Probability and Statistics

Topic 14 - Properties of the Normal Distribution

Aamir Alaud Din, PhD

November 15, 2024

TABLE OF CONTENTS

TABLE OF CONTENTS

RECAP

- For a binomial probability distribution, there are two mutually exclusive outcomes which are referred to as success and failure.
- Although the success and failure are mutually exclusive, but their probabilities are different.
- The probability of obtaining x successes in n independent trials of a binomial experiment is given by

$$P(x) = {}_nC_r p^x (1-p)^{n-x} \qquad x = 0, 1, 2, \ldots, n$$

  where p is the probability of success.
- A binomial experiment with n independent trials and probability of success p has a mean and standard deviation given by the formulas

$$\mu_X = np \qquad \text{and} \qquad \sigma_X = \sqrt{np(1-p)}$$

TABLE OF CONTENTS

OBJECTIVES

After learning this topic and studying, you should be able to:

1. Use the uniform probability distribution
2. Graph a normal curve
3. State the properties of the normal curve
4. Explain the role of area in the normal density function

TABLE OF CONTENTS

THE WHY SECTION

-

TABLE OF CONTENTS

USE THE UNIFORM PROBABILITY DISTRIBUTION

- First, we discuss a uniform distribution to see the relation between area and probability.

EXAMPLE 1

Imagine that a friend of yours is always late. Let the random variable X represent the time from when you are supposed to meet your friend until he shows up. Suppose your friend could be on time ($x = 0$) or up to 30 minutes late ($x = 30$), with all intervals of equal time between $x = 0$ and $x = 30$ being equally likely. For example, your friend is just as likely to be 3–4 minutes late as he is to be 25–26 minutes late. The random variable X can be any value in the interval from 0 to 30, that is, $0 \leq x \leq 30$. Because any two intervals of equal length between 0 and 30, inclusive, are equally likely, the random variable X is said to follow a **uniform probability distribution**. •

USE THE UNIFORM PROBABILITY DISTRIBUTION

- When computing probabilities for discrete random variables, we usually substitute the value of the random variable into a formula.
- Things are not as easy for continuous random variables.
- Because an infinite number of outcomes are possible for continuous random variables, the probability of observing one particular value is zero.
- For example, the probability that your friend is exactly 12.9438823 minutes late is zero.
- This result is based on the fact that classical probability is found by dividing the number of ways an event can occur by the total number of possibilities: there is one way to observe 12.9438823, and there are an infinite number of possible values between 0 and 30.
- To resolve this problem, we compute probabilities of continuous random variables over an interval of values.

USE THE UNIFORM PROBABILITY DISTRIBUTION

- For example, we might compute the probability that your friend is between 10 and 15 minutes late.
- To find probabilities for continuous random variables, we use probability density functions.

Probability Density Function

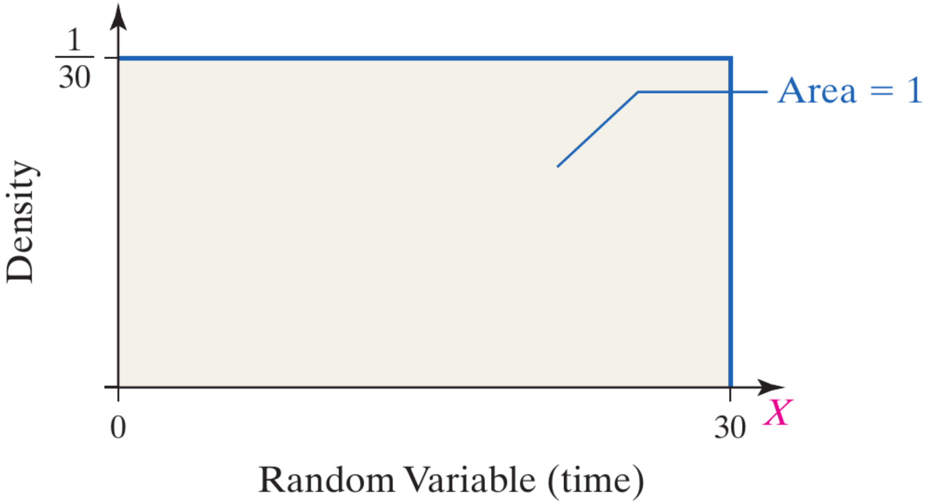
A probability density function (pdf) is an equation used to compute probabilities of continuous random variables. It must satisfy the following two properties:

1. The total area under the graph of the equation over all possible values of the random variable must equal 1.
2. The height of the graph of the equation must be greater than or equal to 0 for all possible values of the random variable.

