The Learning Hub

Data Science Fundamentals

Random Variables

The Learning Hub

Random Variables

What are Random Variables?

- **Random Variables** : They are variables whose values are unknown or vary over time or from one individual to another due to an underlying random process.

- Random variables are often classified as **discrete**, which are variables that have a finite or countable set. or **continuous**, which are variables that can have any values containing a whole interval of numbers

What are Random Variables?

Definition: Let Ω be a sample space corresponding to some experiments ξ and let $x : \Omega \rightarrow R$ be a function from the sample space to real line.

Then X is called a **random variable**.

What are Random Variables?

- Consider the experiment of rolling a fair die twice. The sample space Ω of this experiment has 36 sample points.
- Consider now the sum of the two rolls and call it (X), X is now dependent on the value of the sample points that prevails when the experiment is conducted. That is, X is a function of that event ω.

Example

- From a box containing 4 red balls and 3 black balls, two balls are randomly chosen in succession without replacement. assume Y is the number of red balls, find the possible outcomes and the values y of the random variable ?

Sample Space	y
RR	2
RB	1
BR	1
BB	0

Probability Mass Function & Probability Density Function

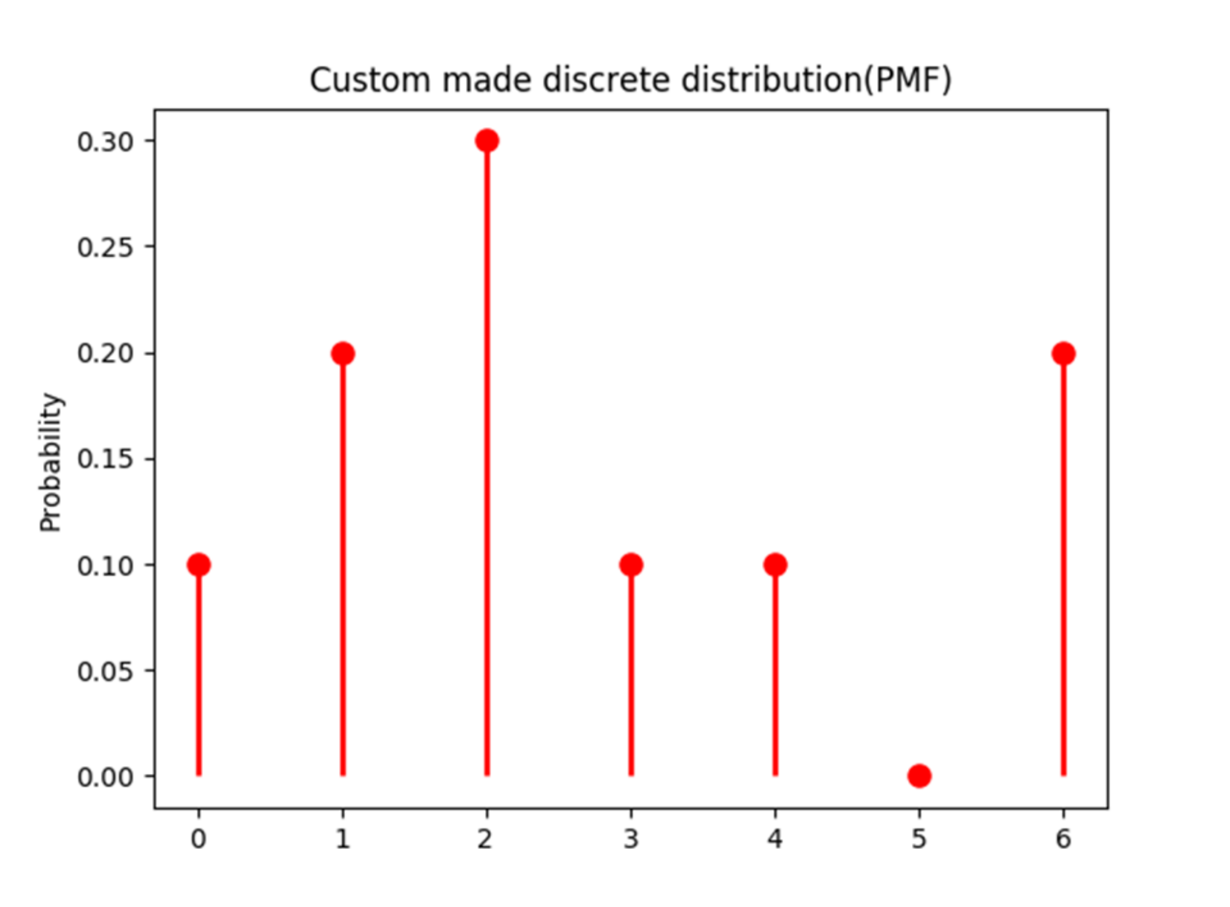
Probability Mass Function (PMF)

