1 Methods for More Than Two Random Variables

1.1 Joint Distributions and Independence

- Independent Random variables: Random variables $X_1, X_2, \ldots, X_n$ are independent, if for all $(x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$,

$$F_{X_1,X_2,\ldots,X_n}(x_1,x_2,\ldots,x_n) = F_{X_1}(x_1)F_{X_2}(x_2)\cdots F_{X_n}(x_n).$$

- If $X_1, X_2, \ldots, X_n$ are discrete, then they are independent if for all $(x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$, we have

$$P_{X_1,X_2,\ldots,X_n}(x_1,x_2,\ldots,x_n) = P_{X_1}(x_1)P_{X_2}(x_2)\cdots P_{X_n}(x_n).$$

- If $X_1, X_2, \ldots, X_n$ are continuous, then they are independent if for all $(x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$, we have

$$f_{X_1,X_2,\ldots,X_n}(x_1,x_2,\ldots,x_n) = f_{X_1}(x_1)f_{X_2}(x_2)\cdots f_{X_n}(x_n).$$

- If random variables $X_1, X_2, \ldots, X_n$ are independent, then we have

$$E[X_1X_2\cdots X_n] = E[X_1]E[X_2]\cdots E[X_n].$$

Random variables $X_1, X_2, \ldots, X_n$ are said to be independent and identically distributed (i.i.d.) if they are independent, and they have the same marginal distributions:

$$F_{X_1}(x) = F_{X_2}(x) = \ldots = F_{X_n}(x), \text{ for all } x \in \mathbb{R}.$$
1.2 Sums of Random Variables

- A random variable $Y$ is given by
  
  $$ Y = X_1 + X_2 + \ldots + X_n $$

- The linearity of expectation gives
  
  $$ EY = EX_1 + EX_2 + \cdots + EX_n. $$

- We can find the variance of $Y$

  $$ Var(Y) = Cov\left(\sum_{i=1}^{n} X_i, \sum_{j=1}^{n} X_j\right) $$

  $$ = \sum_{i=1}^{n} \sum_{j=1}^{n} Cov(X_i, X_j) $$

  $$ = \sum_{i=1}^{n} Var(X_i) + 2 \sum_{i<j} Cov(X_i, X_j). $$

  In general

  $$ Var\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} Var(X_i) + 2 \sum_{i<j} Cov(X_i, X_j) $$

  If $X_1, X_2, \ldots, X_n$ are independent,

  $$ Var\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} Var(X_i). $$

2 Limit Theorems

For i.i.d. random variables $X_1, X_2, \ldots, X_n$, the **sample mean**, denoted by $\overline{X}$, is defined as

$$ \overline{X} = \frac{X_1 + X_2 + \ldots + X_n}{n}. $$
The weak law of large numbers (WLLN)

Let $X_1, X_2, ..., X_n$ be i.i.d. random variables with expected value $EX_i = \mu < \infty$. Then, for any $\epsilon > 0$,

$$\lim_{n\to\infty} P(|\bar{X} - \mu| \geq \epsilon) = 0.$$ 

The Central Limit Theorem (CLT)

Let $X_1, X_2, ..., X_n$ be i.i.d. random variables with expected value $EX_i = \mu < \infty$ and variance $0 < \text{Var}(X_i) = \sigma^2 < \infty$. Then, the random variable

$$Z_n = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} = \frac{X_1 + X_2 + ... + X_n - n\mu}{\sqrt{n}\sigma}$$

converges in distribution to the standard normal random variable as $n$ goes to infinity, that is

$$\lim_{n\to\infty} P(Z_n \leq x) = \Phi(x), \quad \text{for all } x \in \mathbb{R},$$

where $\Phi(x)$ is the standard normal CDF.
How to Apply The Central Limit Theorem (CLT)

Here are the steps that we need in order to apply the CLT:

1. Write the random variable of interest, \( Y \), as the sum of \( n \) i.i.d. random variable \( X_i \)'s:
   \[
   Y = X_1 + X_2 + ... + X_n.
   \]

2. Find \( EY \) and \( \text{Var}(Y) \) by noting that
   \[
   EY = n\mu, \quad \text{Var}(Y) = n\sigma^2,
   \]
   where \( \mu = EX_i \) and \( \sigma^2 = \text{Var}(X_i) \).

3. According to the CLT, conclude that
   \[
   \frac{Y - EY}{\sqrt{\text{Var}(Y)}} = \frac{Y - n\mu}{\sqrt{n\sigma}}
   \]
   is approximately standard normal; thus, to find \( P(y_1 \leq Y \leq y_2) \), we can write
   \[
   P(y_1 \leq Y \leq y_2) = P\left( \frac{y_1 - n\mu}{\sqrt{n\sigma}} \leq \frac{Y - n\mu}{\sqrt{n\sigma}} \leq \frac{y_2 - n\mu}{\sqrt{n\sigma}} \right)
   \approx \Phi\left( \frac{y_2 - n\mu}{\sqrt{n\sigma}} \right) - \Phi\left( \frac{y_1 - n\mu}{\sqrt{n\sigma}} \right).
   \]
3 Problems

1. Let $X_1, X_2, X_3$ be i.i.d. random variables, where $X_i \sim \text{Bernoulli}(p)$. Define

   
   \[ Y_1 = X_1 X_2, \]
   \[ Y_2 = X_2 X_3, \]
   \[ Y_3 = X_3 X_1. \]

   If $Y = Y_1 + Y_2 + Y_3$, find

   (a) $E[Y]$.
   (b) $\text{Var}(Y)$. 


2. There are 36 students in ECE314. On the basis of years of experience, the instructor knows that the time needed to grade a randomly chosen exam paper is a random variable with mean $EX_i = \mu = 10$ min and standard deviation $\sqrt{\text{Var}(x)} = \sigma = 6\text{min}$. Assume that grading times are independent. The instructor starts grading at 4 P.M. and grades continuously.

(a) Find the probability that the grading is done by 9 : 30 P.M.
(b) Find the probability the grading is finished sometime between 9 : 30 P.M. and 10 : 30 P.M.