1 Problem 1

1.1 Deriving Bayes Theorem

To derive Bayes Theorem we start with the law of total probability,

$$P(A) = P(A|B_1)\,P(B_1) + ... + P(A|B_k)\,P(B_k), \tag{1}$$

and the multiplicative rule of probability,

$$P(A \cap B_i) = P(A|B_i)\,P(B_i). \tag{2}$$

Then, using the fact that the intersection operator is commutative (ie. $P(A \cap B_i) = P(B_i \cap A)$) and following from eq. 2 we can write,

$$P(A|B_i)\,P(B_i) = P(A \cap B_i) = P(B_i \cap A) = P(B_i|A)\,P(A). \tag{3}$$

Rearranging eq. 3,

$$P(B_i|A) = \frac{P(A|B_i)\,P(B_i)}{P(A)}, \tag{4}$$

which is the most basic form of Bayes Theorem. Then, substituting in eq. 1 for $P(A)$ we get the more general form of Bayes Theorem,

$$P(B_i|A) = \frac{P(A|B_i)\,P(B_i)}{P(A|B_1)\,P(B_1) + ... + P(A|B_k)\,P(B_k)}, \tag{5}$$

where i is a given event number and there are k events.

1.2 Discussing Bayes Theorem

- Bayes Theorem is not able to be used simply "as is". Instead, as the statistician/scientist we must decide on the best values for the prior and likelihood functions. The prior quantifies our past understanding of the material and the likelihood function defines the probability of the data distribution under a certain set of model paramters. Both the prior and the likelihood functions require a level of subjectivity on the users part to make use of Bayes Theorem.

- There are a lot of challenges in using Bayes Theorem and it appears to vary between fields. Some of the more common challenges are [1, 2, 4, 7]

  1. It requires a model of the physical system, which could be very complex.
  2. If the model is computationally complex it can make it nearly impossible (or at least take a very long time) for solutions to Bayes Theorem to converge.
  3. Putting all relevant past knowledge into one simple prior function.
  4. Different statisticians/scientists agreeing on the "correct" prior function for a scenario.

- A frequentist would argue that Bayes Theorem is suboptimal because it requires the person using it to input their opinion on the subject as the prior [1]. This could also be lead to the argument that the results of a calculation using Bayes Theorem can be misleading because it is not purely objective mathematics. For example, one could tune the prior function until they get a favorable, but misleading, result.

1.3 Multivariate Gaussian Likelihood Function

The multivariate gaussian likelihood function is given by [5]

$$\mathcal{L}(\mu, C) = \frac{1}{(2\pi)^{n/2}\,|C|^{1/2}}\ \exp\left[-\frac{1}{2}(x-\mu)^T C^{-1}(x-\mu)\right] \tag{6}$$

The variables in eq. 6 are given in Table 1.3.

Table 1:

Variable	Description
x	A vector representing the data set that we are trying to find the likelihood of.
C	The covariance matrix of the data set x with dimensions nxn.
μ	The mean of the data set x.
n	The length of x or the dimension of C.

x is a given subset of the population data and is found by surveying that population. In astronomy, this would be like taking data on targets that make up a subset of a larger population. μ is computed by taking the mean of the data set x. n is simply calculated by taking the length of the data set, which is also the dimension of C.

C is the covariance matrix of x and is more complex to calculate. The covariance matrix tells us how much each vector in the data set is related. Ideally, the covariance matrix would be known exactly from the model. But, if it is not, the covariance matrix can be estimated from the data using eq. 7 on every combination of column vectors in the data set. In eq. 7, n is the length of both of the sample vectors, x_i is a given point in the first vector, y_i is a given value in the second vector, $\bar{x}$ is the sample mean of the first vector, and $\bar{y}$ is the sample mean of the second vector.

$$C = \sum_{i=1}^{n} \frac{(x_i - \bar{x})(y_i - \bar{y})}{n-1} \tag{7}$$

As one could imagine, performing this operation on a very large data set is very computational expensive ($\mathcal{O}(n^2)$).