- **Definition:** Let $x\ :\ \Omega\ \rightarrow R$ be a discrete random variable taking a finite or countably infinite number of values $x_1, x_2, x_3, ..$
- The probability distribution or the probability mass function (PMF) of X is the function

$$p(x) = P(X = x), x = x_1, x_2, x_3, .. \, and\ p(x) = 0\ otherwise.$$

Probability Mass Function (PMF)

- In any probability mass function, $p(x) \geq 0 \; for \; any \; x, and \; \sum_i p(x_i) = 1$
- **NOTE:** Not all random variables are discrete. Some have uncountably many possible values like the value of a random fractional number, those are called **continuous random variables.**



Probability Mass Function (PMF)

Example: If a car agency sells 50% of its inventory of a particular foreign car equipped with side airbags, Find out Probability mass function (PMF) of the next 4 cars to be sold?

Solution: Let X be number of cars with airbags among the next 4 cars sold by the agency.

f(x) = $\binom{4}{x} 0.5^x (1-0.5)^{4-x} = \binom{4}{x} 0.5^x (0.5)^{4-x} = \binom{4}{x} 0.5^4 = \frac{1}{16}\binom{4}{x}$

x=0,1,2,3,4

Probability Mass Function (PMF)

- $f(0) = = \frac{1}{16}\binom{4}{0} = \frac{1}{16}$
- $f(1) = = \frac{1}{16}\binom{4}{1} = \frac{4}{16}$
- $f(2) = = \frac{1}{16}\binom{4}{2} = \frac{6}{16}$
- $f(3) = = \frac{1}{16}\binom{4}{3} = \frac{4}{16}$
- $f(4) = = \frac{1}{16}\binom{4}{4} = \frac{1}{16}$

