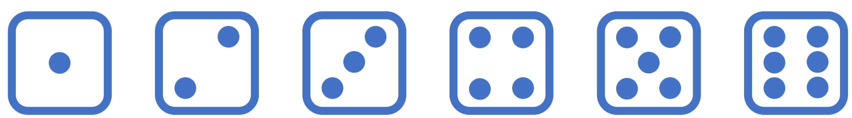
Preliminaries: Probability

CS 418. Introduction to Data Science

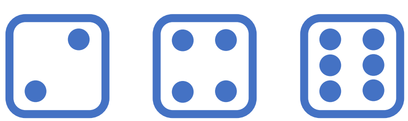
© 2018 by Gonzalo A. Bello

Preliminaries
Terminology

- An **experiment** is a procedure that results in one of a number of possible **outcomes**.
  - ***Example*:** rolling a dice.
- The **sample space** of the experiment is the set of all possible **outcomes**.
  - ***Example*:** $\{1, 2, 3, 4, 5, 6\}$



- An **event** is a subset of the sample space.
  - ***Example*:** $\{2, 4, 6\}$



Preliminaries
Probability of an Event

- **Probability** is the measure of the **likelihood** of an event occurring.

> If S is a finite nonempty sample space of **equally likely outcomes**, and $\boldsymbol{E \subseteq S}$ is an **event**, then the **probability** of $\boldsymbol{E}$ is
>
> $$\boldsymbol{P(E) = \frac{|E|}{|S|}}$$

- The **probability** of an event is **between 0 and 1**.
- ***Example*:** Two dice are rolled.

  What is the probability that the numbers on the two dice are the same?

  $\boldsymbol{P(E) =} \frac{6}{36} = \frac{1}{6}$

Preliminaries
Exercise 3.1

THINK PAIR SHARE

What is the probability that a hand of 5 cards drawn from a standard deck of 52 cards contains a full house (3 cards of one rank and 2 of another rank)?

4♠	4♥	4♦	J♠	J♣

$$\frac{\binom{13}{1}\binom{4}{3}\binom{12}{1}\binom{4}{2}}{\binom{52}{5}}$$

Each card has a **rank** and a **suit**. There are **13 possible ranks**, **4 possible suits**, and **13 cards for each suit** (one for each rank).

Preliminaries
Complement of an Event

- The **complement** of an event E, denoted $\boldsymbol{E'}$, consists of all outcomes that are **not** in E.

Let E be an event in a sample space S.
The **probability of the complement** of E is

$$\boldsymbol{P(E') = 1 - P(E)}$$

- ***Example***: Two dice are rolled.

What is the probability that the numbers on the two dice are not the same?

$P(E') = 1 - \frac{1}{6} = \frac{5}{6}$

Preliminaries
Mutually Exclusive Events

- The **joint probability** of E and F, denoted $P(E \cap F)$, is the probability of **both** E and F occurring at the same time.
- Two events are **mutually exclusive** if they cannot occur at the same time.

The events E and F are **mutually exclusive** if and only if

- ***Example***: Two dice are rolled. E is the event that the numbers on the two dice are the same, F is the event that the number on the first die is even, and G is the event that the number on the first die is odd.

Are E and F mutually exclusive? No

Are F and G mutually exclusive? Yes

Preliminaries
Conditional Probability

- The **conditional probability** of E given F, denoted $\boldsymbol{P(E|F)}$, is the probability of E occurring **given that** F has already occurred.

Let $\boldsymbol{E}$ and $\boldsymbol{F}$ be events with $\boldsymbol{P(F) > 0}$. The **conditional probability of $\boldsymbol{E}$ given $\boldsymbol{F}$**, denoted $\boldsymbol{P(E|F)}$, is given by

$$\boldsymbol{P(E|F) = \frac{P(E \cap F)}{P(F)}}$$

- ***Example*:** Two dice are rolled.

What is the probability that the numbers on the two dice are the same <u>given that</u> the number on the first die is even?

$$\boldsymbol{P(E|F)} = \frac{3/36}{1/2} = \frac{1}{6}$$

Preliminaries
Exercise 3.2

THINK PAIR SHARE

What is the probability that a hand of 5 cards drawn from a standard deck of 52 cards contains a full house <u>given that</u> the first three cards drawn are the ace of hearts, the ace of diamonds, and the ace of spades?

