Review of Probability

Dr. Patrick Toche

Textbook:

James H. Stock and Mark W. Watson, *Introduction to Econometrics*, 4th Edition, Pearson.

Other references:

Joshua D. Angrist and Jörn-Steffen Pischke, *Mostly Harmless Econometrics: An Empiricist's Companion*, 1st Edition, Princeton University Press.
Jeffrey M. Wooldridge, *Introductory Econometrics: A Modern Approach*, 7th Edition, Cengage Learning.

The textbook comes with online resources and study guides. Other references will be given from time to time.

Contents

In this lesson you will learn ...

- random variables and probabilities, probability distribution and probability density
- expected value, standard deviation, and variance of a random variable
- joint probability and conditional probability of two random variables
- random sampling and the central limit theorem

In this lesson you will learn …

- random variables and probabilities, probability distribution and probability density
- expected value, standard deviation, and variance of a random variable
- joint probability and conditional probability of two random variables
- random sampling and the central limit theorem

In this lesson you will learn ...

- random variables and probabilities, probability distribution and probability density
- expected value, standard deviation, and variance of a random variable
- joint probability and conditional probability of two random variables
- random sampling and the central limit theorem

In this lesson you will learn ...

- random variables and probabilities, probability distribution and probability density
- expected value, standard deviation, and variance of a random variable
- joint probability and conditional probability of two random variables
- random sampling and the central limit theorem

Definitions

Basic Definitions

- **Random Experiment**: A repeatable procedure that has a well-defined set of outcomes.
- **Outcomes**: The mutually exclusive potential results of a random process.
- **Sample Space**: The set $\mathcal{S}$ of the possible outcomes of an experiment.
- **Event**: A subset of the sample space, $\mathcal{E} \subseteq \mathcal{S}$.
- **Random variable**: function from set $\mathcal{S}$ to a real number.
- **Probability**: A mapping from all subsets of the sample space $\mathcal{S}$ to $[0, 1]$ with these properties:
  - $\Pr(\mathcal{S}) = 1$
  - $0 \le \Pr(\mathcal{E}) \le 1$ for all $\mathcal{E} \subseteq \mathcal{S}$
  - If $\mathcal{E}_1, \mathcal{E}_2, \ldots$ are disjoint events, then $\Pr(\mathcal{E}_1 \cup \mathcal{E}_2 \cup \ldots) = \Pr(\mathcal{E}_1) + \Pr(\mathcal{E}_2) + \ldots$
- The probability of an outcome is the long-run frequency that the outcome occurs.

Basic Definitions

- **Random Experiment**: A repeatable procedure that has a well-defined set of outcomes.
- **Outcomes**: The mutually exclusive potential results of a random process.
- **Sample Space**: The set $\mathcal{S}$ of the possible outcomes of an experiment.
- **Event**: A subset of the sample space, $\mathcal{E} \subseteq \mathcal{S}$.
- **Random variable**: function from set $\mathcal{S}$ to a real number.
- **Probability**: A mapping from all subsets of the sample space S to $[0, 1]$ with these properties:
  - $\Pr(\mathcal{S}) = 1$
  - $0 \leq \Pr(\mathcal{E}) \leq 1$ for all $\mathcal{E} \subseteq \mathcal{S}$
  - If $\mathcal{E}_1, \mathcal{E}_2, \ldots$ are disjoint events, then $\Pr(\mathcal{E}_1 \cup \mathcal{E}_2 \cup \ldots) = \Pr(\mathcal{E}_1) + \Pr(\mathcal{E}_2) + \ldots$
- The probability of an outcome is the long-run frequency that the outcome occurs.

Basic Definitions

- **Random Experiment**: A repeatable procedure that has a well-defined set of outcomes.
- **Outcomes**: The mutually exclusive potential results of a random process.
- **Sample Space**: The set $\mathcal{S}$ of the possible outcomes of an experiment.
- **Event**: A subset of the sample space, $\mathcal{E} \subseteq \mathcal{S}$.
- **Random variable**: function from set $\mathcal{S}$ to a real number.
- **Probability**: A mapping from all subsets of the sample space $\mathcal{S}$ to $[0, 1]$ with these properties:
- The probability of an outcome is the long-run frequency that the outcome occurs.

Basic Definitions

- **Random Experiment**: A repeatable procedure that has a well-defined set of outcomes.
- **Outcomes**: The mutually exclusive potential results of a random process.
- **Sample Space**: The set $\mathcal{S}$ of the possible outcomes of an experiment.
- **Event**: A subset of the sample space, $\mathcal{E} \subseteq \mathcal{S}$.
- **Random variable**: function from set $\mathcal{S}$ to a real number.
- **Probability**: A mapping from all subsets of the sample space $\mathcal{S}$ to $[0, 1]$ with these properties:
- The probability of an outcome is the long-run frequency that the outcome occurs.