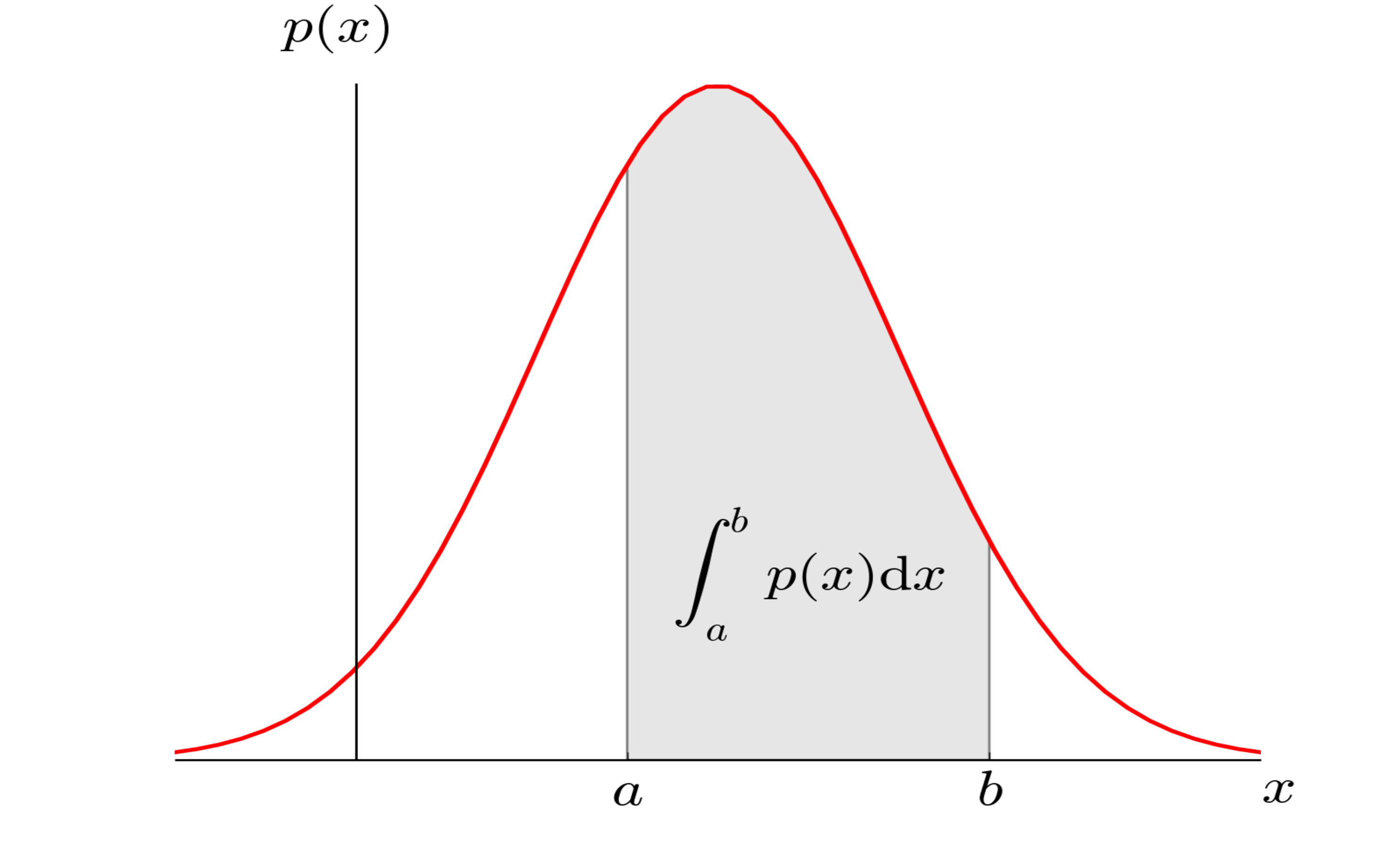
12

Probability Density Function (PDF)

- Consider the sample space $\Omega = [0,1]$, and define a random variable X by the function

$$X:\ \Omega \to R \ as\ X(\omega), \omega \in \Omega$$

- if we define a function $f(x)\ on\ [0,1]\ as\ f(x) = 1\ then\ we\ have$:
- $P(a \le X \le b) = b - a = \int_a^b f(x)dx$



The Learning Hub

Cumulative Distribution Function (CDF)

Cumulative Distribution Function (CDF)

- **Theorem:** A Function F(x) is the CDF of some real-valued variable X if and only if it satisfies all the following properties:

  - $\mathbf{0 \leq F(x) \leq 1 \; \forall \; x \in R}$
  - $\boldsymbol{F(x) \rightarrow 0 \; as \; x \longrightarrow -\infty, and \; F(x) \rightarrow 1 \; as \; x \rightarrow \infty}$
  - $\boldsymbol{Given \; any \; real \; number \; a, F(x) \downarrow F(a) as \; x \downarrow a}$
  - $\boldsymbol{Given \; any \; two \; real \; numbers \; x, y. x < y, F(x) \leq F(y)}$

