APM496 Assignment 1

Brian Lee 1002750855

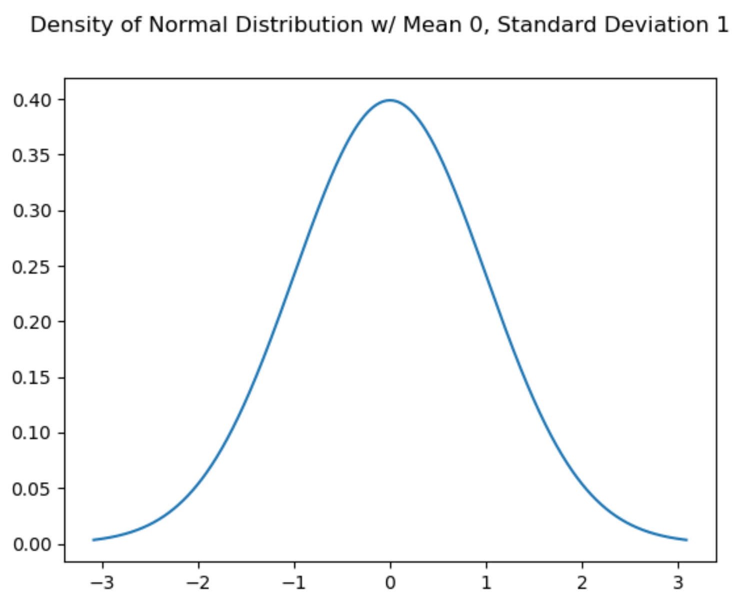
1 Problem 1

(a) My favourite unvariate distribution is the normal distribution. A normal distribution with mean μ and variance σ^2 (i.e. $N(\mu, \sigma^2)$) has the probability density

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right)$$

The two parameters are the location parameter μ (the mean), and the scale parameters σ (standard deviation).

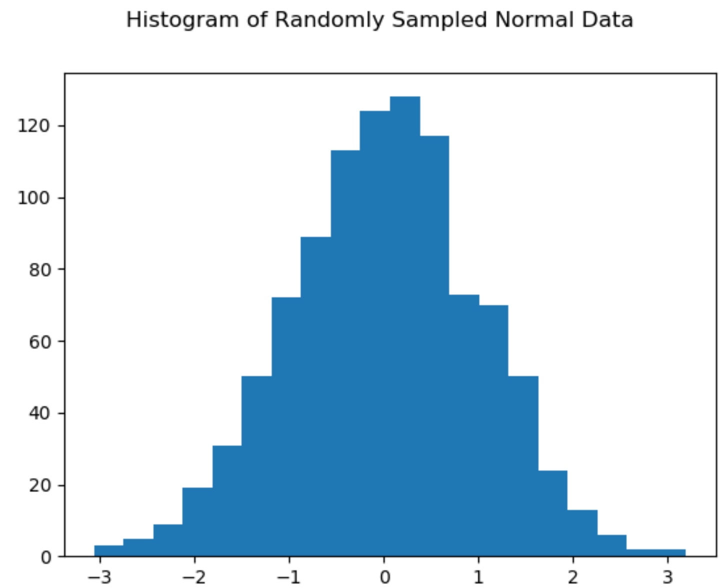
(b) A plot of the density of a normal distribution with $\mu = 0$ and $\sigma = 1$ (i.e. the standard normal distribution) is given below



(c) Due to the central limit theorem, we expect that for independent data with some regularity conditions (eg. identically distributed or satisfies some Lyapunov or Lindenberg condition) that the average of the data, given a large enough sample, will be normally distributed. This means that by under some regularity conditions on your population, if you sample repeatedly and take an average, the distribution of this average will be normally distributed.
For some more concrete examples of normally distributed data in real life, the distribution of heights or the distribution of test scores would approximately normally distributed given a large enough sample size. A normal distribution is appropriate to model heights, as you would expect most people's height to be around average (so there's concentration about the mean, as in a normal distribution), only a small number of people are very tall or very short (decay in the tails), and you would expect the number of people that are taller or shorter than the average to be approximately equal (symmetry of the gaussian). The same kind of logic applies to test scores, given a large enough sample size.

2 Problem 2

(a) We randomly sampled 1000 data points from a standard normal distribution, and the histogram is given below



(b) Using the data estimators function, we get that for this sample,

$$\text{Mean} = 0.0119, \quad \text{Standard deviation} = 0.9956, \quad \text{Skew} = -0.0594, \quad \text{Kurtosis} = -0.0510$$

(Note that this kurtosis is excess kurtosis, so that a normal distribution should have kurtosis 0)

(c) We generated another 1000 samples from the same distribution by running the code again, and the resulting sample mean is

$$\text{Mean} = 0.05343$$

The sample mean here is different from the sample mean above. This is because we used a different random sample; due to the 'randomness' in the way we sample data points from the distribution, and since we're only taking 1000 samples, it's expected that there is variance in these results.