A♥	A♦	A♠	?	?

$$\frac{\binom{12}{1}\binom{4}{2}}{\binom{49}{2}}$$

Conditional probability

$$\boxed{P(E|F) = \frac{P(E \cap F)}{P(F)}}$$

Each card has a **rank** and a **suit**. There are **13 possible ranks**, **4 possible suits**, and **13 cards for each suit** (one for each rank).

Preliminaries
Independent Events

- Two events are **independent** if knowing that one occurred does **not** affect the probability of the other.

The events E and F are **independent** if and only if

$$P(E|F) = P(E)$$

$$P(F|E) = P(F)$$

$$P(E \cap F) = P(E) \cdot P(F)$$

All these statements are **equivalent**

- ***Example*:** Two dice are rolled. E is the event that the numbers on the two dice are the same and F is the event that the number on the first die is even.

Are E and F independent? Yes

Preliminaries
Bayes' Theorem (I)

- Conditional probabilities can be correctly "reversed" using **Bayes' theorem**.

Let $\boldsymbol{A}$ and $\boldsymbol{B}$ be events with $\boldsymbol{p(A) \neq 0}$ and $\boldsymbol{p(B) \neq 0}$. Then

$$\boldsymbol{P(A|B) = \frac{P(B|A)\,P(A)}{P(B)} = \frac{P(B|A)\,P(A)}{P(B|A)\,P(A) + P(B|\neg A)\,P(\neg A)}}$$

- ***Example***: Suppose that 1 in 100,000 people has a rare disease. 99.0% of the people who have the disease test positive and 99.5% of the people who do not have the disease test negative. **What is the probability that a person who tests positive has the disease?**

$$\boldsymbol{P(disease|+) = \frac{P(+|disease)\,P(disease)}{P(+|disease)\,P(disease) + P(+|\neg disease)\,P(\neg disease)} \approx 0.002}$$

Preliminaries
Bayes' Theorem (II)

Let A and B be events with $p(A) \neq 0$ and $p(B) \neq 0$. Then

$$P(A|B) = \frac{P(B|A)\,P(A)}{P(B)} = \frac{P(B|A)\,P(A)}{P(B|A)\,P(A) + P(B|\neg A)\,P(\neg A)}$$

- ***Example***: Suppose that we have found that the word "free" occurs in 250 out of 2000 emails known to be spam and in 5 out of 1000 emails known not to be spam. Assume that it is equally likely that an incoming message is spam or not spam. **What is the probability that an incoming message containing the word "free" is spam?**

$$P(spam|free) = \frac{P(free|spam)\,P(spam)}{P(free|spam)\,P(spam) + P(free|\neg spam)\,P(\neg spam)} \approx 0.962$$

Preliminaries
Random Variables

- A **random variable** X is a function from the sample space of an experiment to the set of real numbers.
- The **range of a random variable** X is the set of all possible values of X.
- A **random variable** can be **discrete** (if its range is countable) or **continuous** (if its range is uncountable).
- ***Example***:
  - Let the **random variable** X be the **number of heads that come up** when a coin is flipped 3 times.

    What are the possible values of X?

    $\{0, 1, 2, 3\}$
  - Let the **random variable** Y be the **time used by a student** to complete a timed 60-minute exam.

    What are the possible values of Y?

    $[0, 60]$

Preliminaries
Discrete Random Variables

- The probability distribution of a **discrete** random variable $\boldsymbol{X}$ can be described by a **probability mass function** (**pmf**).
- A **probability mass function** $\boldsymbol{P(X)}$ assigns a probability to each possible value of $\boldsymbol{X}$.
  - Each probability is **between zero and one**, inclusive.
  - The **sum of the probabilities is one**.
- ***Example***:
  - Let the **random variable** $\boldsymbol{X}$ be the **number of heads that come up** when a coin is flipped 3 times.

    What is the probability mass function of X?