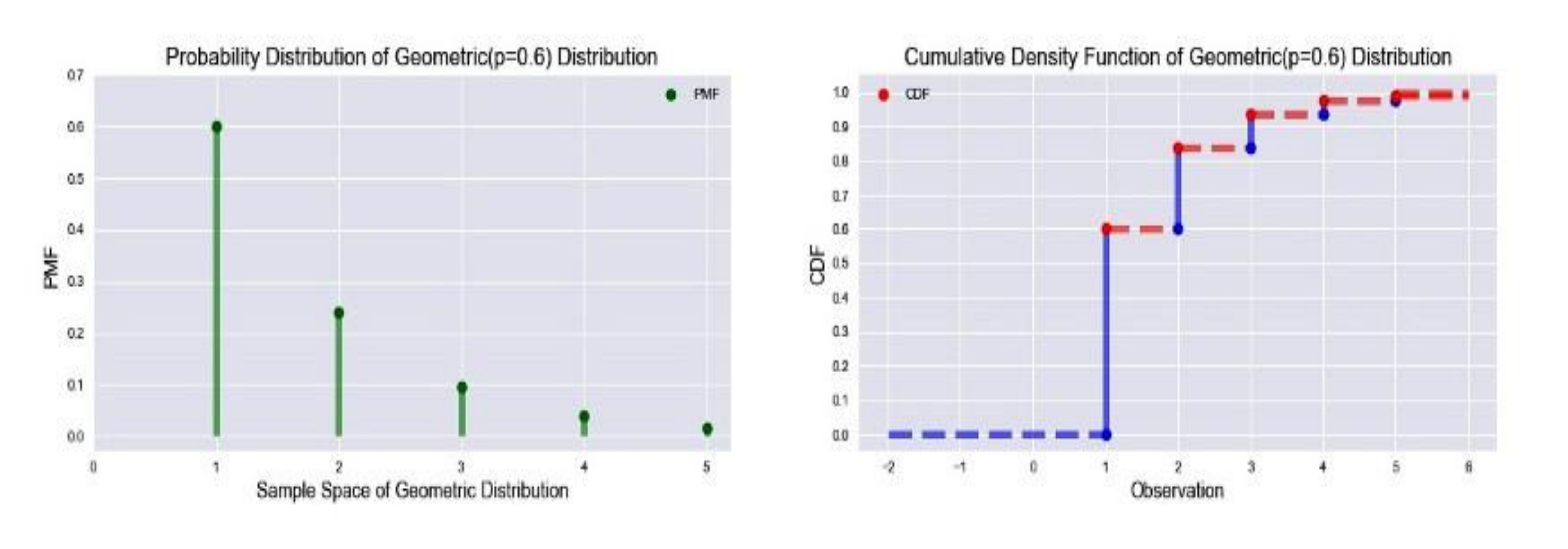
Cumulative Distribution Function (CDF)

- The CDF gives the probability that any random variable X is less than or equal to any given number x.
- It measures the probability that has been accumulated up to and including a given number x. Hence the name is **cumulative**.
- The notion of CDF is universal. It applies to all random variables, **discrete or continuous**.

Definition: The cumulative distribution function (CDF) of a random variable X is the function

$$F(x) = P(X \leq x),\ x \in \mathbf{R}$$

PMF VS CDF


Probability Distribution of Geometric(p=0.6) Distribution
PMF
Sample Space of Geometric Distribution
Cumulative Density Function of Geometric(p=0.6) Distribution
CDF
Observation

Independence of Random Variables

- **Definition:**

  Let $X_1, X_2, \dots, X_k$ be discrete random variables defined on the same sample space Ω.

  we say that $X_1, X_2, \dots, X_k$ are independent if $P(X_1 = x_1, X_2 = x_2, \dots, X_m = x_m) = P(X_1 = x_1)P(X_2 = x_2) \dots P(X_m = x_m), \forall\, m \le k$ and all $x_1, x_2, \dots, x_m$

- In other words, X_1 and X_2 are independent if the results of X_1 has no effect on X_2
- Example: Let X_1 and X_2 be two tosses of a coin. After X_1, the chances of X_2 being either heads or tails is unchanged hence the independence of the two random variables.

The Learning Hub

Continuous Random Variables

Continuous Random Variables properties

- Distribution that takes all values in a **nonempty interval**. In another word, the **entire real line**.
- Instead of discrete probability we use integrals, and instead of probability mass functions we use **density functions**.
- For a continuous random variable, any single value x is infinitely unlikely and has a **probability zero**.
- $P(X = x) = 0$ for each specific number x. Intervals with nonzero length will not usually however have a probability of zero.
- We calculate the probability of an interval as the integral of a density function of the interval.

20

Independence of Continuous Random Variables

- **Definition:**

$Let\ X_1, X_2, \dots, X_n\ be\ n\ random\ variables\ defined\ on\ some\ sample\ space\ \Omega.$

$X_1, X_2, \dots, X_n\ are\ independent\ if:$

$$P(X_1 \le x_1, X_2 \le x_2, \dots, X_n \le x_n) = \prod_{i=1}^{n} p(X_i \le x_i) \forall\, x_1, x_2, \dots, x_n \iff$$

$$P(X_1 > x_1, X_2 > x_2, \dots, X_n > x_n) = \prod_{i=1}^{n} p(X_i > x_i) \forall\, x_1, x_2, \dots, x_n$$

Example of Continuous Random Variables

- Consider the function:
- $$f(x) = \begin{cases} 48x(1-4x) & , 0 \le x \le \frac{1}{4} \\ 0 & , x < 0 \text{ or } x > 1 \text{ or } \frac{1}{4} < x < \frac{3}{4} \\ 48(1-x)(4x-3) & , \frac{3}{4} \le x \le 1 \end{cases}$$

- Function is always non-negative.

