Continuous Probability Distribution

Counting the Uncountable

Suppose we make \( n \) observations of a continuous variable between the values \( a \) and \( b \). To organize these data, we create a relative frequency histogram using \( m \) uniform intervals.

Let the function \( h(x) \) be defined such that \( h(x_i) \) is the relative frequency of the \( i^{th} \) interval.

The probability that the random variable \( X \) falls in the interval \((a,b)\) is the sum of the probabilities of the relative frequency histogram. That is,

\[
P(a < X < b) = P(x_1) + P(x_2) + P(x_3) + \ldots + P(x_m).
\]

Notice that the probability \( P(x_i) \) is the area of the bar in the relative frequency histogram. We know the height of the bar, \( h(x_i) \). Let the width of the bar be denoted

\[
\Delta x = \frac{b - a}{m}
\]

where \( m \) is the number of intervals in the histogram. Thus,

\[
P(a < X < b) = h(x_1)\Delta x + h(x_2)\Delta x + h(x_3)\Delta x + \ldots + h(x_m)\Delta x = \sum_{i=1}^{m} h(x_i)\Delta x
\]
To infinity and beyond…

Let us now consider what happens when the number of observations \( n \) increases without bound, and the width of the interval \( \Delta x \) subsequently decreases without bound.

The first assumption is that the relative frequency histogram \( h(x) \) will approach a continuous function \( f(x) \) that gives the true probability of selecting a continuous variable from \( a \) to \( b \).

Given that is the case, then

\[
P(a < X < b) = \lim_{\Delta x \to 0} \sum_{i=1}^{m} f(x_i) \Delta x
\]

Hopefully, the right hand side of the equation looks familiar. This is the definition of a definite integral of \( f \) from \( a \) to \( b \). Thus,

\[
P(a < X < b) = \int_{a}^{b} f(x) \, dx
\]

Continuous Random Variables

When dealing with a continuous random variable, a function \( f(x) \) must be known from either theory or observation that determines the true probability within each subinterval of all possible values. The function \( f(x) \) must be nonnegative for all possible values of the continuous random variable. The total area between the function \( f(x) \) and the \( x \)-axis must be 1 for all possible values. Finally, the probability of randomly selecting the variable \( X \) in the interval \( (a,b) \) is determined by the area bounded by the function \( f(x) \), the \( x \)-axis, and the vertical lines at \( x = a \) and \( x = b \).

The following definition states these ideas formally.

**Definition:** The **probability density function (p.d.f.)** of a continuous random variable \( X \) with sample space \( S \) that is an interval or union of intervals is an integrable function \( f(x) \) satisfying the following conditions:

a. \( f(x) \geq 0, \ x \in S \)

b. \( \int_{S} f(x) \, dx = 1 \)

c. If \( (a,b) \subseteq S \), then the probability of the event \( \{a < X < b\} \) is

\[
P(a < X < b) = \int_{a}^{b} f(x) \, dx
\]
1. The following graph represents a probability density function \( f(x) \) of the continuous random variable \( X \). Write an expression for the probability that \( 1 < X < 4 \). Identify on the graph the probability that \( 1 < X < 4 \).

2. Let the continuous random variable \( X \) be modeled by the probability density function \( g(x) = e^{-x} \) for values of \( x \) greater than 0. Determine the probability that \( 1 < X < 3 \). Represent the probability using the graph of \( g(x) \).