    $P(X = 0) = 1/8$

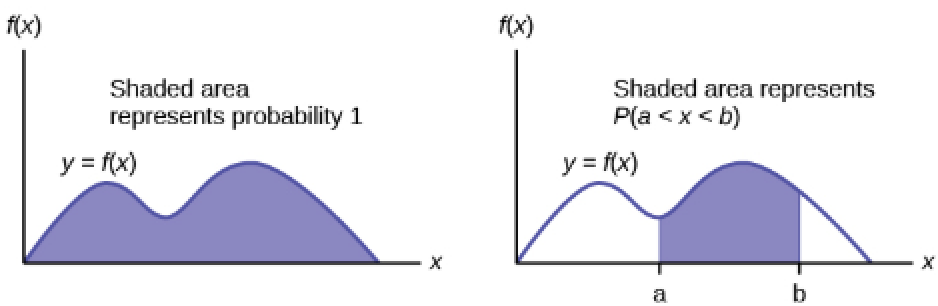
    $P(X = 1) = 3/8$

    $P(X = 2) = 3/8$

    $P(X = 3) = 1/8$

Preliminaries
Continuous Random Variables

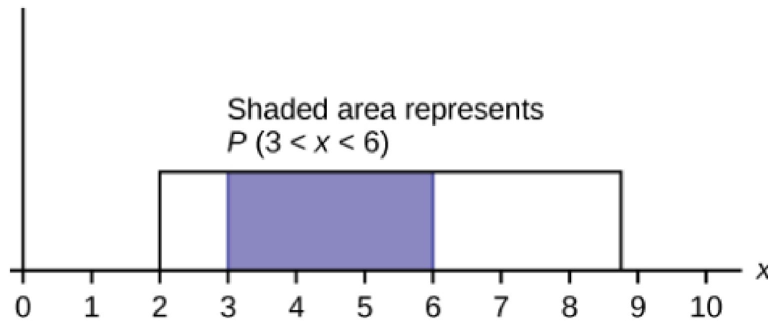
- The probability distribution of a **continuous** random variable can be described by a **probability density function** (**pdf**).
- A **probability mass function** $\boldsymbol{f(x)}$ describes the relative likelihood of all possible values of $\boldsymbol{X}$.
  - $\boldsymbol{f(x) \geq 0}$.
  - The **total area under the curve** $\boldsymbol{f(x)}$ **is one**.



Source: OpenStax, *Introductory Statistics* (2016)

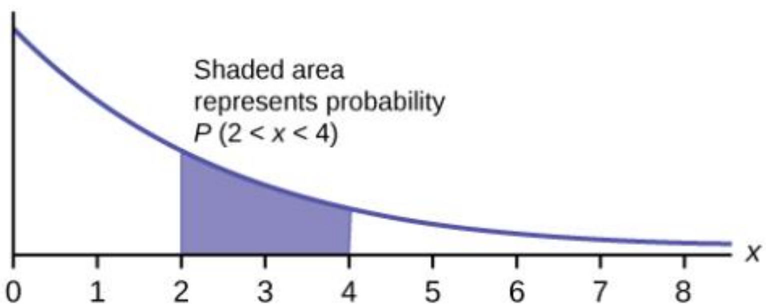
Preliminaries
Continuous Probability Distribution

- There are many **continuous probability distributions**.

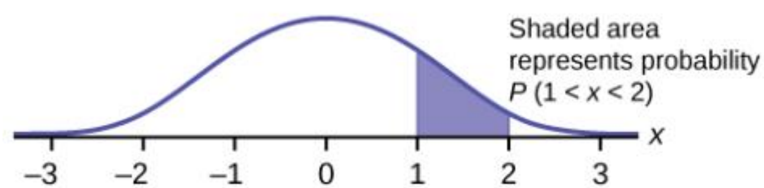


The uniform distribution

The **uniform distribution** is concerned with **outcomes that are equally likely to occur**