2 Problem 2

Five probability distributions are:

1. Normal Distribution

2. Poisson Distribution
3. Binomial Distribution
4. Exponential Distribution
5. Uniform Distribution

Information about each of these distributions, including the equation and plots of them, are in the subsections below.

2.1 Normal Distribution

The equation for the normal (gaussian) distribution is give by eq. 8[5]. A plot of the normal distribution is in Figure 1.

$$P(x|\mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \tag{8}$$

Table 2: Normal Distributions Parameters

Variable	Description
x	The data set that is distributed over this probability distribution.
μ	The mean of the sample data set x. This affects the location of the center of the distribution.
σ	The standard deviation of the sample data set x. This affects how widespread the distribution is.

2.2 Poisson Distribution

The equation for the Poisson Distribution is given in eq. 9[5]. The meaning and impact of each parameter on the distribution is in Table 2.2. A plot of the poisson distribution is in Figure 2.

$$P(x|\lambda) = e^{-\lambda}\frac{\lambda^x}{x!} \tag{9}$$

Table 3: Poisson Distributions Parameters

Variable	Description
x	The number of a given event.
λ	The average number of the given event which effects the skewness of the distribution and the location of the peak. A higher λ means the distribution is less skewed and the peak is shifted to higher x values.

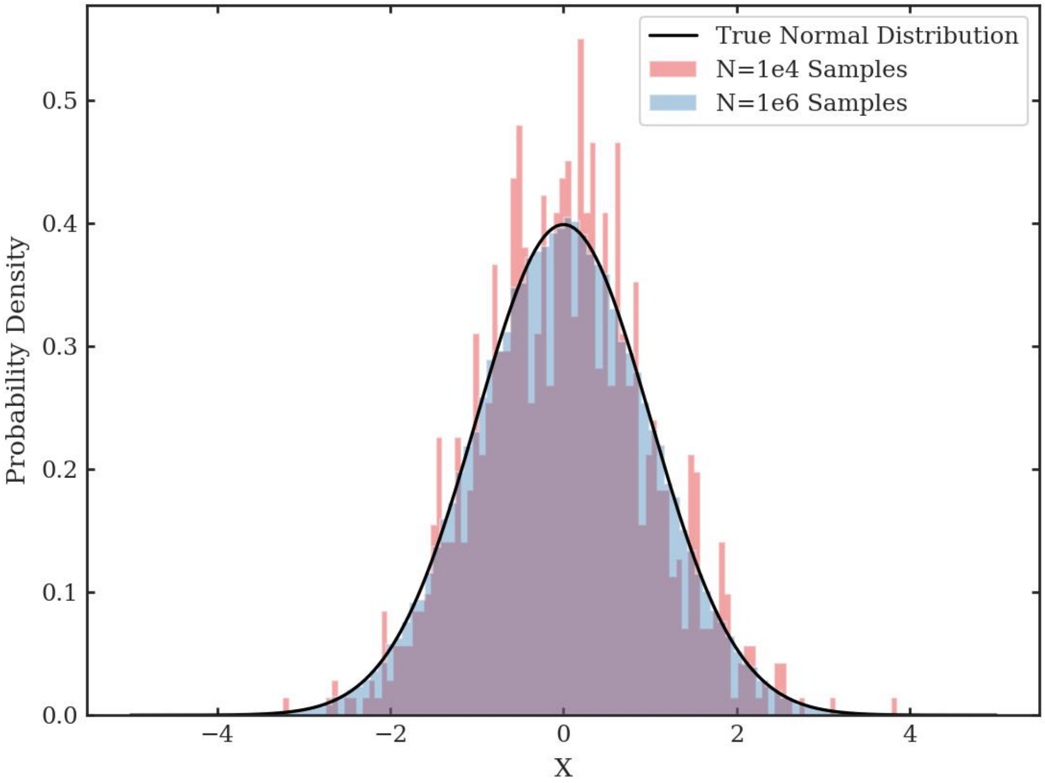

Figure 1: The true normal distribution with two different random samples of it overplotted.