Basic Definitions

- **Random Experiment**: A repeatable procedure that has a well-defined set of outcomes.
- **Outcomes**: The mutually exclusive potential results of a random process.
- **Sample Space**: The set $\mathcal{S}$ of the possible outcomes of an experiment.
- **Event**: A subset of the sample space, $\mathcal{E} \subseteq \mathcal{S}$.
- **Random variable**: function from set $\mathcal{S}$ to a real number.
- **Probability**: A mapping from all subsets of the sample space S to $[0, 1]$ with these properties:
  - $\Pr(S) = 1$
  - $0 \le \Pr(\mathcal{E}) \le 1$ for all $\mathcal{E} \subseteq S$
  - If $\mathcal{E}_1, \mathcal{E}_2, \ldots$ are disjoint events, then $\Pr(\mathcal{E}_1 \cup \mathcal{E}_2 \cup \ldots) = \Pr(\mathcal{E}_1) + \Pr(\mathcal{E}_2) + \ldots$
- The probability of an outcome is the long-run frequency that the outcome occurs.

Basic Definitions

- **Random Experiment**: A repeatable procedure that has a well-defined set of outcomes.
- **Outcomes**: The mutually exclusive potential results of a random process.
- **Sample Space**: The set $\mathcal{S}$ of the possible outcomes of an experiment.
- **Event**: A subset of the sample space, $\mathcal{E} \subseteq \mathcal{S}$.
- **Random variable**: function from set $\mathcal{S}$ to a real number.
- **Probability**: A mapping from all subsets of the sample space S to $[0, 1]$ with these properties:
  - $\Pr(\mathcal{S}) = 1$
  - $0 \leq \Pr(\mathcal{E}) \leq 1$ for all $\mathcal{E} \subseteq \mathcal{S}$
  - If $\mathcal{E}_1, \mathcal{E}_2, \ldots$ are disjoint events, then $\Pr(\mathcal{E}_1 \cup \mathcal{E}_2 \cup \ldots) = \Pr(\mathcal{E}_1) + \Pr(\mathcal{E}_2) + \ldots$
- The probability of an outcome is the long-run frequency that the outcome occurs.

Basic Definitions

- **Random Experiment**: A repeatable procedure that has a well-defined set of outcomes.
- **Outcomes**: The mutually exclusive potential results of a random process.
- **Sample Space**: The set $\mathcal{S}$ of the possible outcomes of an experiment.
- **Event**: A subset of the sample space, $\mathcal{E} \subseteq \mathcal{S}$.
- **Random variable**: function from set $\mathcal{S}$ to a real number.
- **Probability**: A mapping from all subsets of the sample space S to $[0, 1]$ with these properties:
  - $\Pr(\mathcal{S}) = 1$
  - $0 \leq \Pr(\mathcal{E}) \leq 1$ for all $\mathcal{E} \subseteq \mathcal{S}$
  - If $\mathcal{E}_1, \mathcal{E}_2, \ldots$ are disjoint events, then $\Pr(\mathcal{E}_1 \cup \mathcal{E}_2 \cup \ldots) = \Pr(\mathcal{E}_1) + \Pr(\mathcal{E}_2) + \ldots$
- The probability of an outcome is the long-run frequency that the outcome occurs.

Basic Definitions

- **Random Experiment**: A repeatable procedure that has a well-defined set of outcomes.
- **Outcomes**: The mutually exclusive potential results of a random process.
- **Sample Space**: The set $\mathcal{S}$ of the possible outcomes of an experiment.
- **Event**: A subset of the sample space, $\mathcal{E} \subseteq \mathcal{S}$.
- **Random variable**: function from set $\mathcal{S}$ to a real number.
- **Probability**: A mapping from all subsets of the sample space S to $[0, 1]$ with these properties:
  - $\Pr(\mathcal{S}) = 1$
  - $0 \leq \Pr(\mathcal{E}) \leq 1$ for all $\mathcal{E} \subseteq \mathcal{S}$
  - If $\mathcal{E}_1, \mathcal{E}_2, \ldots$ are disjoint events, then $\Pr(\mathcal{E}_1 \cup \mathcal{E}_2 \cup \ldots) = \Pr(\mathcal{E}_1) + \Pr(\mathcal{E}_2) + \ldots$
- The probability of an outcome is the long-run frequency that the outcome occurs.

