Homework # 13
Due 5/7

For all questions below, let X and Y be jointly distributed r.v. Suppose we sample jointly from X and Y, collecting the samples $(\hat{X}_i, \hat{Y}_i)$ for $i = 1, 2, \ldots, N$. Let X_e, Y_e be the r.v.'s defined by the empirical distribution of the samples.

1. Show through LLN arguments that $\text{Cov}(X_e, Y_e) \to \text{Cov}(X, Y)$ as $N \to \infty$.

2. Consider the regression model $y_i \sim ax_i + b + \epsilon_i$ where the ϵ_i are iid versions of $\epsilon \sim \mathcal{N}(0, \sigma^2)$.

   (a) Define the likelihood function $L(a, b, \sigma^2)$ by,

   $$L(a, b, \sigma^2) = \prod_{i=1}^{N} f(y_i - ax_i b) \tag{1}$$

   where $f()$ is the pdf of $\mathcal{N}(0, \sigma^2)$. Explain why maximizing the likelihood is a reasonable way to select the "best fit" a, b, σ^2.

   (b) Explain why the values of a, b, σ^2 that maximize $L(a, b, \sigma^2)$ are the same as those that maximize the log-likelihood, $\log L(a, b, \sigma^2)$.

   (c) Find a, b, σ^2 that maximize $\log L(a, b, \sigma^2)$ by computing $\nabla \log L(a, b, \sigma^2)$ and then solving $\nabla \log L(a, b, \sigma^2) = 0$. In particular, you should find that the optimal a is given by

   $$a = \frac{\text{Cov}(X_e, Y_e)}{V[X_e]} \tag{2}$$

3. Use the formula of a in Problem 2 to provide the following examples.

   (a) Through a plot of the samples, provide two examples of $(\hat{X}_i, \hat{Y}_i)$ for $i = 1, 2, \ldots, N$, one of which has $\text{Cov}(X_e, Y_e) = 1$ and one of which has $\text{Cov}(X_e, Y_e) = 0$.

   (b) For each of your examples in (a), provide jointly distributed r.v. (X, Y) that are consistent with the samples you plotted.