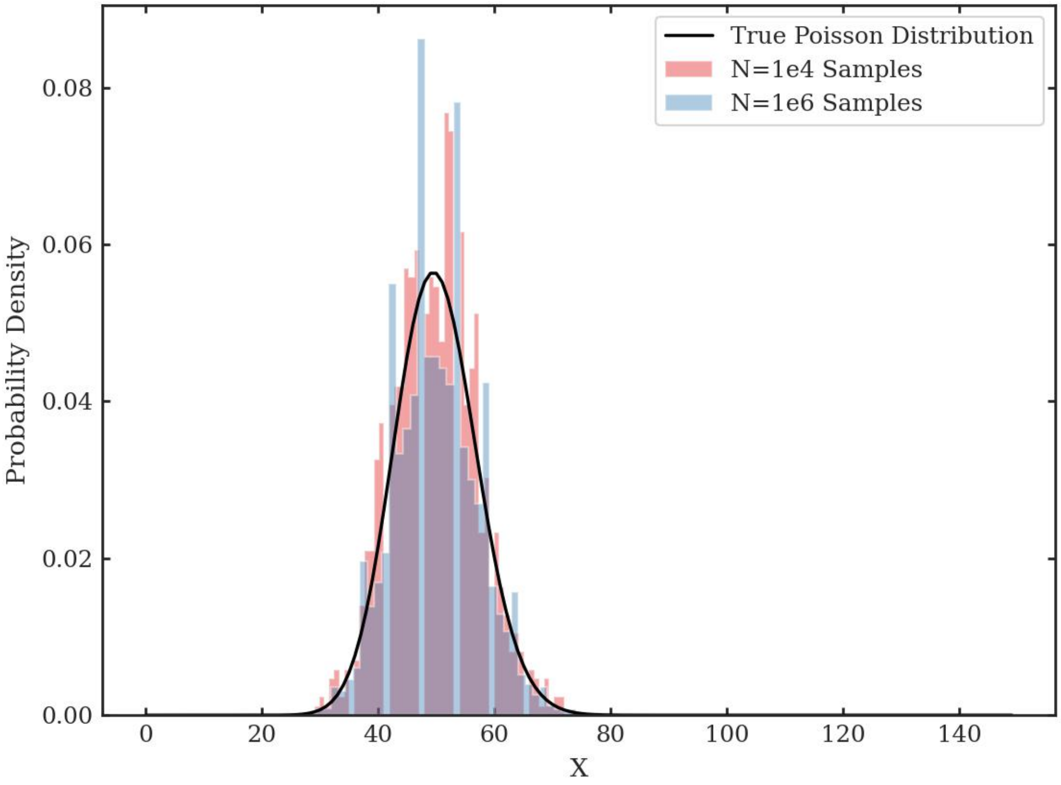

Figure 2: The true poisson distribution with two different random samples of it over plotted.

