Homework 4

1. Generate a sample of size $n = 10$ from a χ^2_5 distribution. We are trying to estimate the mean with a confidence interval. Construct a 95% confidence interval for the mean based on the central limit theorem. Note that n is quite small. Check if the interval contains the true mean, 5. Now construct a bootstrap interval based on 100 bootstrap samples and perform same check. Carry out the above procedure (starting with generating the sample of size 10) 1000 times and verify which of the intervals (based on CLT or Bootstrap) have coverage closer to 95%.

2. The aim is to generate observations from the density

$$p(x) \propto (2 + \sin(5x) + \sin(2x)) * \exp(-x^2).$$

   Do this by setting up a Metropolis Hastings algorithm with transition density $q(x, y) = \mathcal{N}(x - y, 1)$.

3. Generate 100 data points from the model $y = x + \epsilon)$ with $\epsilon \sim \mathcal{N}(0, 0.5^2)$ and x is the subset of (0,1) with 100 equally spaced points starting at 0.01. Fit a kernel regression with normal kernel and bandwidth 0.25. Find the 5-fold cross-validation error. Repeat the process by changing the bandwidth from 0.02 to 0.5 in increments of 0.02 and find the optimal bandwidth. Plot the data overlayed with the kernel fit for the best choice of bandwidth. Also plot the CV-error as a function of bandwidth.