Basic Definitions

- **Random Experiment**: A repeatable procedure that has a well-defined set of outcomes.
- **Outcomes**: The mutually exclusive potential results of a random process.
- **Sample Space**: The set $\mathcal{S}$ of the possible outcomes of an experiment.
- **Event**: A subset of the sample space, $\mathcal{E} \subseteq \mathcal{S}$.
- **Random variable**: function from set $\mathcal{S}$ to a real number.
- **Probability**: A mapping from all subsets of the sample space S to $[0, 1]$ with these properties:
  - $\Pr(\mathcal{S}) = 1$
  - $0 \leq \Pr(\mathcal{E}) \leq 1$ for all $\mathcal{E} \subseteq \mathcal{S}$
  - If $\mathcal{E}_1, \mathcal{E}_2, \ldots$ are disjoint events, then $\Pr(\mathcal{E}_1 \cup \mathcal{E}_2 \cup \ldots) = \Pr(\mathcal{E}_1) + \Pr(\mathcal{E}_2) + \ldots$
- The probability of an outcome is the long-run frequency that the outcome occurs.

Basic Definitions

- **Random Experiment**: A repeatable procedure that has a well-defined set of outcomes.
- **Outcomes**: The mutually exclusive potential results of a random process.
- **Sample Space**: The set $\mathcal{S}$ of the possible outcomes of an experiment.
- **Event**: A subset of the sample space, $\mathcal{E} \subseteq \mathcal{S}$.
- **Random variable**: function from set $\mathcal{S}$ to a real number.
- **Probability**: A mapping from all subsets of the sample space S to $[0, 1]$ with these properties:
  - $\Pr(\mathcal{S}) = 1$
  - $0 \leq \Pr(\mathcal{E}) \leq 1$ for all $\mathcal{E} \subseteq \mathcal{S}$
  - If $\mathcal{E}_1, \mathcal{E}_2, \ldots$ are disjoint events, then $\Pr(\mathcal{E}_1 \cup \mathcal{E}_2 \cup \ldots) = \Pr(\mathcal{E}_1) + \Pr(\mathcal{E}_2) + \ldots$
- The probability of an outcome is the long-run frequency that the outcome occurs.

Examples

- Example of a random experiment: Flipping a coin.
- Sample space: $S = \{H, T\}$.
- Equivalent representation:

$$X = \begin{cases} 1 & \text{if} \quad H \\ 0 & \text{if} \quad T \end{cases}$$

- The assignment is arbitrary, so another equivalent representation:

$$X = \begin{cases} 1 & \text{if} \quad T \\ 0 & \text{if} \quad H \end{cases}$$

- Other examples of random experiments: rolling a die and observing the face; rolling two dice and observing the sums; drawing colored balls from an urn; car registration numbers from passing cars.

Examples

- Example of a random experiment: Flipping a coin.
- Sample space: $\mathcal{S} = \{H, T\}$.
- Equivalent representation:

$$X = \begin{cases} 1 & \text{if} \quad H \\ 0 & \text{if} \quad T \end{cases}$$

- The assignment is arbitrary, so another equivalent representation:

$$X = \begin{cases} 1 & \text{if} \quad T \\ 0 & \text{if} \quad H \end{cases}$$

- Other examples of random experiments: rolling a die and observing the face; rolling two dice and observing the sums; drawing colored balls from an urn; car registration numbers from passing cars.

Examples

- Example of a random experiment: Flipping a coin.
- Sample space: $\mathcal{S} = \{H, T\}$.
- Equivalent representation:

$$X = \begin{cases} 1 & \text{if} \quad H \\ 0 & \text{if} \quad T \end{cases}$$

- The assignment is arbitrary, so another equivalent representation:

$$X = \begin{cases} 1 & \text{if} \quad T \\ 0 & \text{if} \quad H \end{cases}$$

- Other examples of random experiments: rolling a die and observing the face; rolling two dice and observing the sums; drawing colored balls from an urn; car registration numbers from passing cars.

Examples

- Example of a random experiment: Flipping a coin.
- Sample space: $\mathcal{S} = \{H, T\}$.
- Equivalent representation:

$$X = \begin{cases} 1 & \text{if} \quad H \\ 0 & \text{if} \quad T \end{cases}$$

- The assignment is arbitrary, so another equivalent representation:

$$X = \begin{cases} 1 & \text{if} \quad T \\ 0 & \text{if} \quad H \end{cases}$$

- Other examples of random experiments: rolling a die and observing the face; rolling two dice and observing the sums; drawing colored balls from an urn; car registration numbers from passing cars.