USE THE UNIFORM PROBABILITY DISTRIBUTION

- Property 1 is similar to the rule for discrete probability distributions that stated the sum of the probabilities must add up to 1.
- Property 2 is similar to the rule that stated all probabilities must be greater than or equal to 0.
- Figure 1 illustrates these properties for Example 1.
- Since any value of the random variable between 0 and 30 is equally likely, the graph of the probability density function is a rectangle.
- Because the random variable is any number between 0 and 30 inclusive, the width of the rectangle is 30.

USE THE UNIFORM PROBABILITY DISTRIBUTION



USE THE UNIFORM PROBABILITY DISTRIBUTION

- Since the area under the graph of the probability density function must equal 1, and the area of a rectangle equals height times width, the height of the rectangle must be $\frac{1}{3}$.
- A pressing question remains: How do we use density functions to find probabilities of continuous random variables?
- The area under the graph of a density function over an interval represents the probability of observing a value of the random variable in that interval.
- The following example illustrates this statement.

EXAMPLE 2

Refer to the situation in Example 1.

(a) What is the probability your friend will be between 10 and 20 minutes late?

USE THE UNIFORM PROBABILITY DISTRIBUTION

EXAMPLE 2 (CONTINUED)

(b) It is 10 A.M. There is a 20% probability your friend will arrive within the next ______ minutes.

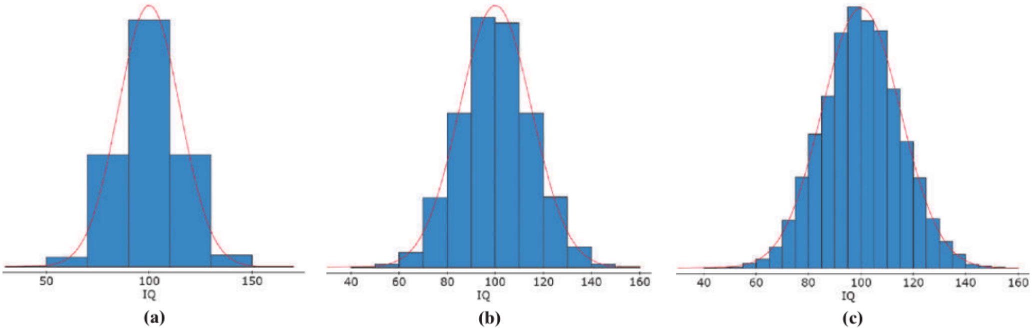
- We introduced the uniform density function so we could associate probability with area.
- We are now better prepared to discuss the most frequently used continuous distribution, the normal distribution.

TABLE OF CONTENTS

GRAPHING A NORMAL CURVE

- A rectangle is used to find the probability of observing an interval of numbers (such as 10–20 minutes after 10 a.m.) for a uniform random variable.
- However, not all continuous random variables follow a uniform distribution.
- For example, continuous random variables such as IQ scores and birth weights of babies have distributions that are symmetric and bell-shaped.
- Consider the histograms in Figure 3, which represent the IQ scores of 10,000 randomly selected adults.
- Notice that as the class width of the histogram decreases, the histogram becomes closely approximated by the smooth red curve.
- For this reason, we can use the curve to model the probability distribution of this continuous random variable.

GRAPHING A NORMAL CURVE



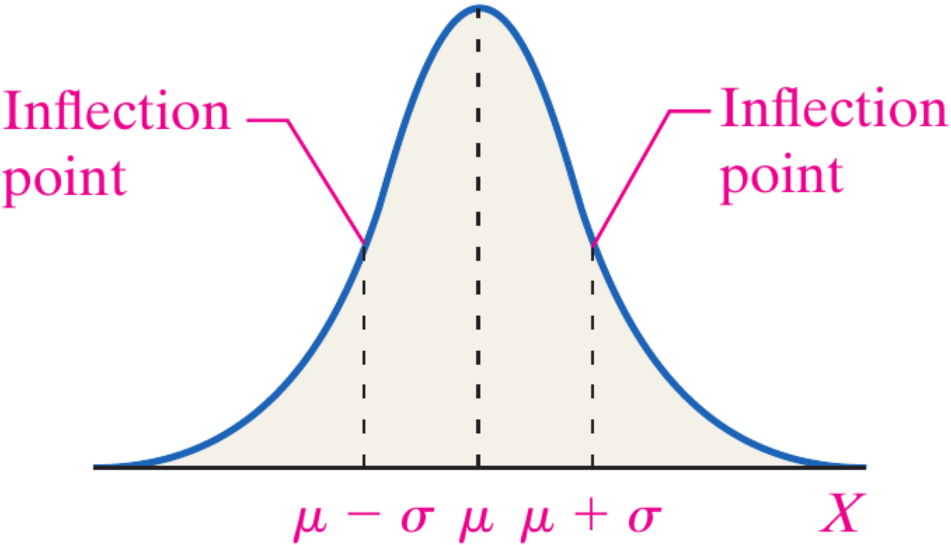
- In mathematics, a model is an equation, table, or graph used to describe reality.
- The red curve in Figure 3 is a model called the normal curve, which is used to describe continuous random variables that are said to be normally distributed.

