Project 4: Distinguishing AGN Mass Populations Based on Redshift

Craig Brooks

Due: May 8 2023

Introduction

Active galactic nuclei (AGN) are highly active regions located at the centers of galaxies, characterized by the presence of a supermassive black hole accreting large amounts of matter. They exhibit a range of phenomena across different wavelengths, including powerful jets, intense electromagnetic radiation, and high-energy particles. These objects are believed to be fueled by the gravitational energy released as matter falls into the central black hole. The study of AGN plays a crucial role in understanding galaxy evolution processes and the interplay between black holes and their host galaxies. [3] [2]

AGNs are also some of the highest redshift objects in the universe. Since they form the core of galaxies, observing them at various redshifts may give us insight into how a galaxy's mass may evolve over time.

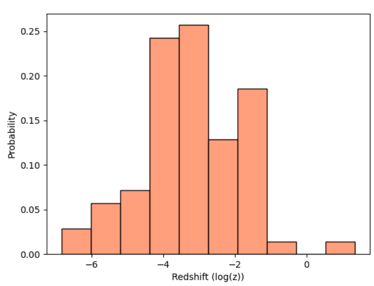
In this project, we would like to do is obtain the log-likelihood that the AGN with high masses is are located relatively late in cosmic time (low redshift and nearer) or in earlier at high redshift (high redshift and farther)

Exploratory Data Analysis

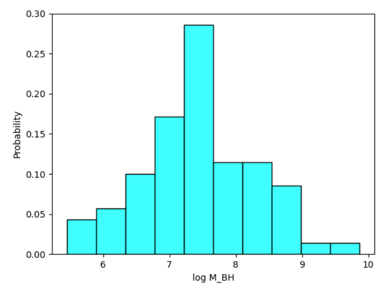
We use a dataset from *AGN Black Hole Mass Database* based on work by Katz and Bentz [1]. This database contains the masses, positions, and redshifts of approximately 90 AGNs whose masses are obtained using reverberation mapping. Reverberation mapping is a technique to estimate masses based on the velocity of the matter around the central black hole and a parameter f. It follows the relation

$$GM_{BH} = fR_{BLR}(\Delta\ V)^2$$

where M_{BH} is the mass of the central black hole, R_{BLR} is the radius of the broad line region, $\Delta\ V$ is the RMS velocity of the gas near the broad line emission region of the black hole. The analysis in this paper can be performed for different values of the f parameter (data files included)