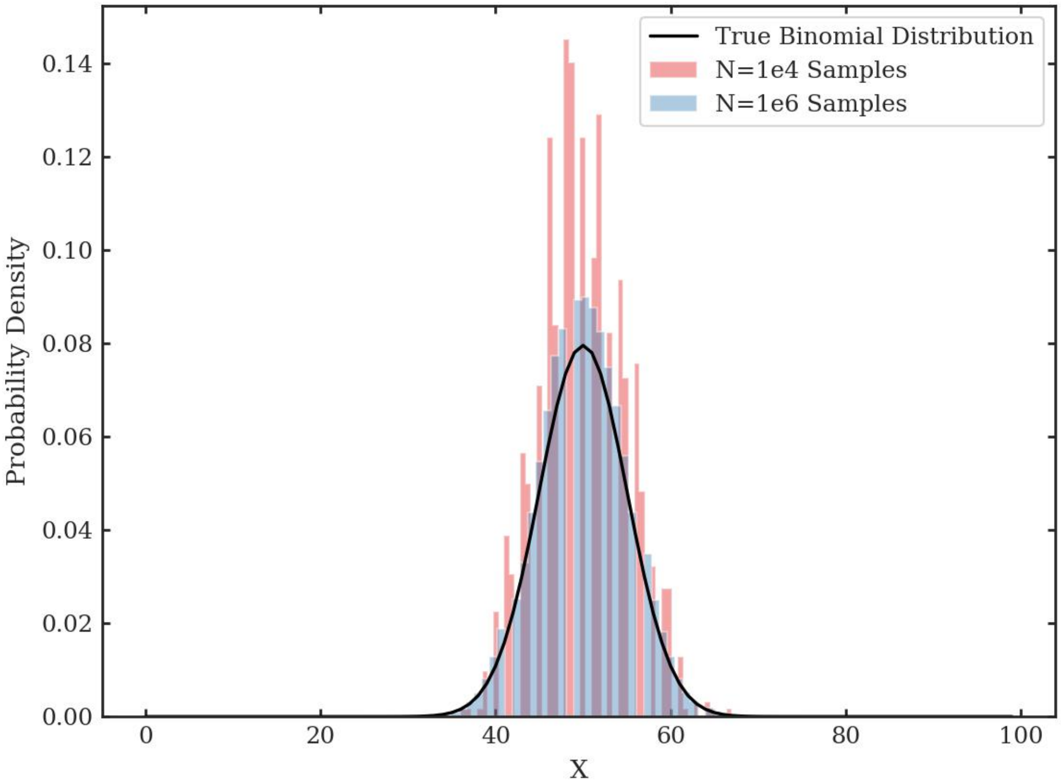

Figure 3: The true binomial distribution with two different random samples of it over plotted.

2.3 Binomial Distribution

The equation for the Binomial Distribution is given in eq. 10[5]. The meaning and impact of each parameter on the distribution is in Table 2.3. A plot of the binomial distribution is in Figure 3.

$$P(x|f,N) = \binom{N}{x} f^x (1-f)^{N-x} \tag{10}$$

Table 4: Binomial Distributions Parameters

Variable	Description
r	An integer
f	The bias parameter which effects the steepness (or "pointyness") of the binomial distribution. A higher value of f will make the PDF taller and narrower.
N	The total number of trials. This simply changes the location of the PDF.

2.4 Exponential Distribution

The equation for the Exponential Distribution is given in eq. 11[5]. The meaning and impact of each parameter on the distribution is in Table 2.4. A plot of the exponential distribution

