Homework 1.1.2

1. Using my random seed, 779 / 10000 (~7.8%) averages were in the range [-0.01, 0.01]

2. Starting at n = 100, if we decrease *n* then the quoted fraction in 1. will be smaller. If we increase *n*, then the quoted fraction will be larger.

3. Using my random seed, the value of epsilon covering 99% of the data is 0.2652. This value of epsilon was found by trial and error.

4. Starting at n = 100, if n increases then the same value of epsilon will cover more than 99% of the data. Hence, if *n* increases, then we should have a smaller value of epsilon to cover 99% of the data. Similarly, if *n* decreases then the same epsilon (from n = 100) will cover less than 99% of the data. Hence, if *n* decreases we should have a larger value of epsilon to cover 99% of the data.

5. I interpret this question as asking "What happens to the fraction of X_bar s in the range [-0.01, 0.01] and the value of epsilon covering 99% of the data when we parameterize the Gaussian distribution differently" (the trends identified in questions 2 and 4 should still hold if we parameterize the Gaussian differently). To make the comparisons easier, let's keep mu = 0 and only change sigma. Using sigma = 10, we see that 81/10000 (~0.8%) averages are in the range [-0.01, 0.01], and an epsilon value of 0.2652 covers ~21% of the data. Since making sigma larger will "spread out" our data more, this result makes sense. Similarly, if we set sigma = 0.5, then 1571/10000 (~16%) averages are in the range [-0.01 0.01] and an epsilon of 0.2652 covers 100% of the data. Making sigma smaller will "decrease the spread" of our data, so again this result makes sense.

Another interpretation of this question is "What happens if we use a non-Gaussian distribution to answer questions 1-4" (e.g. a gamma distribution or beta distribution). There are many different parameterizations of the gamma and beta distributions that give very different probability density functions, so it would be harder to draw meaningful general comparisons. Hence, I chose to parameterize my Gaussian differently when answering question 5., since I thought the comparison would be more meaningful.

As discussed in class and in solns, questions could have been worded better. The main take-away is that X_bar distribution will be Gaussian even if distribution of X_i is not Gaussian.