GRAPHING A NORMAL CURVE

Normal Probability Distribution

A continuous random variable is normally distributed, or has a normal probability distribution, if its relative frequency histogram has the shape of a normal curve.

- Figure below shows a normal curve, demonstrating the roles that the mean m and standard deviation s play in drawing the curve.
- The mode represents the "high point" of the graph of any distribution.
- The median represents the point where 50% of the area under the distribution is to the left and 50% is to the right.
- The mean represents the balancing point of the graph of the distribution (see Figure 2 on page 148 in Section 3.1).


Inflection point
Inflection point
$\mu - \sigma$
μ
$\mu + \sigma$
X

GRAPHING A NORMAL CURVE

- For symmetric distributions with a single peak, such as the normal distribution, the mean $=$ median $=$ mode.
- Because of this, the mean, m, corresponds to the high point of the graph of the distribution.
- The points at $x = \mu - \sigma$ and $x = \mu + \sigma$ are the inflection points on the normal curve, the points on the curve where the curvature of the graph changes.
- To the left of $x = \mu - \sigma$ and to the right of $x = \mu + \sigma$, the curve is drawn upward.
- Between $x = \mu - \sigma$ and $x = \mu + \sigma$, the curve is drawn downward.
- Figure below shows how changes in μ and σ change the position or shape of a normal curve.

GRAPHING A NORMAL CURVE

- In Figure (a), one density curve has $\mu = 0$, $\sigma = 1$ and the other has $\mu = 3$, $\sigma = 1$.
- We can see that increasing the mean from 0 to 3 caused the graph to shift three units to the right but maintained its shape.
- In Figure (b), one density curve has $\mu = 0$, $\sigma = 1$ and the other has $\mu = 0$, $\sigma = 2$.
- We can see that increasing the standard deviation from 1 to 2 caused the graph to become flatter and more spread out but maintained its location of center.
- Notice that the points inflections are always at first standard deviation about the mean.
- Whatever the of standard deviation is doesn't change the point of inflection which is the beauty of normal distribution curve.

TABLE OF CONTENTS

PROPERTIES OF THE NORMAL CURVE

- The normal probability density function satisfies all the requirements of probability distributions.
- We list the properties of the normal density curve next.

Properties of the Normal Curve

1. The normal curve is symmetric about its mean, μ.
2. Because mean $=$ median $=$ mode, the normal curve has a single peak and the highest point occurs at $x = \mu$.
3. The normal curve has inflection points at $x = \mu - \sigma$ and $x = \mu + \sigma$.
4. The area under the normal curve is 1.

PROPERTIES OF THE NORMAL CURVE

Properties of the Normal Curve (Continued ...)

5. The area under the normal curve to the right of m equals the area under the curve to the left of μ, which equals $\frac{1}{2}$.
6. As x increases without bound (gets larger and larger), the graph approaches, but never reaches, the horizontal axis. As x decreases without bound (gets more and more negative), the graph approaches, but never reaches, the horizontal axis.
7. The Empirical Rule:
    - Approximately 68% of the area under the normal curve is between $x = \mu - \sigma$ and $x = \mu + \sigma$;
    - approximately 95% of the area under the normal curve is between $x = \mu - 2\sigma$ and $x = \mu + 2\sigma$;
    - approximately 99.7% of the area under the normal curve is between $x = \mu - 3\sigma$ and $x = \mu + 3\sigma$.