Nonzero between 0 and 0.25 and between 0.75 and 1 , and zero otherwise.

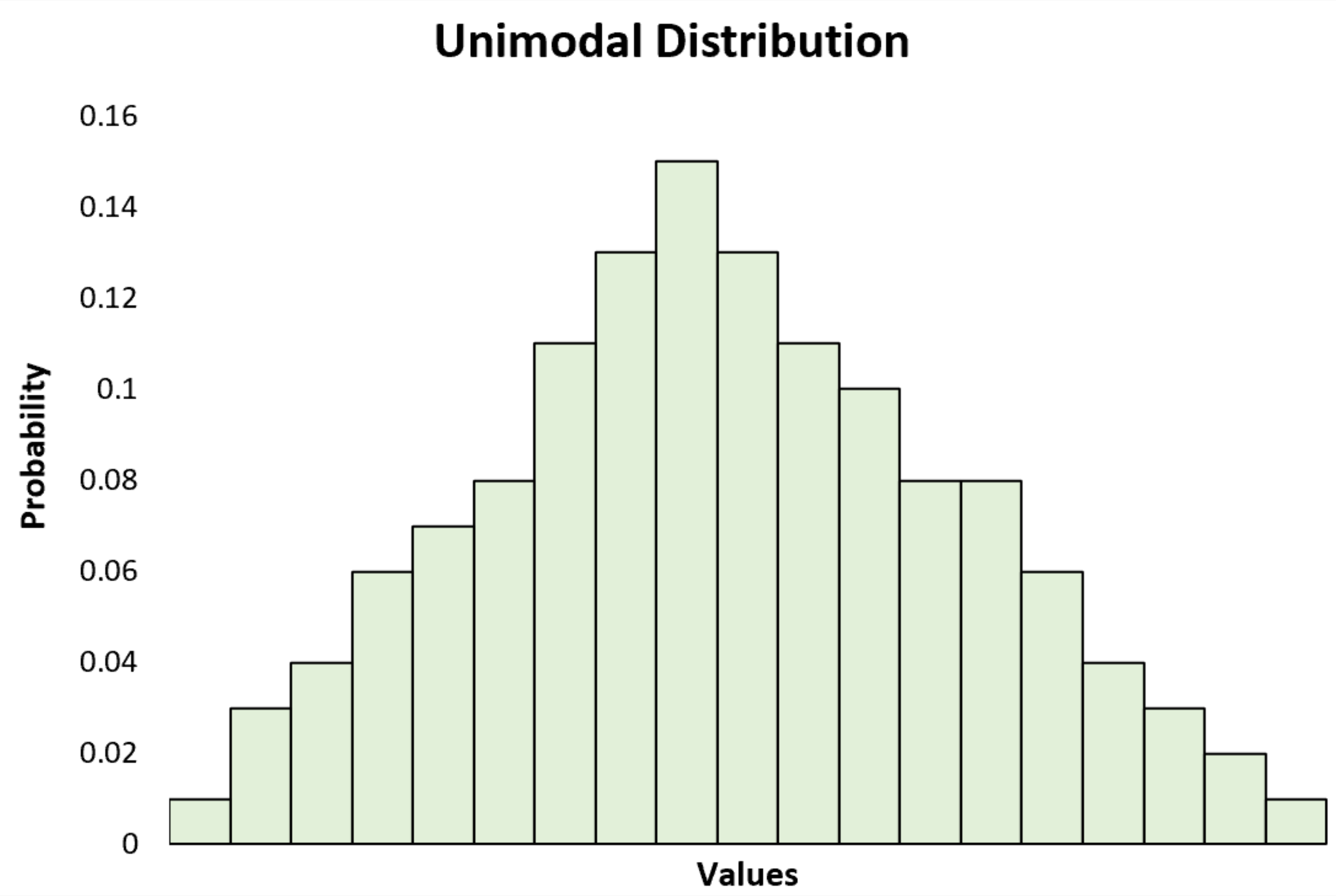
Density Functions

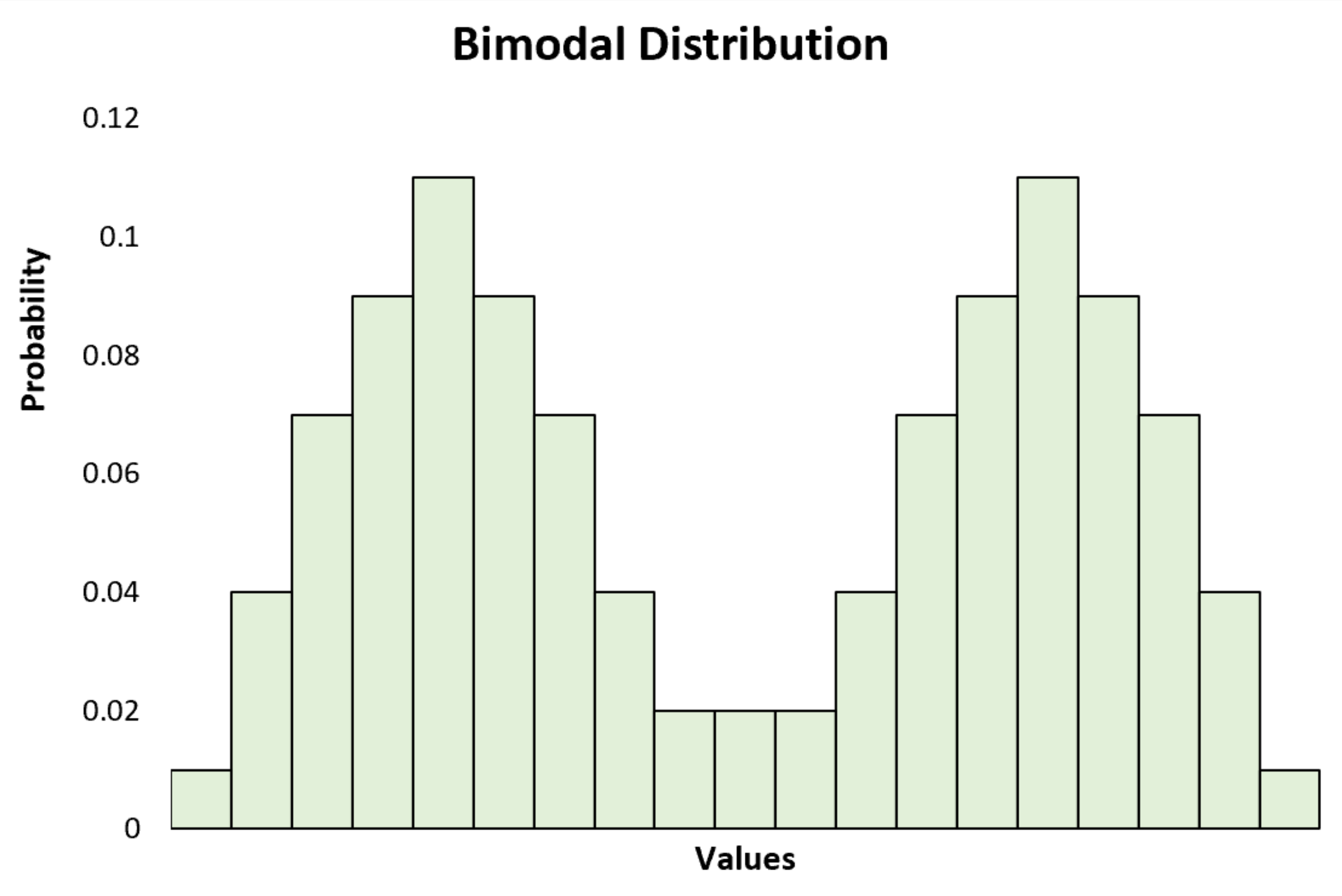
- The standard exponential density.
  - $f(x) = \begin{cases} e^{-x} & if,\ 0 \leq x < \infty \\ 0 & if\ ,\ x < 0 \end{cases}$
- The standard normal density.
  - $f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, -\infty < x < \infty$
- The standard double exponential density.
  - $f(x) = \frac{1}{2} e^{-|x|}, -\infty < x < \infty$
- The standard Cauchy density.
  - $f(x) = \frac{1}{\pi} \frac{1}{1+x^2}, -\infty < x < \infty$

Unimodal Vs Bimodal Distribution

Definition:

A density function $f(x)$ is called strictly unimodal at (or around) a number M if $f(x)$ is increasing for $x < M$ and decreasing for $x > M$.