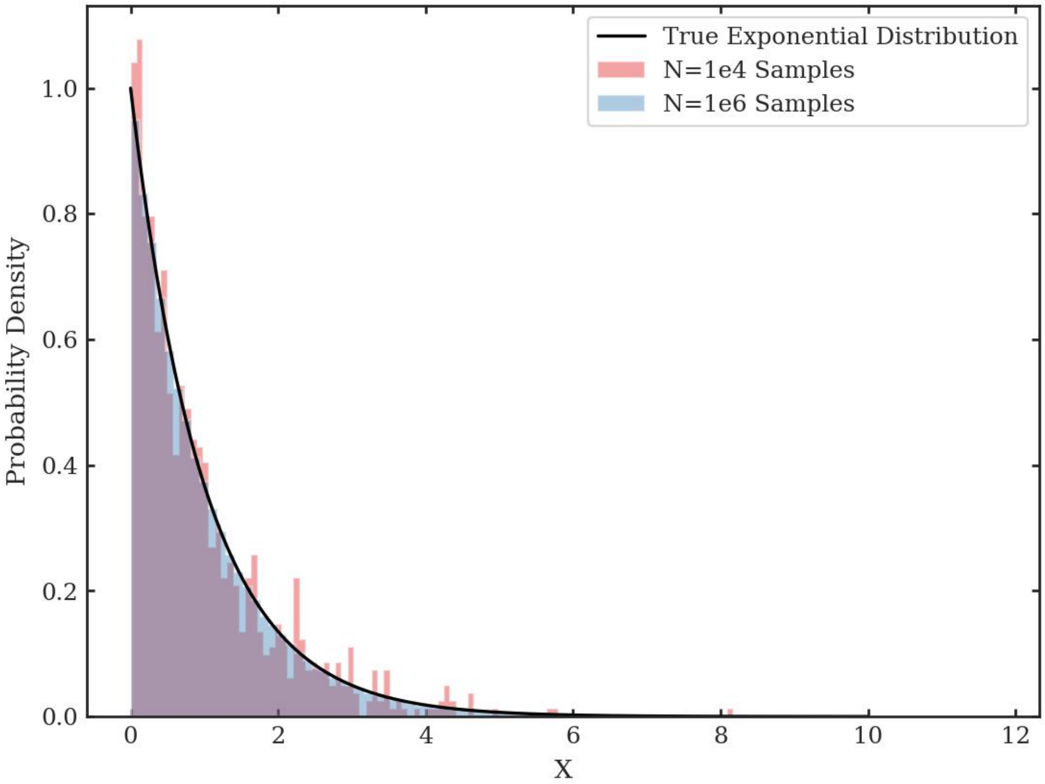

Figure 4: The true exponential distribution with two different random samples of it over plotted.

is in Figure 4.

$$P(x|\lambda) = \lambda e^{-x\lambda} \tag{11}$$

Table 5: Exponential Distributions Parameters

Variable	Description
x	The number of occurrences of an event.
λ	The bias or rate parameter. This effects the steepness of the PDF where a larger λ makes the PDF fall towards zero faster.

2.5 Uniform Distribution

The equation for the Uniform Distribution is given in eq. 12[5]. The meaning and impact of each parameter on the distribution is in Table 2.5. A plot of the uniform distribution is in Figure 5.

