Shiv Patel Astro 410 – Midterm April 3rd 2023

1.1) For this question, we can use Bayes Theorem to derive the join posterior distribution of the parameters, μ, α_d, and A, with the observed data D.

$$P(\mathrm{A}|\mathrm{B}) = \frac{P(B|A) * P(A)}{P(B)}$$

$$P(\mu, \alpha_\mathrm{d}, \mathrm{A}|\mathrm{y_i}, \mathrm{x_i}) = \frac{P(\mathrm{y_i}|\mu, \alpha_\mathrm{d}, \mathrm{A}, \mathrm{x_i}) * P(\mu, \alpha_\mathrm{d}, \mathrm{A})}{P(\mathrm{y_i}, \mathrm{x_i})}$$

$$P(\mu, \alpha_\mathrm{d}, \mathrm{A}|\mathrm{y_i}, \mathrm{x_i}) = \frac{\prod_{\mathrm{i}=1}^{\mathrm{n}} P(\mathrm{y_i}|\mu, \alpha_\mathrm{d}, \mathrm{A}, \mathrm{x_i}) * P(\mu)\mathrm{P}(\alpha_\mathrm{d})\mathrm{P}(\mathrm{A})}{P(\mathrm{y_i}, \mathrm{x_i})}$$

$$P(\mu, \alpha_\mathrm{d}, \mathrm{A}|\mathrm{y_i}, \mathrm{x_i}) = \frac{\prod_{\mathrm{i}=1}^{\mathrm{n}} P(\mathrm{y_i}|\mathrm{m_i}, \sigma_\mathrm{i}) P(\mu)\mathrm{P}(\alpha_\mathrm{d})\mathrm{P}(\mathrm{A})}{P(\mathrm{y_i}, \mathrm{x_i})}$$

The Metropolis algorithm is a type of Markov Chain Monte Carlo that allows us to ample from a probability distribution without using the denominator. Because the metropolis algorithm does not care if the probabilities that are fed into it is normalized or not.

$$P(\mu, \alpha_\mathrm{d}, \mathrm{A}|\mathrm{y_i}, \mathrm{x_i}) = \prod_{\mathrm{i}=1}^{\mathrm{n}} P(\mathrm{y_i}|\mathrm{m_i}, \sigma_\mathrm{i}) P(\mu)\mathrm{P}(\alpha_\mathrm{d})\mathrm{P}(\mathrm{A})$$

$$P(\mu, \alpha_\mathrm{d}, \mathrm{A}|\mathrm{y_i}, \mathrm{x_i}) = \prod_{\mathrm{i}=1}^{\mathrm{n}} \frac{1}{\sqrt{2\pi}\sigma_\mathrm{i}} e^{-\left(\frac{\mathrm{y_i}\text{-}\mathrm{m_i}}{\sqrt{2\pi}\sigma_\mathrm{i}}\right)^2} * \frac{1}{\sqrt{2\pi}\sigma_\mu} e^{-\left(\frac{\mathrm{y_i}\text{-}\mu_\mu}{\sqrt{2\pi}\sigma_\mu}\right)^2} * \frac{1}{\sqrt{2\pi}\sigma_{\alpha_\mathrm{d}}} e^{-\left(\frac{\mathrm{y_i}\text{-}\mu_{\alpha_\mathrm{d}}}{\sqrt{2\pi}\sigma_{\alpha_\mathrm{d}}}\right)^2} * \frac{1}{\sqrt{2\pi}\sigma_\mathrm{A}} e^{-\left(\frac{\mathrm{y_i}\text{-}\mu_\mathrm{A}}{\sqrt{2\pi}\sigma_\mathrm{A}}\right)^2}$$

