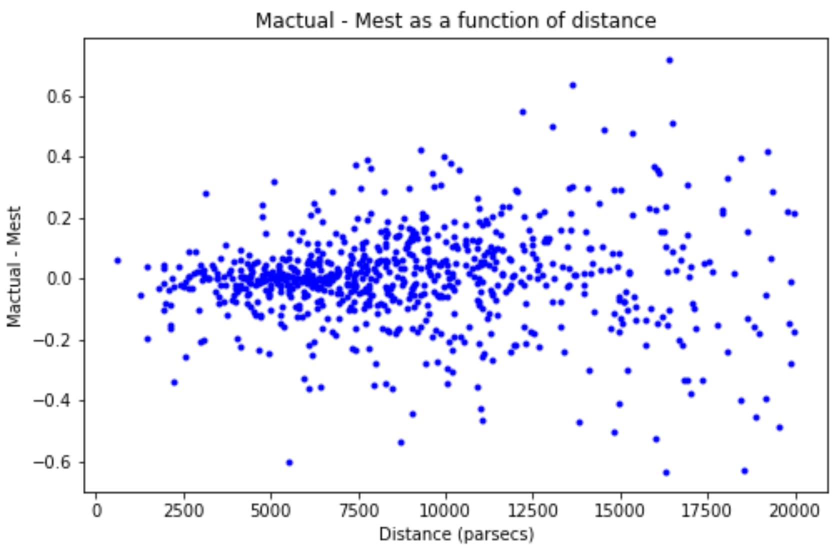


Making a histogram of the entire sample of stars not limited by apparent magnitude reveals the same trend that was noticed before, but much more dramatic. By limiting our selection to apparent magnitude of 14, we are excluding the vast majority of stars in the sample.

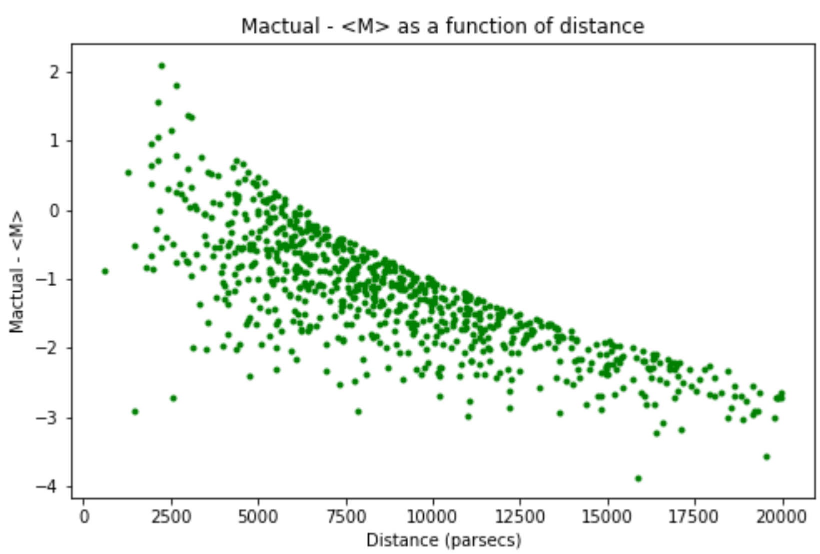


With the same cut in apparent magnitude but a larger distribution of M, we see that this selection effect has become more dramatic. At larger distances, only the stars with negative absolute magnitudes have been selected. The effects of Malmquist bias has been exacerbated with the larger variations in absolute magnitude.

4. **Now assume that you can measure a second parameter (P_{magic}) that can estimate the deviation of each star's absolute magnitude (M) from the average. For example, adopt $M_{est} = a \times P_{magic} + b$ with a precision of about 0.1 in sigma$_M$. Hint: here you can invert the relationship to compute P_{magic} and then add random noise to either P_{magic} or just add it M_{est}. Put another way, when you "measure" P_{magic} at the telescope it predicts/estimates M to ± 10%. Now make plots of M_{actual} - M_{est} vs. distance and M_{actual} - <M> vs. distance in order to show the effects of the Malmquist bias.**