$$P(x) = C \tag{12}$$

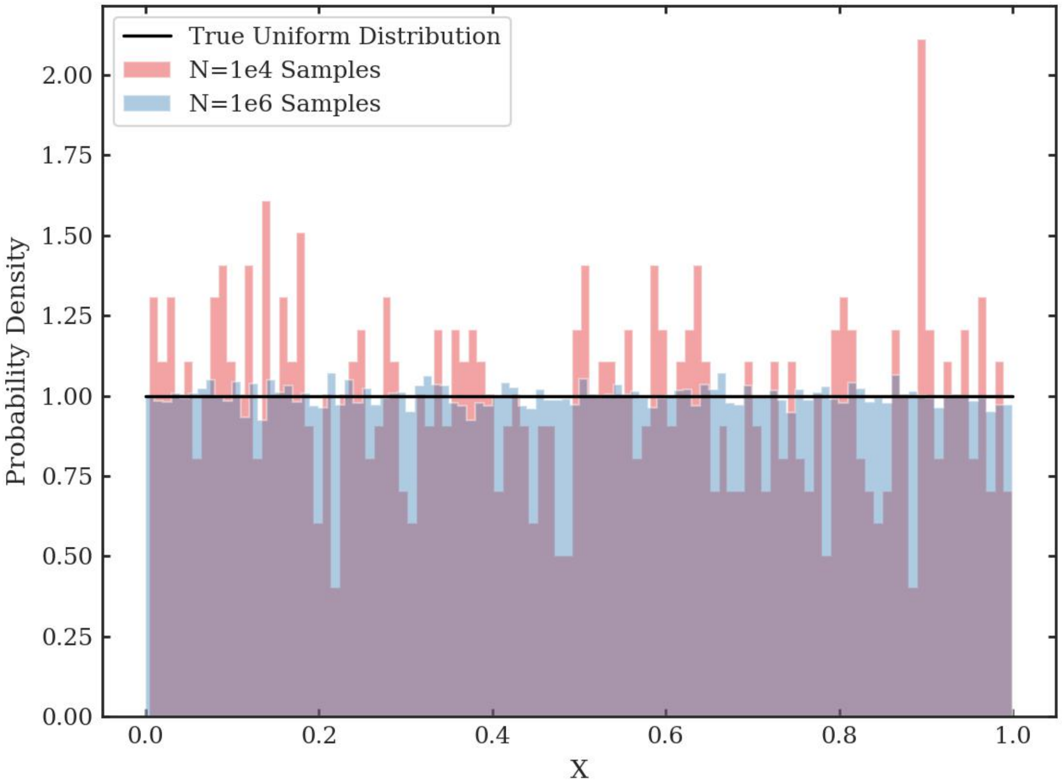

Figure 5: The true uniform distribution with two different random samples of it over plotted.

Table 6: Uniform Distributions Parameters

Variable	Description
x	The data vector.
C	A constant probability such that $C = 1/\ell$ where ℓ is the length of x. That way the area under the probability distribution normalizes to 1. Changing C simply raises or lowers the probability density and changes the probability of each value of x accordingly.

2.6 Closed Form Moments

2.6.1 Exponential Distribution Moments

The moments of the exponential distribution can be computed with eq. 13,

$$M_n = \int_0^\infty (x - \lambda^{-1})^n \lambda e^{-x\lambda} dx, \tag{13}$$

where all of the variables are the same as in Table 2.4 and M_n is the nth moment of the exponential distribution. The first four closed form moments are in Table 7 by solving eq. 13 for $n = 1 \cdots 4$. Table 7 also lists the true and estimated moments for $\lambda = 1$. This shows that as the number of random samples increases so does the accuracy of the moments. It also shows that the higher moments are more vulnerable to small fluctuations in the curve than the lower moments.

Table 7: Exponential Distribution Moments

Moment No.	Moment Closed Form Expression	True Moment	Estimated Moment (N=1e4)	Estimated Moment (N=1e6)
1	0	0	0	0
2	$1/\lambda^2$	2	0.94	0.98
3	$2/\lambda^3$	6	1.71	1.95
4	$9/\lambda^4$	9	7.34	8.67

2.6.2 Normal Distribution Moments

The moments of the normal distribution can be computed with eq. 14,

$$M_n = \int_{-\infty}^{\infty} \frac{(x-\mu)^n}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx, \tag{14}$$

where, similar to above, all of the variables are the same as in Table 2.1 and M_n is the nth moment of the normal distribution. The first four closed form moments of the normal distribution are well known[1] and are in Table 8 for $n = 1 \cdots 4$. Similar to above, as the moment numbe increases so does the vulnerability of the resulting value to small fluctuations in the dataset.

Table 8: Poisson Distribution Moments

Moment No.	Moment Closed Form Expression	True Moment	Estimated Moment (N=1e4)	Estimated Moment (N=1e6)
1	0	0	0	0
2	σ^2	1	0.96	1.00
3	0	0	0.11	0.001
4	$3\sigma^4$	3	2.81	3.03

2.7 Astronomical Use Cases

2.7.1 Case 1: Binomial Distribution

An example of a use case for the binomial distribution in astronomy is in [6][2].

- The goal of this study is to reanalyze the population of spin directions in spiral galaxies in the publicly classified Galaxy Zoo data set. This is motivated by a hot debate about if the spin parity of spiral galaxies is violated, in other words, is there an uneven distribution of clockwise and counterclockwise spins in spiral galaxies.