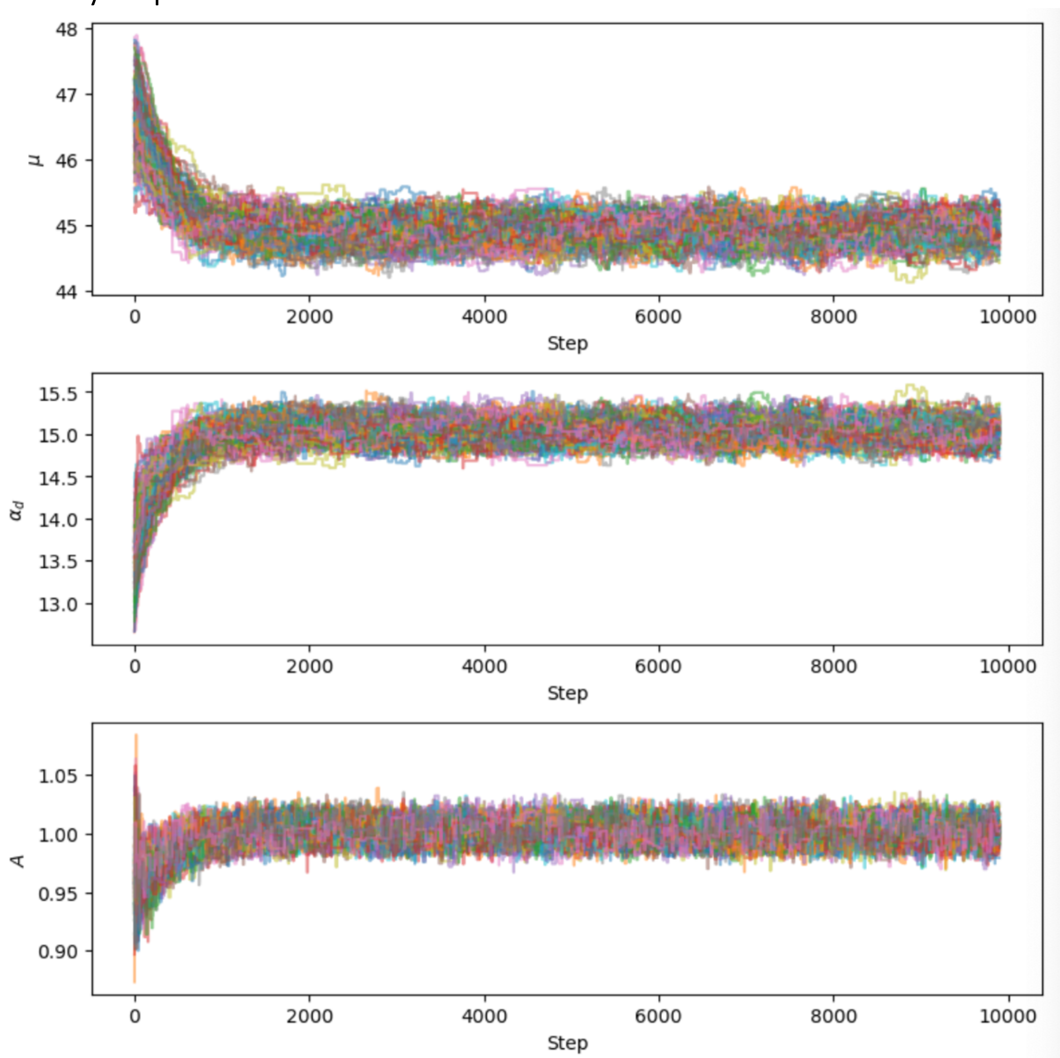
The likelihood function represents the probability of observing the data with the model parameters. The prior function denoted by the log_prior() function in my code, represents the prior belief about the value of these parameter before observing the data. We also assume that the prior probability uses a gaussian distribution centered at the mean values of $\mu 0, \alpha_\mathrm{d} 0$ and $\mathrm{A}0$ with standard deviations of σμ, σαD, and σA. My code implements the logarithm of the posterior probability which is proportional to the product of the log of the likelihood and the log of the probability. By doing the log of the posterior probability my code avoids numerical under/overflow as a safeguard to dealing with very small or large probabilities. For more context I also implemented a norm_pdf() function that makes it a little easier to code. This function calculates the logarithm of the PDF of a normal distribution with the given mean and standard deviation because it is proportional to the negative square of the different between the input variable and the mean divided by the square of the standard deviation. My code uses bayes theorem to determine the most probable values of the parameters by considering the prior and post assumptions about them.