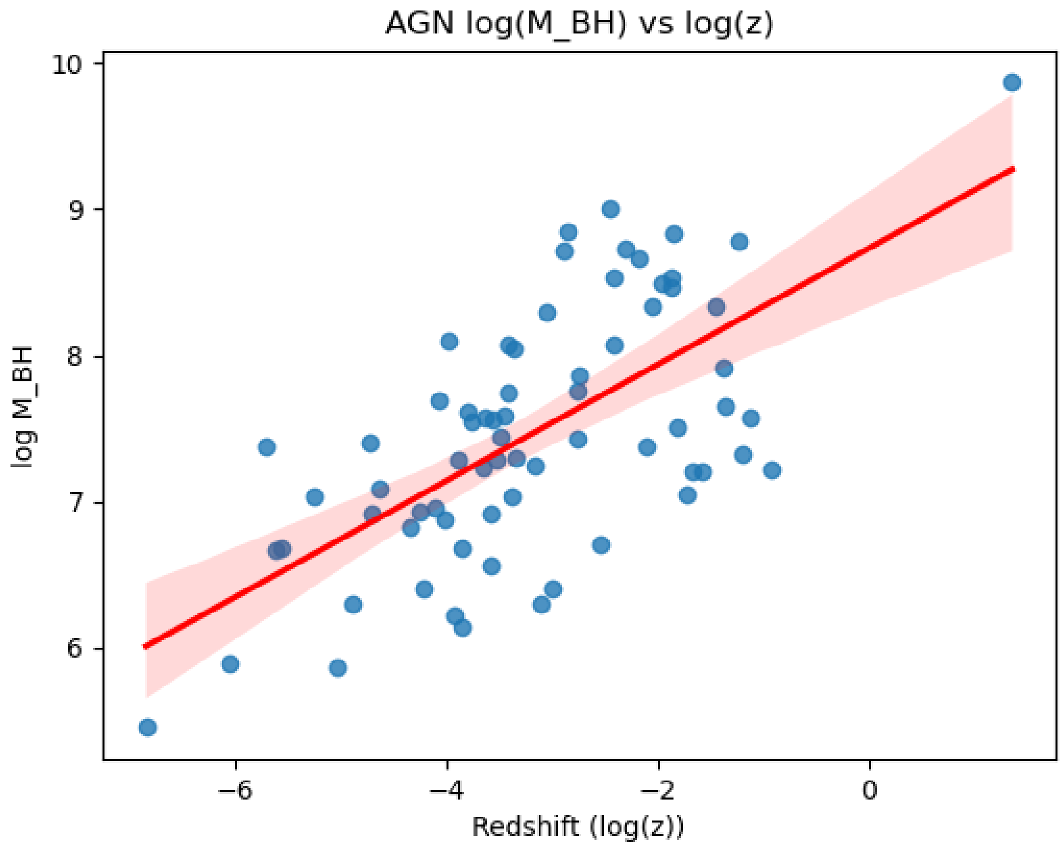
(a) Log Z



(b) Log of black hole mass

Figure 1: Log Z and Log mass histograms

Figure 1 shows histograms of each AGN by the log of redshift z and log of mass. We use log-log plots since there is a high dynamic range of the data. Prima facie, both seem to follow a normal distribution. Furthermore, Figure 2 plots the log mass vs log redshift, and fitted to a regression line. It does appear that since the trend is approximately linear, we will assume the normal distribution in our modelling.



We can build a likelihood function by declaring x_i as the log-mass of the AGN, P_A is the probability that $log\ z > -4$, and $P_B = 1 - P(A)$ The log-likelihood ration therefore becomes

$$\sum_i \log\left(\frac{P(X_i|P_A)}{P(X_i|P_B)}\right)$$

Our probability function in both cases will follow the normal distribution

$$X \sim \mathcal{N}(\mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

Let us start by computing $P(A) = P(\langle \log\ (Z)\rangle > 4)$ by first computing $\langle \log\ (Z)\rangle$ and σ_Z, then compute the z-score. The mean of $\log(Z)$ is -3.19, and the standard deviation σ_Z is 1.40. Therefore

$$z(Z = -4) = \frac{-4 - (-3.19)}{1.40} = -.5786$$

this corresponds to $P(Z > -4) = .7192$. Therefore, P(A) = .7192 and P(B) = 1-P(A) = .2808. Therefore, based on redshift alone, the probability of having a $\log(Z) > -4$ is 71.92%

If we examine the log of the mass of the AGN black holes, let us examine low mass black holes at high redshift. since we have a dynamic range of roughly $6 \leq \log(Z) \leq 10$, in other words, black holes of $10^6 M_\odot - 10^{10} M_\odot$ Perhaps we would like to determine the likelihood of finding a black hole of $10^6 M_\odot$ in the *late* universe, that is, at low redshift. As our sample is approximately normally distributed, we will treat the masses as they were in a normal distribution as we did the redshifts.

Let's define μ_{bh} as $\langle \log M_{BH} \rangle$, $\lambda_z = \langle \log(Z) \rangle$ and σ_{bh} is the standard deviation of $\log M_{BH}$. In our dataset, before bootstrapping, μ_{bh} is 7.46, and the standard deviation $\sigma_{bh} = .85$. Therefore, $z(\log M_{BH} = 6)$ is

$$z(Z = 6) = \frac{6 - 7.46}{.85} = -1.7182$$

This corresponds to $P(Z > 6) = .9571$, or, 95.71% of the black holes in this sample are larger than $10^6 M_\odot$ based on the mass alone, we would expect the overwhelming majority of AGN to be larger than 1 million solar masses.

Now we would like to determine the likelihood of a black hole having a mass of $10^6 M_\odot$ given $\log(Z) \geq -4$. We can represent this as.

$$P(\mu_{bh} \leq 6 | \lambda_z \geq -4) = \frac{P(\lambda_z \geq -4 | \mu_{bh} \leq 6) \cdot P(\mu_{bh} \leq 6)}{P(\lambda_z \geq -4)}$$

Two of these terms we have calculated. What is left is to determine $P(\lambda_z \geq -4 | \mu_{bh} \leq 6)$.

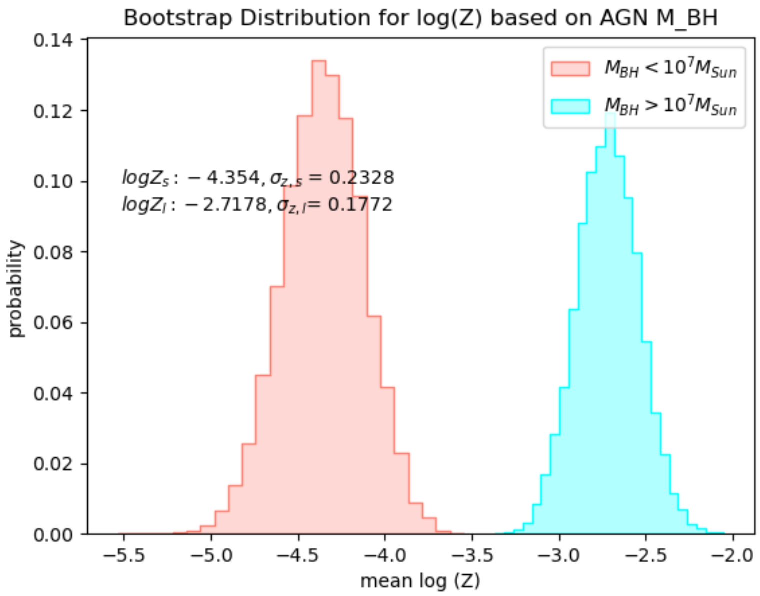
Hypothesis

For our Hypothesis, we have the following

$$H_0 : \lambda_z = -2.7 \sim \mu_{bh} \geq 7 \quad , \quad H_1 : \langle \lambda_z \rangle = -4 \sim \mu_{bh} < 7$$

Experiment

Because we are separating the AGN population based on a single mass cut selection ($\log M_B H = 7$), we can use a '2-urn model' to group them, one group for $\log(M_B H) < 7$ and $\log(M_B H) >= 7$. Since our dataset consists of small number of samples, we will use bootstrap sampling to get more meaningful test statistics such as $\langle \log(Z) \rangle$ in each population. Essentially, bootstrapping is the simulation of a larger population from a smaller dataset.



