• (30 points) Below are descriptions of the joint distribution over two random variables, $X$ and $Y$. For each of these distributions, answer the following questions: Are $X$ and $Y$ independent? (yes/no) Are they correlated? (yes/no) What is their correlation coefficient? (give the number).

1. (5pts) $X$ is distributed according to the normal distribution $\mathcal{N}(5, 2)$, $Y = 3X$.
   
   **Answer:**
   $X$ and $Y$ are dependent, and the correlation coefficient is 1.

2. (5pts) $X$ is distributed uniformly in the range $[0, 1]$, $Y = -3X$.
   
   **Answer:**
   $X$ and $Y$ are dependent, and the correlation coefficient is $-1$.

3. (10pts) $X$ is distributed uniformly in the range $[-1, 1]$. 
   
   $P(Y|X) = \begin{cases} 
   1/2 & \text{if } Y = 3X \\
   1/2 & \text{if } Y = -3X \\
   0 & \text{otherwise}
   \end{cases}$
   
   **Answer:**
   $X$ and $Y$ are dependent. The correlation coefficient is zero because of symmetry. More formally, because $E(Y|X)$ is identically zero.
4. (10pts) The joint distribution of $X$ and $Y$ is defined by the density function

$$p_{X,Y}(x,y) = \frac{1}{Z} \begin{cases} xy + x - y - 1 & \text{if } 9 \leq x \leq 10 \text{ and } 0 \leq y \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

Were $Z$ is a normalization factor so that the integral of the density function over all $X, Y$ is 1.

**Answer:**
Although this is not obvious from the way the distribution is defined, $X$ and $Y$ are independent.

As $xy + x - y - 1 = (x - 1)(y + 1)$ we can write the joint density $p_{X,Y}(x,y)$ where $p_X(x)p_Y(y)$

$$p_X(x) = \frac{1}{Z} \begin{cases} x - 1 & \text{if } 9 \leq x \leq 10 \\ 0 & \text{otherwise} \end{cases}$$

$$p_Y(y) = \frac{1}{Z} \begin{cases} y + 1 & \text{if } 0 \leq y \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

As $X$ and $Y$ are independent, their correlation coefficient is zero.
• (30 points) Let $X_1, X_2, \ldots, X_{1000}$ be independent random variables distributed uniformly in the range $[0, 1]$. Let $S = \sum_{i=1}^{1000} X_i$.

1. (15pts) What are the mean and standard deviation of $S$.

**Answer:**

The mean of $X_i$ is $1/2$, therefore the mean of $S$ is

$$E(S) = E\left(\sum_{i=1}^{1000} X_i\right) = \sum_{i=1}^{1000} E(X_i) = 500$$

The variance of $X_i$ can be calculated by integrating over the uniform density:

$$V(X_i) = \int_{-\infty}^{\infty} (x - 1/2)^2 p_U(x)dx = \int_{0}^{1} (x - 1/2)^2 dx = \int_{-1/2}^{1/2} x^2 dx = \left[\frac{x^3}{3}\right]_{-1/2}^{1/2} = \frac{1}{12}$$

Using that we can calculate the variance of $S$:

$$V(S) = V\left(\sum_{i=1}^{1000} X_i\right) = 1000V(X_i) = \frac{1000}{12}$$

and the standard deviation of $S$ is $\sqrt{V(S)} = \sqrt{\frac{1000}{12}}$

2. (15pts) Give an approximate formula for the CDF of $S$.

**Answer:**

As $S$ is the sum of a large number of random variables, we can approximate it’s distribution as a normal distribution. We calculated the mean and the variance of this normal distribution in the previous part of the question, and the mean and the variance fully specify the normal distribution. Therefore the CDF of the distribution of $S$ is:

$$F_S(a) = \int_{-\infty}^{a} \exp\left(-\frac{12(s - 500)^2}{500}\right)$$