Sampling Distributions and the Central Limit Theorem

Recall: A (univariate) random variable is defined to be a function from a sample space S into the real numbers, ℝ. We've talked a lot about distributions of random variables (including multivariate random variables), and we've used these distributions to find probabilities and other quantities like expected values that can be used to make decisions.

Often, we are not interested in the random variable(s) themselves, but instead in functions of the random variable(s), such as the **sample mean.**

Specifically, we can focus on functions of the variables $Y_1, Y_2, \dots, Y_n$ **observed in a random sample** selected from the population of interest.

In such scenarios, we need the probability distribution of these functions of random variables in order to make inferences about population parameters, and, in turn, make decisions.

Y_i can be Discrete or Continuous.

Definition: (Random Sample)

$F(y)$

$Y_1, Y_2, \dots, Y_n$ are a **random sample** of size n from a population with distribution (CDF) ~~$F(\cdot)$~~ if $Y_1, Y_2, \dots, Y_n$ are:

- **independent** ⇒ Unrelated
- **identically distributed** (have the same probability distribution)

This is also called an **iid (independent and identically distributed)** sequence.

Example 1: Let Y be the number of calls received per hour at a customer service center. The observed values of Y in randomly selected 8 hours are given below:

$y_1 = 3, \quad y_2 = 1, \quad y_3 = 2, \quad y_4 = 1, \quad y_5 = 10, \quad y_6 = 3, \quad y_7 = 0, \quad y_8 = 12$

Here, Y_i is the number of calls received per hour and it can take any value from $\{0, 1, 2, \dots\}$. Y_i has a Poisson distribution with parameter λ which is the average number of calls per hour.

Since these 8 hours are randomly selected, Y_i are independent here.