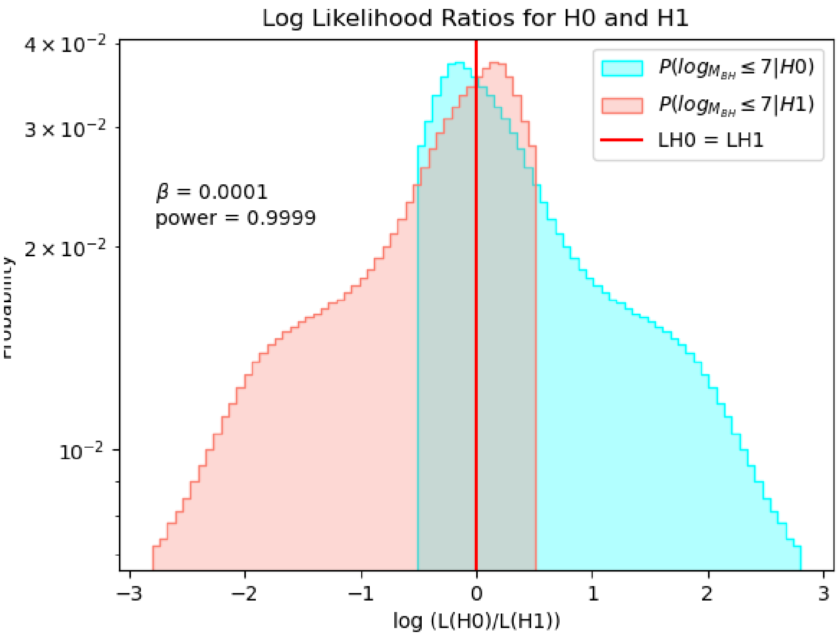
(a) $\langle log(Z) \rangle$

Figure 2: Histogram for bootstrap for $\langle \log(Z) \rangle$

Figure 2 shows the distribution of $\log(Z)$ and the standard deviation from the bootstrapped samples that were drawn from the masses alone. In the right panel of Figure 2, we can clearly see that the populations are quite distinct from each other based on the mass criterion we set for the experiment. The maximum of each histogram in Figure 2 is what we will take as the *true* $\langle log(Z) \rangle$ and σ_z. Therefore, The mean redshift for $\log(M_{BH}) < 7$ is $z = -4.35$ with $\sigma_{z,s} = .2351$ and for $\log(M_{BH}) \geq 7$ is $z = -2.72$ with $\sigma_{z,s} = .1748$

After the AGN's are sorted into mass categories, we will draw produce redshift samples and calculate the likelihood of the AGN populations having log-mass < 7 given the average redshift in the urn. We will also do the same for the complement of the dataset to compute the log-likelihood ratio.

Results



(a)

Figure 3: Log-likelohood ratios for low and high redshift

Figure 3 shows the log likelihood ratios for H_0 and H_1 for the same test statistic $\mu = \langle \log(Z) \rangle$. We found that μ had the higher likelihood ratio under H_0 than for H_1. This effectively means that at higher redshifts, you are more likely to find higher mass black holes than at lower redshifts. The average $\log(M_{BH}) = 7.86\ M_\odot$ in the high redshift sample and $\log(M_{BH}) = 6.48\ M_\odot$ for the low redshift black holes.

Therefore, we conclude that we have evidence that at higher redshifts, the AGN black hole masses are larger.

Confidence interval and power of the test

The 95 % confidence interval for the redshifts for $\log(M_BH) < 7 = [-4.8472, -3.9327]$ and for $\log(M_BH) > 7 = [-3.05442, -2.3620]$. The power of the test $1 - \beta = .999$.

References

[1] Misty C. Bentz and Sarah Katz. The agn black hole mass database. *Publications of the Astronomical Society of the Pacific*, 127(947):67, jan 2015.

[2] Fulvio Melia. *High-energy astrophysics*. Princeton series in astrophysics. Princeton University Press, Princeton, N.J, 2009. OCLC: ocn228632818.

[3] P. Padovani, D. M. Alexander, R. J. Assef, B. De Marco, P. Giommi, R. C. Hickox, G. T. Richards, V. Smolcic, E. Hatziminaoglou, V. Mainieri, and M. Salvato. Active Galactic Nuclei: what's in a name? 2017.