PROPERTIES OF THE NORMAL CURVE

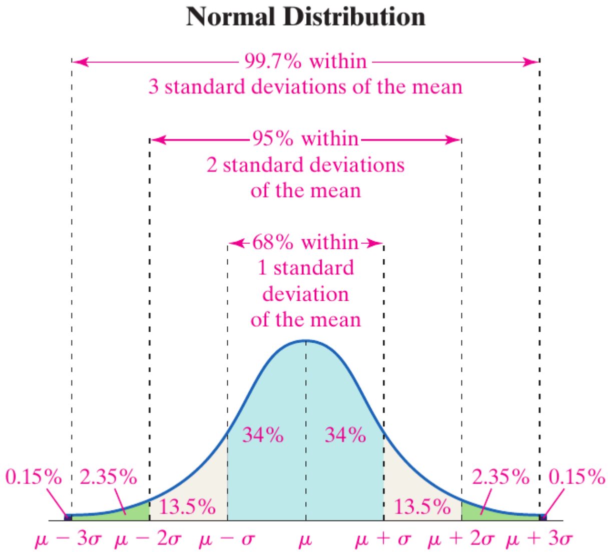


TABLE OF CONTENTS

- Let's look at an example of a normally distributed random variable.

EXAMPLE 3

The relative frequency distribution given in Table 1 represents the heights of a pediatrician's three-year-old female patients. The raw data indicate that the mean height of the patients is $\mu = 38.72$ inches with standard deviation $\sigma = 3.17$ inches.

(a) Draw a relative frequency histogram of the data. Comment on the shape of the distribution.

(b) Draw a normal curve with $\mu = 38.72$ inches and $\sigma = 3.17$ inches on the relative frequency histogram. Compare the area of the rectangle for heights between 40 and 40.9 inches to the area under the normal curve for heights between 40 and 40.9 inches.

ROLE OF AREA IN NORMAL DENSITY FUNCTION

Table 1

Height (inches)	Relative Frequency
29.0–29.9	0.005
30.0–30.9	0.005
31.0–31.9	0.005
32.0–32.9	0.025
33.0–33.9	0.02
34.0–34.9	0.055
35.0–35.9	0.075
36.0–36.9	0.09
37.0–37.9	0.115

Table 1

Height (inches)	Relative Frequency
38.0–38.9	0.15
39.0–39.9	0.12
40.0–40.9	0.11
41.0–41.9	0.07
42.0–42.9	0.06
43.0–43.9	0.035
44.0–44.9	0.025
45.0–45.9	0.025
46.0–46.9	0.005
47.0–47.9	0.005

ROLE OF AREA IN NORMAL DENSITY FUNCTION

- The equation (or model) used to determine the probability of a continuous random variable is called a probability density function (or pdf).
- The normal probability density function is given by

$$y = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}} = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

  where μ is the mean and σ is the standard deviation of the normal random variable.
- Do not feel threatened by this equation, because we will not be using it in this text.
- Instead, we will use the normal distribution in graphical form by drawing the normal curve.
- We now summarize the role area plays in the normal curve.

Area Under a Normal Curve

Suppose that a random variable X is normally distributed with mean μ and standard deviation σ. The area under the normal curve for any interval of values of the random variable X represents either

- the proportion of the population with the characteristic described by the interval of values or
- the probability that a randomly selected individual from the population will have the characteristic described by the interval of values.

ROLE OF AREA IN NORMAL DENSITY FUNCTION

Area Under a Normal Curve

The serum total cholesterol for males 20–29 years old is approximately normally distributed with mean $\mu = 180$ and $\sigma = 36.2$, based on data obtained from the National Health and Nutrition Examination Survey.

(a) Draw a normal curve with the parameters labeled.

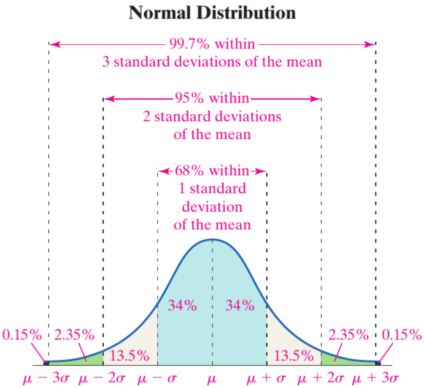
(b) An individual with total cholesterol greater than 200 is considered to have high cholesterol. Shade the region under the normal curve to the right of $x = 200$.

(c) Suppose that the area under the normal curve to the right of $x = 200$ is 0.2903. (You will learn how to find this area in Section 7.2.) Provide two interpretations of this result.

TABLE OF CONTENTS

SUMMARY

- Area under a normal curve is equal to 1.
- In a uniform distribution, the point of inflection is the first standard deviation.




Thank You!