Statistics for Data Science -1

Continuous Random Variables

Usha Mohan

Indian Institute of Technology Madras

Learning objectives

Learning objectives

1. Define what is a continuous random variable.

Learning objectives

1. Define what is a continuous random variable.
2. Probability distribution function and examples

Learning objectives

1. Define what is a continuous random variable.
2. Probability distribution function and examples
3. Cumulative distribution function, graphs, and examples.

Learning objectives

1. Define what is a continuous random variable.
2. Probability distribution function and examples
3. Cumulative distribution function, graphs, and examples.
4. Expectation and variance of random variables.

Probability density function, graph, and examples

Probability density function

Discrete and Continuous random variables

Discrete and Continuous random variables

Definition

A random variable that can take on at most a countable number of possible values is said to be a discrete random variable.

Discrete and Continuous random variables

Definition

A random variable that can take on at most a countable number of possible values is said to be a discrete random variable.

Definition

When outcomes for random event are numerical, but cannot be counted and are infinitely divisible, we have continuous random variables.

Discrete and continuous random variable

Discrete and continuous random variable

- A discrete random variable is one that has possible values that are discrete points along the real number line.

Discrete and continuous random variable

- A discrete random variable is one that has possible values that are discrete points along the real number line.
- A continuous random variable is one that has possible values that form an interval along the real number line. In other words, a continuous random variable can assume any value over an interval or intervals.

Probability density function (pdf)

- Every continuous random variable X has a curve associated with it.

Probability density function (pdf)

- Every continuous random variable X has a curve associated with it.
- The probability distribution curve of a continuous random variable is also called its probability density function. It is denoted by $f(x)$

Area under a pdf

- Consider any two points a and b, where a is less than b.
- The probability that X assumes a value that lies between a and b is equal to the area under the curve between a and b. That is,

$P(X \in [a, b]) = P(a \leq X \leq b)$ is area under curve between a and b

$$P(a \leq X \leq b) = \int_a^b f(x)dx$$

Properties of pdf

Properties of pdf

1. The area under the probability distribution curve of a continuous random variable between any two points is between 0 and 1.

Properties of pdf

Properties of pdf

2. Total area under the probability distribution curve of a continuous random variable is always 1.

Properties of pdf

- The area under the graph of the probability density function between points a and b is the same regardless of whether the endpoints a and b are themselves included:

$$P(a \leq X \leq b) = P(a < X < b)$$

Properties of pdf

- The area under the graph of the probability density function between points a and b is the same regardless of whether the endpoints a and b are themselves included:

$$P(a \leq X \leq b) = P(a < X < b)$$

- The probability density curve of a random variable X is a curve that never goes below the $x-$ axis

Example

Figure below is a probability density function for the random variable that represents the time (in minutes) it takes a repairer to service a television. The numbers in the regions represent the areas of those regions.

What is the probability that the repairer takes

1. Less than 20

Example

Figure below is a probability density function for the random variable that represents the time (in minutes) it takes a repairer to service a television. The numbers in the regions represent the areas of those regions.

What is the probability that the repairer takes

1. Less than 20 =0.29

Example

Figure below is a probability density function for the random variable that represents the time (in minutes) it takes a repairer to service a television. The numbers in the regions represent the areas of those regions.

What is the probability that the repairer takes

1. Less than 20 =0.29
2. Less than 40

Example

Figure below is a probability density function for the random variable that represents the time (in minutes) it takes a repairer to service a television. The numbers in the regions represent the areas of those regions.

What is the probability that the repairer takes

1. Less than 20 =0.29
2. Less than 40 =0.56

Example

Figure below is a probability density function for the random variable that represents the time (in minutes) it takes a repairer to service a television. The numbers in the regions represent the areas of those regions.

What is the probability that the repairer takes

1. Less than 20 $=0.29$
2. Less than 40 $=0.56$
3. More than 50

Example

Figure below is a probability density function for the random variable that represents the time (in minutes) it takes a repairer to service a television. The numbers in the regions represent the areas of those regions.

What is the probability that the repairer takes

1. Less than 20 =0.29
2. Less than 40 =0.56
3. More than 50 =0.33

Example

Figure below is a probability density function for the random variable that represents the time (in minutes) it takes a repairer to service a television. The numbers in the regions represent the areas of those regions.

What is the probability that the repairer takes

1. Less than 20 =0.29
2. Less than 40 =0.56
3. More than 50 =0.33
4. Between 40 and 70 minutes to complete a repair?

Example

Figure below is a probability density function for the random variable that represents the time (in minutes) it takes a repairer to service a television. The numbers in the regions represent the areas of those regions.

What is the probability that the repairer takes

1. Less than 20 $=0.29$
2. Less than 40 $=0.56$
3. More than 50 $=0.33$
4. Between 40 and 70 minutes to complete a repair? $=0.27$

Cumulative distribution function

For a continuous random variable X

$$F(a) = P(X \leq a) = \int_{-\infty}^{a} f(x)dx$$

Since the probability that a continuous random variable X assumes a single value is always zero, we have

$$P(X < a) = P(X \leq a) = \int_{-\infty}^{a} f(x)dx$$

Expectation and Variance

- Expected value: $E(X) = \int x\ f(x)dx$.
- Variance: $Var(X) = \int (x - E(X))^2\ f(x)dx$

Section summary

- Probability density function and its properties.
- cdf, expectation, and variance of continuous random variables.

Introduction

- A random variable is said to be uniformly distributed over the interval $[0, 1]$ if its probability density function is given by
  $$f(x) = \begin{cases} 1 & 0 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}$$