1.2) Construct a Markov Chain to sample from the posterior, assuming flat prior for μ, αD and A, and using a uniform proposal distribution. Write a computer program to Implement the Markov Chain.

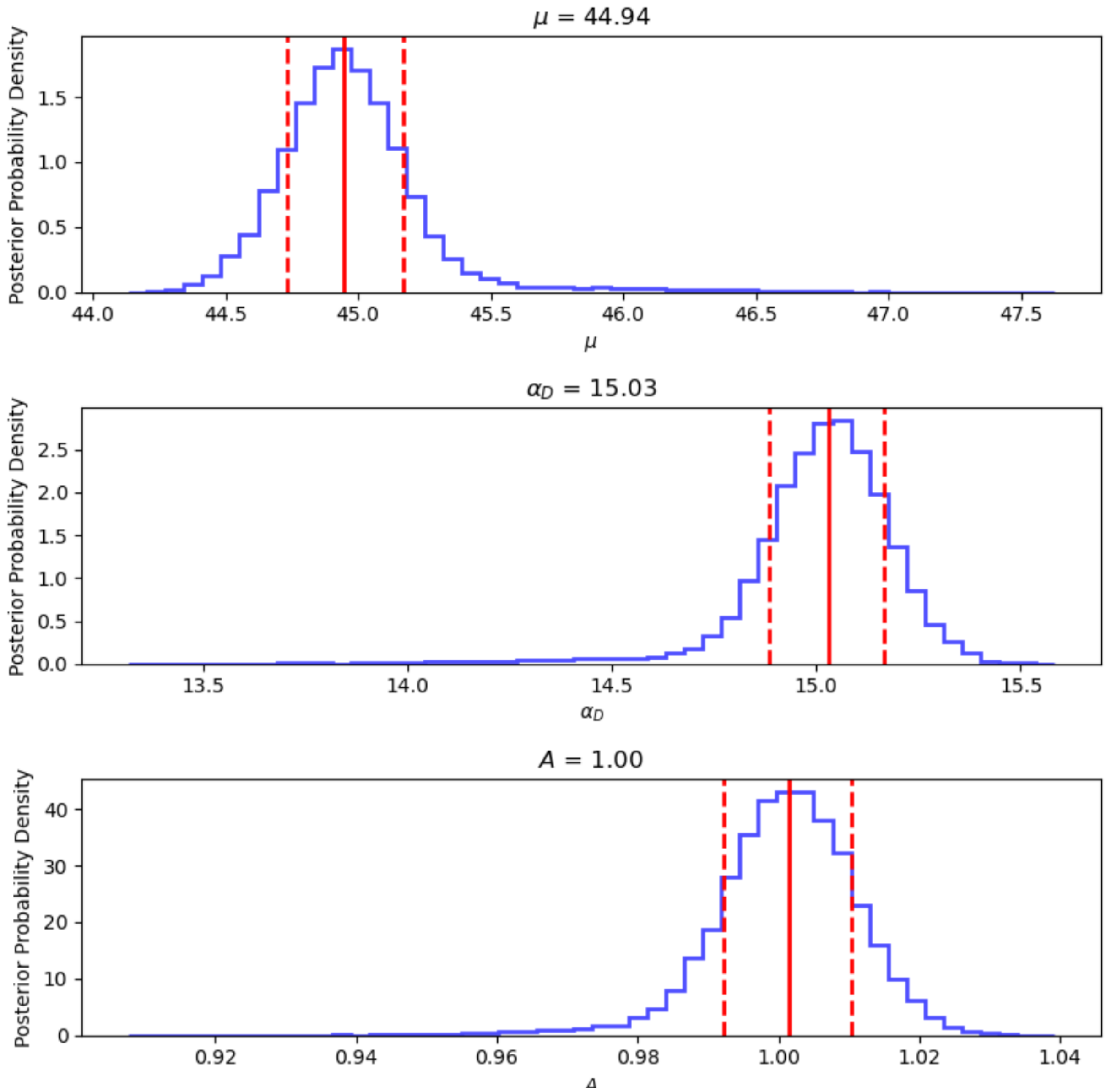
Implemented in the Code.