The Learning Hub

Expected Value

Definition of Expected Value

- **Definition:**

  Let X be a discrete random variable. We say that the expected value of X exists

  $$\text{if } \sum_i |x_i| p(x_i) < \infty,$$

  in which case the expected value is defined as:

  $$\mu = E(X) = \sum_i x_i p(x_i)$$

- Also known as the expectation or the **mean** of X.

26

Properties of Expectations

- If there exists a finite constant c such that $P(X=c)=1$, then $E(X) = c$.
- If X and Y are random variables defined on the same sample space Ω with finite expectations, and if $P(X \leq Y) = 1$, then $E(X) \leq E(Y)$.
- If X has a finite expectation, and if $P(X \geq c) = 1$, then $E(X) \geq c$. If $P(X \leq c) = 1$,
- then $E(X) \leq c$.
- Linearity of Expectations: $E(cX) = cE(X)$ and $E(X1 + X2) =E(X1) + E(X2)$.
- Suppose X1, X2, X3,...Xk are independent random variables, then provided each expectation exists, $E(X1, X2,..Xk) = E(X2)E(X2)..E(Xk)$.

Example of Expectations

- **Example:** Let X be the sum of the two rolls when a fair die is rolled twice?
  - $p(2) = p(12) = \frac{1}{36}$ $\quad$ $p(3)=p(11)=\frac{2}{36}$
  - $p(4) = p(10) = \frac{3}{36}$ $\quad$ $p(5)=p(9)= \frac{4}{36}$
  - $p(6) = p(8) = \frac{5}{36}$ $\quad$ $p(7)=\frac{6}{36}$

$$E(X) = (2 * \frac{1}{36}) + (3 * \frac{2}{36}) + (4 * \frac{3}{36}) + \ldots (12 * \frac{1}{36}) = 7$$

- **Note:** What if we roll the die ten times? We can use the linearity of expectations then. Let Xi = the face obtained on the ith roll. Then E(X) = E(X1+X2+...+X10) = E(X1) + E(X2) +..+ E(X10)

Variance

Definition of Variance

- **Definition:**

Let a random variable X have a finite mean μ.

The variance of X is defined as : $\sigma^2 = E[(X - \mu)^2]$,

and the

Standard deviation of X is defined as $\sigma = \sqrt{\sigma^2}$

- Variance measures average squared deviation from the expected value.

30

Properties of Variance

- $var(cX) = c^2X\ for\ any\ ral\ c$
- $var(X + k) = var(X) for\ any\ real\ k$
- $var\ (X) \geq 0\ for\ any\ random\ variable\ X\ and\ equals\ zero\ only\ if$
- $P(X = c) = 1\ for\ some\ real\ constant\ c.$
- $var\ (X) = E(X^2) - \mu^2$

31

Example of Variance

- **Example:** Consider the experiment of two tosses of a fair coin and let X be the number of heads obtained.

  - $p(0) = p(2) = \frac{1}{4}$ *and* $p(1) = \frac{1}{2}$
  - $E(x^2) = (0 \ * \ \frac{1}{4}) + (1 * \frac{1}{2}) + (4 * \frac{1}{4}) = \frac{3}{2} \quad and \ E(X) = 1$
  - $var(X) = E(X^2) - \mu^2 = (\frac{3}{2} - 1) = \frac{1}{2}$
  - $standard \ deviation \ \sigma = \sqrt{0.5} = 0.707$

References

1. Javed, O.A. (n.d.). Probability and Statistics for Engineers and Scientist - 9th Edition (by Walpole, Mayers, Ye).
2. What is a Unimodal Distribution? (Definition & Example) (statology.org)

THANK YOU