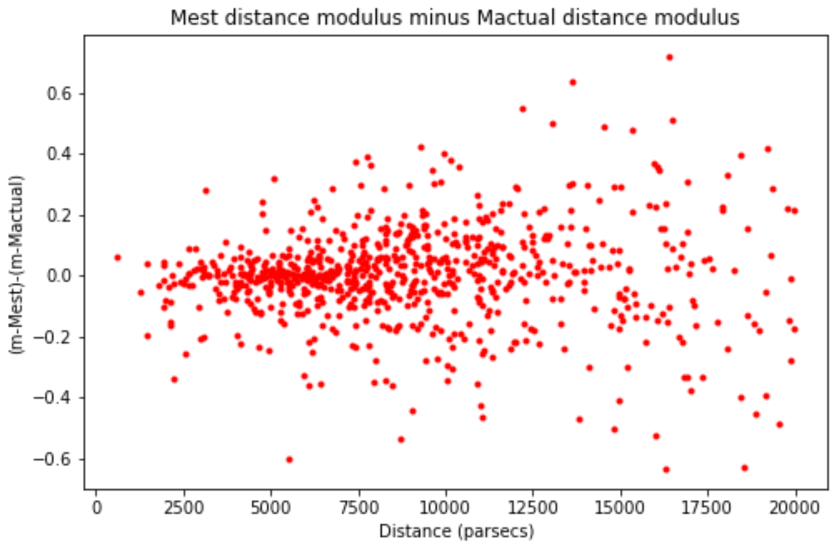


Plotting the difference between the actual absolute magnitude of the selected stars between the absolute magnitude calculated from P_{magic} reveals a Gaussian distribution of magnitudes centered on 0. The noise in these values is independent of distance and is the same for the entire sample.

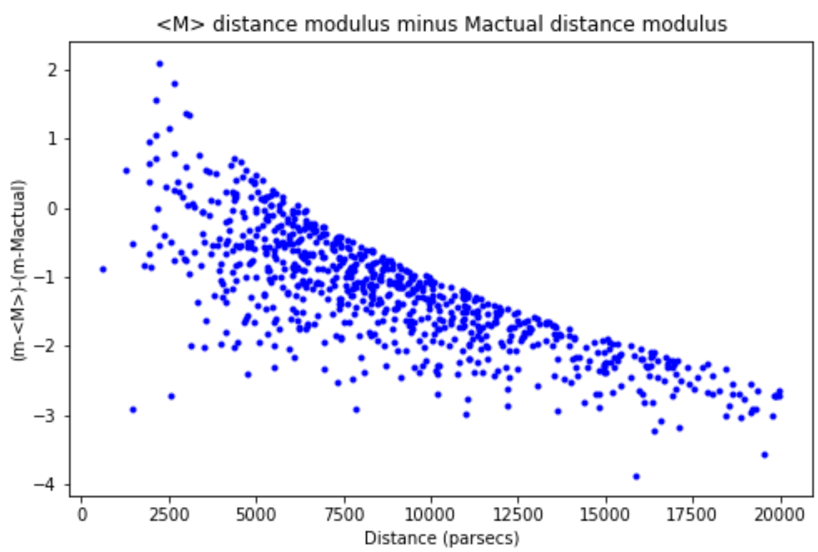


Plotting the difference between the actual absolute magnitudes of the stars selected by apparent magnitude and the average absolute magnitude of the sample reveals that for larger distances, the sample is biased towards intrinsically brighter stars.

5. **Note this will also effect any estimated distance modulus for your sample members since they all have measurable apparent magnitudes given their actual distance and their actual M. Do this by either assuming M_{est} = <M> or using parameter P_{magic} to compute M_{est}. Make two plots of $(m\text{-}M)_{est}$ - $(m\text{-}M)_{actual}$ vs.distance (actual) for your two methods of estimating M.**



Here we can see the estimated distance modulus using P_{magic} is approximately the same as the actual distance modulus for all distances. There is a Gaussian distribution of noise about 0 which is to be expected as P_{magic} gives the actual absolute magnitude to within 10% error.



The distance modulus estimated from the average Absolute magnitude is approximately 0 for very close stars, but quickly becomes more and more inaccurate for larger distances because of the selection effects of Malmquist bias. At larger distances, the distance modulus becomes more inaccurate and a worse predictor of the actual distance modulus.

6. **Summarize your procedure and your results and compare what you find in part-4 with what you found in part-3. Discuss why they are different. Comment on your results on the estimated distance modulus from part 5.**

I populated the sky with object with a Gaussian distribution of magnitudes centered on 0 for varying sigmas about the mean. Plotting a histogram of the absolute magnitudes of each of these stars reveals a Gaussian distribution, not so surprisingly. When selecting by apparent magnitude, the effects of Malmquist bias becomes apparent. At larger distances, brighter stars are prefentially chosen out of the random distribution of stars. This effect is more noticeable with a larger variation in the stars Absolute magnitude.

When using the parameter P_{magic} in part 4 to calculate absolute magnitude, we can effectively eliminate the effects of Malmquist bias found in part 3. Even when using a biased sample of intrinsically brighter stars, we found that in part four, we can accurately predict each stars' Absolute magnitude using P_{magic}. Plotting the star's acutal Absolute magnitude minus the Absolute magnitude calculated from P_{magic} versus distance reveals a Gaussian distribution about 0 for all distances indicating the P_{magic} is an accurate predictor of Absolute magnitude.

We can see a similar observation in part 5. The distance modulus calculated from P_{magic} minus the acutal distance modulus plotted over distance reveals a Gaussian distribution about 0 for all distances. This indicates that P_{magic} is an excellent method of calculating the distance modulus. When making the same plot using the average absolute for all distances reveals that at large distances, this method is insufficient for calculating the actual distance modulus and severely underestimates the actual value. This is a result of Malmquist bias.