CTA 200: Random Walks and Simulated Data

Problem 1: Random walks have a big influence on how we think about real world data. We may have some underlying model and is also influenced by noise. This noise effectively shifts the data by a random number. If we were to repeat a measurement many times, this would act like a random walk. The simplest example is a 1-dimensional random walk where we start a point $x_0 = 0$ and then we recursively pick the next point by a random process

$$x_{i+1} = x_i \pm 1 \tag{0.1}$$

where we pick $\pm$ each with probability 1/2.

a) Write a script that generates an N dimensional random walk. If we run the simulation many times, we should find that the average location after N steps should be 0 and the average distance $|x| = \sqrt{N}$. Show this by performing an $N = 10, 100, 1000$ random walk a m times and take look at the average value of x_N and $|x_N|$. As you increase the m, how do these averages compare expected mean values? Show this by making a log-log plot for various values of m.

b) Now, instead of averaging over m random walks, do simulations with very large N and make a log-log plots for at how $|x_N|/\sqrt{N}$ scales with N? From the same random walk, implement a different way of analyzing the data that reproduces the results from part (a)?

d) Now make a gaussian random walk such that

$$x_{i+1} = x_i + \xi \tag{0.2}$$

where ξ is drawn from a gaussian distribution with a standard deviation of 1. How do these results compare with part (a) and (b).

Problem 2: A common situation in cosmology is that both the signal and the are random. The propose of this problem is to simulate these types of data sets.

a) Write a script that will generate a time series $x(t) = \sum_{n=1}^{N} A_n \cos(\omega_n t + \phi_n)$ where $\omega_n = \omega_0 n/N$, ϕ_n is a uniformly distributed random phase and A_n is drawn from a random distribution with a variance

$$\langle A_n A_{n'} \rangle = \delta_{nn'} \left(\frac{A_0}{\omega_n} + B_0 \right) \tag{0.3}$$

In this model A_0 is our signal and B_0 is our noise.

b) Now sample this time series by averaging every 10 successive points. Take the fourier transform of the sampled series and compute A_n^2. Repeat this produce m times and compute the average value of A_n^2 and compute the statistical error in the measurement. Make a plot of A_n^2 versus n and include the error bar on each point.

c) How did the averaging procedure affect the measured value of $\langle A_n A_{n'} \rangle$? Show this by making a plot of A_n^2 as a function of n with several values for the number of points being averaged. Can you correct for it (hint: what is the window function)?

d) Fit your data with the model in equation (0.3) for A_0 and B_0 and compute the error on these parameters (from the covariance matrix) using "scipy.optimize.curve_fit". How do these errors scale with m?