The exponential distribution

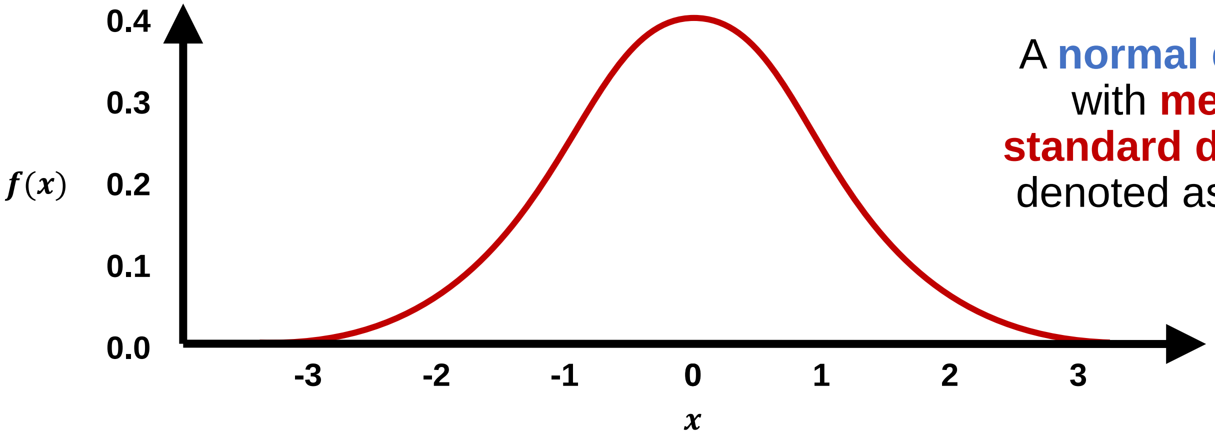


The normal distribution

Source: OpenStax, *Introductory Statistics* (2016)

Preliminaries
Normal Distribution

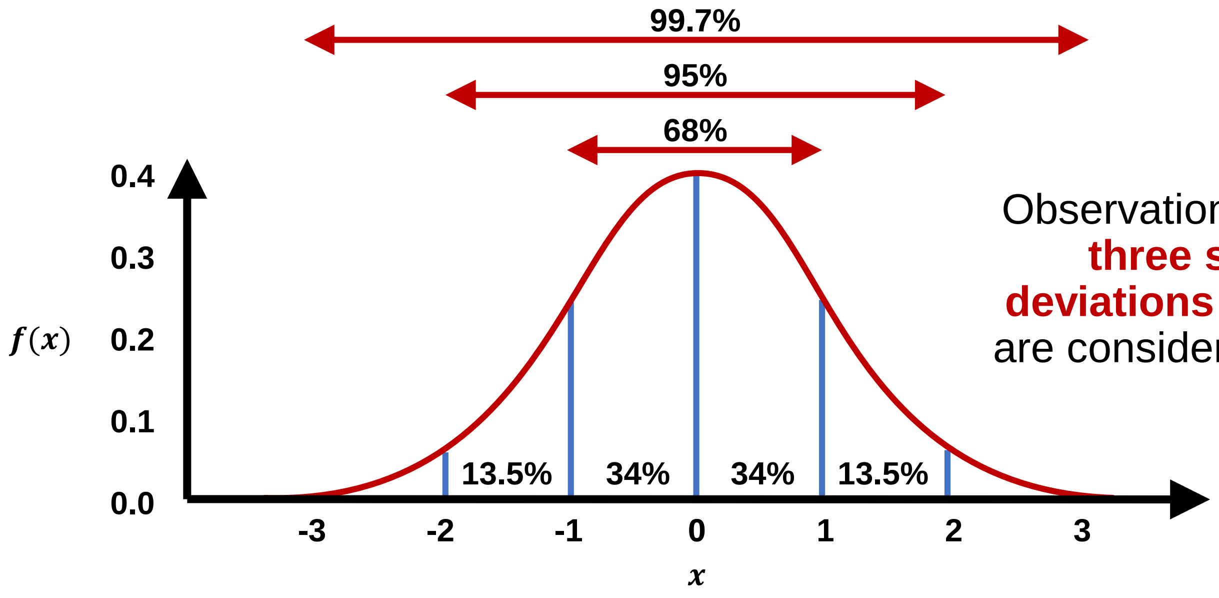
- The **normal** (**Gaussian**) **distribution** is a **continuous** distribution characterized by a **symmetric**, **unimodal**, **bell-shaped** curve.
- The **normal distribution** has two **parameters**: the **mean** μ and the **standard deviation** σ.
- Many processes can be approximated by the **normal distribution**, including exam scores and heights or other physical measurements.



A **normal distribution** with **mean** μ and **standard deviation** σ is denoted as $X \sim N(\mu, \sigma)$.

Normal Distribution
The Empirical Rule

- The **empirical rule** (also known as the **68-95-99.7 rule**) states that, for any **normal distribution**:
  - **68%** of the data falls within **one standard deviation** of the mean.
  - **95%** of the data falls within **two standard deviations** of the mean.
  - **99.7%** of the data falls within **three standard deviations** of the mean.



Observations **more than three standard deviations of the mean** are considered **very rare**.

Preliminaries
References

- David Diez, Christopher Barr, and Mine Çetinkaya-Rundel. *OpenIntro Statistics* (2015).
- Joel Grus. *Data Science from Scratch* (2015).
- OpenStax. *Introductory Statistics* (2016).