Examples

- Example of a random experiment: Flipping a coin.
- Sample space: $\mathcal{S} = \{H, T\}$.
- Equivalent representation:

$$X = \begin{cases} 1 & \text{if} \quad H \\ 0 & \text{if} \quad T \end{cases}$$

- The assignment is arbitrary, so another equivalent representation:

$$X = \begin{cases} 1 & \text{if} \quad T \\ 0 & \text{if} \quad H \end{cases}$$

- Other examples of random experiments: rolling a die and observing the face; rolling two dice and observing the sums; drawing colored balls from an urn; car registration numbers from passing cars.

Set Notation

- **Set:** A collection of objects. These objects are called **elements** of the set.
- A sample space is a set whose elements are the possible outcomes of an experiment.
- The **union** of set A and set B, written $A \cup B$, is the set of every element in either A or B.
- The **intersection** of set A and B, written $A \cap B$, is the set of every element that belongs to both A and B.
- The set with no elements, denoted $\emptyset$, is called the **empty set**.
- Two sets are **disjoint** if their intersection is empty.
- A is a subset of B, denoted $A \subseteq B$, if every element of A is also an element of B.
- If A is a strict subset of B, it is denoted $A \subset B$.

Set Notation

- **Set:** A collection of objects. These objects are called **elements** of the set.
- A sample space is a set whose elements are the possible outcomes of an experiment.
- The **union** of set A and set B, written $A \cup B$, is the set of every element in either A or B.
- The **intersection** of set A and B, written $A \cap B$, is the set of every element that belongs to both A and B.
- The set with no elements, denoted $\emptyset$, is called the **empty set**.
- Two sets are **disjoint** if their intersection is empty.
- A is a subset of B, denoted $A \subseteq B$, if every element of A is also an element of B.
- If A is a strict subset of B, it is denoted $A \subset B$.

Set Notation

- **Set:** A collection of objects. These objects are called **elements** of the set.
- A sample space is a set whose elements are the possible outcomes of an experiment.
- The **union** of set A and set B, written $A \cup B$, is the set of every element in either A or B.
- The **intersection** of set A and B, written $A \cap B$, is the set of every element that belongs to both A and B.
- The set with no elements, denoted $\emptyset$, is called the **empty set**.
- Two sets are **disjoint** if their intersection is empty.
- A is a subset of B, denoted $A \subseteq B$, if every element of A is also an element of B.
- If A is a strict subset of B, it is denoted $A \subset B$.

Set Notation

- **Set:** A collection of objects. These objects are called **elements** of the set.
- A sample space is a set whose elements are the possible outcomes of an experiment.
- The **union** of set A and set B, written $A \cup B$, is the set of every element in either A or B.
- The **intersection** of set A and B, written $A \cap B$, is the set of every element that belongs to both A and B.
- The set with no elements, denoted $\emptyset$, is called the **empty set**.
- Two sets are **disjoint** if their intersection is empty.
- A is a subset of B, denoted $A \subseteq B$, if every element of A is also an element of B.
- If A is a strict subset of B, it is denoted $A \subset B$.

Set Notation

- **Set:** A collection of objects. These objects are called **elements** of the set.
- A sample space is a set whose elements are the possible outcomes of an experiment.
- The **union** of set A and set B, written $A \cup B$, is the set of every element in either A or B.
- The **intersection** of set A and B, written $A \cap B$, is the set of every element that belongs to both A and B.
- The set with no elements, denoted $\emptyset$, is called the **empty set**.
- Two sets are **disjoint** if their intersection is empty.
- A is a subset of B, denoted $A \subseteq B$, if every element of A is also an element of B.
- If A is a strict subset of B, it is denoted $A \subset B$.

Set Notation

- **Set:** A collection of objects. These objects are called **elements** of the set.
- A sample space is a set whose elements are the possible outcomes of an experiment.
- The **union** of set A and set B, written $A \cup B$, is the set of every element in either A or B.
- The **intersection** of set A and B, written $A \cap B$, is the set of every element that belongs to both A and B.
- The set with no elements, denoted $\emptyset$, is called the **empty set**.
- Two sets are **disjoint** if their intersection is empty.
- A is a subset of B, denoted $A \subseteq B$, if every element of A is also an element of B.
- If A is a strict subset of B, it is denoted $A \subset B$.