1.3) Run the MCMC program and experiment with different starting values of the parameters and the width of the proposal distribution function, and choose the most appropriate ones that result in a well-mixed Markov Chain. Plot μ, αD, and A along the Markov Chain history, and determine the burn-in period. Plot the distribution functions of μ, αD, and A using the Markov Chain after discarding the burn-in period.

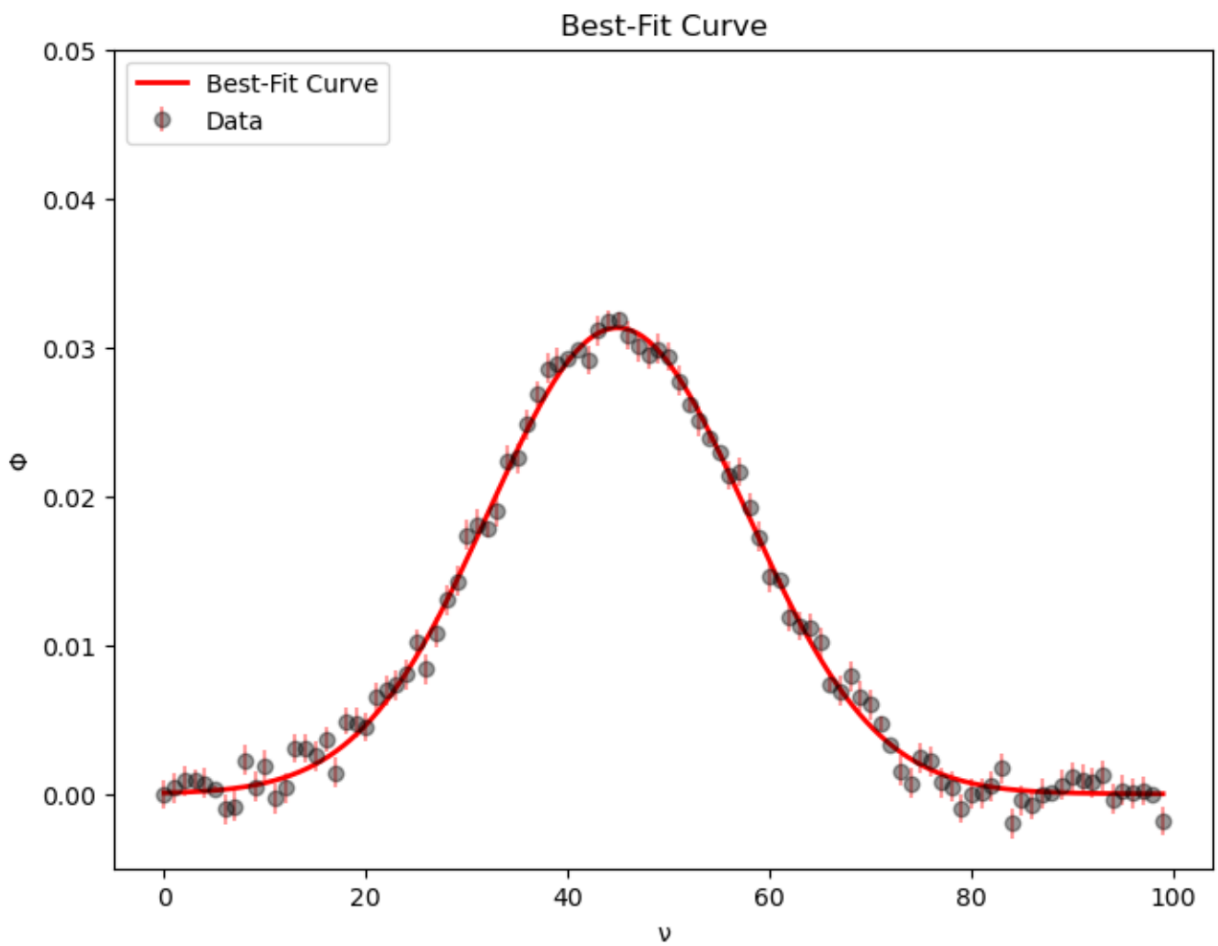
For this question I used an if statement to traverse the data file with the functions I defined in the previous question. And did 10000 iterations (which took 6 mins to run) and my output was:



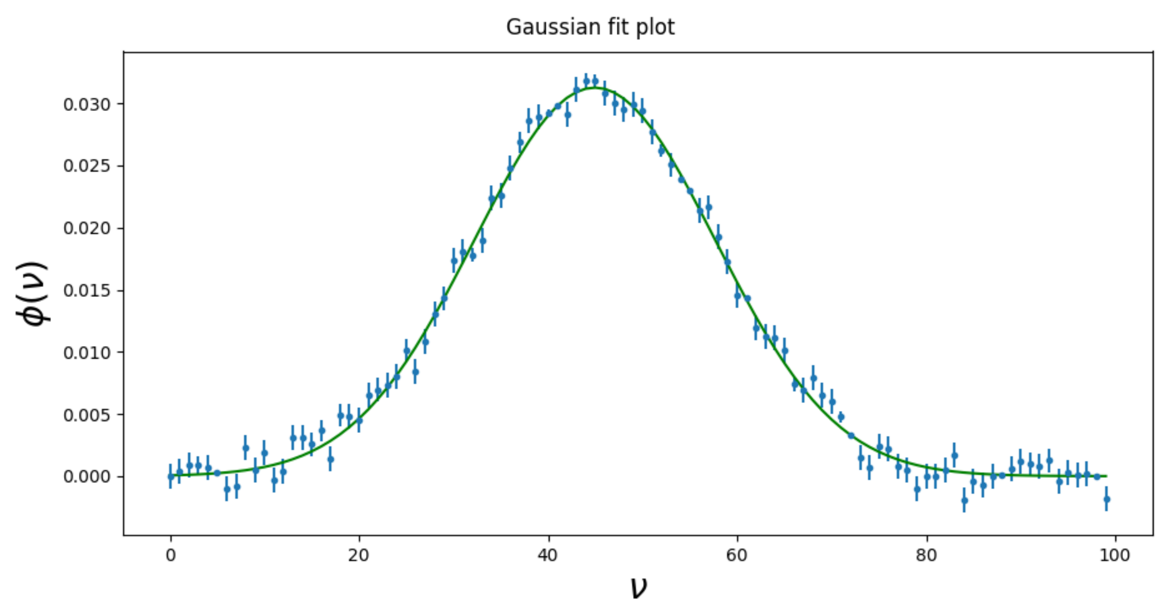
1.4) Show the best fitting parameters and estimate the mean and standard deviation of (μ, α_D, A). Plot the data and the best fit curve. Compare them with the best-fit parameters and errors derived from the non-linear least squares fit in homework 2.



μ = 44.94
Posterior Probability Density
0.0
0.5
1.0
1.5
44.0
44.5
45.0
45.5
46.0
46.5
47.0
47.5
μ
α_D = 15.03
Posterior Probability Density
0.0
0.5
1.0
1.5
2.0
2.5
13.5
14.0
14.5
15.0
15.5
α_D
A = 1.00
Posterior Probability Density
0
10
20
30
40
0.92
0.94
0.96
0.98
1.00
1.02
1.04
A



From homework 2 the best fit line is:



Which is almost identical the Gaussian plot used in this assignment.