[1] https://en.wikipedia.org/wiki/Normal_distribution

[2] https://arxiv.org/pdf/2302.06530.pdf

- The random variable is the spin of a spiral galaxy. The underlying population is that the authors hope to generalize their results to all spiral galaxies.

- The binomial distribution is used for this analysis because there are exactly two possibilities for the spin direction: clockwise or counterclockwise. They do not calculate any moments of the binomial distribution, they simply use it to calculate the p-value of the distribution.

- The main statistical result of this paper is that there is less than a 1% chance that the measured spin parity in their study is by random chance ($p \sim 0.01$). This means that, generalizing to the entire population of spiral galaxies, there is likely a violation of the spin parity of spiral galaxies.

- They discuss a possible bias towards classifying galaxies with a counter clockwise spin with the computational tool they use to measure the galaxy spin direction, `SpArcFire`. This bias may make it so they are underestimating the probability that the measured spin parity is by random chance. However, they attempt to account for this bias by classifying both the regular galaxy images and the mirrored galaxy images. They also discuss possible systematical uncertainty in `SpArcFire`'s classification which they found through visual inspection. They reduce this by cutting out as many non-spiral galaxies as possible using different methods. Finally, they have a good amount of statistical power because their dataset is $\sim 300,000$ spiral galaxies from all parts of the sky.

2.7.2 Case 2: Normal Distribution

An interesting application of the normal distribution in astronomy is found in [3][3].

- The authors goal is to use a novel energy scale calibration for the XMM-Newton telescope to measure the velocity distribution and structure of the intracluster medium. In this work, they apply this technique to measure the velocity structures of the Virgo, Centaurus, and Ophiucus clusters.

- In this case, the random variable is a velocity of a spatially resolved point in the cluster. So, the underlying population is the velocity of a single point in space in the cluster.

- The Normal PDF is used to model the velocity distribution in each cluster and the first two moments of this distribution are computed (ie. the mean and standard deviation).

- The statistical conclusion is that for these three clusters, their velocity PDF's are well fitted to a normal PDF.

- They have large observational uncertainties on their data on the scales of 10kpc and on their velocities of $\sim$ 100km/s. They also have systematic uncertainties due to their S/N cutoff which they tested and show plots with varying degrees of S/N. They don't seem to be concerned with generalizing from three clusters to the entire cluster population

[3] https://arxiv.org/pdf/2307.02576.pdf

as a whole so they do not discuss their statistical power in this sense. They also do not discuss their statistical power when it comes to fitting the velocity distributions, or even give the number of values in their samples. This is slightly concerning given that one of their main conclusions is that the velocity distributions are best fit with a gaussian.

References

[1] Bradley Efron. Bayes' theorem in the 21st century. *Science*, 340(6137):1177–1178, 2013.

[2] Daniel Foreman-Mackey, David W. Hogg, Dustin Lang, and Jonathan Goodman. emcee: The mcmc hammer. *Publications of the Astronomical Society of the Pacific*, 125(925):306–312, March 2013.

[3] Efrain Gatuzz, R. Mohapatra, C. Federrath, J. S. Sanders, A. Liu, S. A. Walker, and C. Pinto. Measuring the hot ICM velocity structure function using XMM-Newton observations. *MNRAS*, 524(2):2945–2953, September 2023.

[4] SandeepK Gupta. Use of bayesian statistics in drug development: Advantages and challenges. *International Journal of Applied and Basic Medical Research*, 2(1):3, 2012.

[5] D.J.C. MacKay. *Information Theory, Inference and Learning Algorithms.* Cambridge University Press, 2003.

[6] Darius Mcadam and Lior Shamir. Reanalysis of the Spin Direction Distribution of Galaxy Zoo SDSS Spiral Galaxies. *Advances in Astronomy*, 2023:1–12, February 2023.

[7] Sanjib Sharma. Markov chain monte carlo methods for bayesian data analysis in astronomy. *Annual Review of Astronomy and Astrophysics*, 55(1):213–259, 2017.