Set Notation

- **Set:** A collection of objects. These objects are called **elements** of the set.
- A sample space is a set whose elements are the possible outcomes of an experiment.
- The **union** of set A and set B, written $A \cup B$, is the set of every element in either A or B.
- The **intersection** of set A and B, written $A \cap B$, is the set of every element that belongs to both A and B.
- The set with no elements, denoted $\emptyset$, is called the **empty set**.
- Two sets are **disjoint** if their intersection is empty.
- A is a subset of B, denoted $A \subseteq B$, if every element of A is also an element of B.
- If A is a strict subset of B, it is denoted $A \subset B$.

Set Notation

- **Set:** A collection of objects. These objects are called **elements** of the set.
- A sample space is a set whose elements are the possible outcomes of an experiment.
- The **union** of set A and set B, written $A \cup B$, is the set of every element in either A or B.
- The **intersection** of set A and B, written $A \cap B$, is the set of every element that belongs to both A and B.
- The set with no elements, denoted $\emptyset$, is called the **empty set**.
- Two sets are **disjoint** if their intersection is empty.
- A is a subset of B, denoted $A \subseteq B$, if every element of A is also an element of B.
- If A is a strict subset of B, it is denoted $A \subset B$.

Discrete Random variables

Discrete

Discrete random variables

- A **random variable** is a function from the sample space $\mathcal{S}$ to a real number.
- A random variable is **discrete** if it takes countably many distinct values, $X \in \{x_1, \ldots, x_n\}$, where the distance between values is not always meaningful.
- Example: Rolling a die, $X \in \{1, 2, 3, 4, 5, 6\}$, where X is the face value.
- **Probability Mass Function:** The **pmf** f of random variable X evaluated at x gives the probability that X equals the discrete value x,

$$p = f(x) = \Pr(X = x)$$

- **Cumulative Distribution Function:** The **cdf** F of random variable X evaluated at x gives the probability that X equals a value at least as large as x,

$$F(x) = \Pr(X \leq x) = \sum_{i=1}^{n} p_i 1\{x_i \leq x\}$$

where $1\{x_i \leq x\}$ is an indicator variable equal to 1 if the condition $x_i \leq x$ is satisfied; 0 otherwise.

Discrete random variables

- A **random variable** is a function from the sample space $\mathcal{S}$ to a real number.
- A random variable is **discrete** if it takes countably many distinct values, $X \in \{x_1, ..., x_n\}$, where the distance between values is not always meaningful.
- Example: Rolling a die, $X \in \{1, 2, 3, 4, 5, 6\}$, where X is the face value.
- **Probability Mass Function:** The **pmf** f of random variable X evaluated at x gives the probability that X equals the discrete value x,

$$p = f(x) = \Pr(X = x)$$

- **Cumulative Distribution Function:** The **cdf** F of random variable X evaluated at x gives the probability that X equals a value at least as large as x,

$$F(x) = \Pr(X \leq x) = \sum_{i=1}^{n} p_i 1\{x_i \leq x\}$$

where $1\{x_i \leq x\}$ is an indicator variable equal to 1 if the condition $x_i \leq x$ is satisfied; 0 otherwise.

Discrete random variables

- A **random variable** is a function from the sample space $\mathcal{S}$ to a real number.
- A random variable is **discrete** if it takes countably many distinct values, $X \in \{x_1, ..., x_n\}$, where the distance between values is not always meaningful.
- Example: Rolling a die, $X \in \{1, 2, 3, 4, 5, 6\}$, where X is the face value.
- **Probability Mass Function:** The **pmf** f of random variable X evaluated at x gives the probability that X equals the discrete value x,

$$p = f(x) = \Pr(X = x)$$

- **Cumulative Distribution Function:** The **cdf** F of random variable X evaluated at x gives the probability that X equals a value at least as large as x,

$$F(x) = \Pr(X \leq x) = \sum_{i=1}^{n} p_i 1\{x_i \leq x\}$$

where $1\{x_i \leq x\}$ is an indicator variable equal to 1 if the condition $x_i \leq x$ is satisfied; 0 otherwise.

Discrete random variables

- A **random variable** is a function from the sample space $\mathcal{S}$ to a real number.
- A random variable is **discrete** if it takes countably many distinct values, $X \in \{x_1, ..., x_n\}$, where the distance between values is not always meaningful.
- Example: Rolling a die, $X \in \{1, 2, 3, 4, 5, 6\}$, where X is the face value.
- **Probability Mass Function:** The **pmf** f of random variable X evaluated at x gives the probability that X equals the discrete value x,

$$p = f(x) = \Pr(X = x)$$

- **Cumulative Distribution Function:** The **cdf** F of random variable X evaluated at x gives the probability that X equals a value at least as large as x,

$$F(x) = \Pr(X \leq x) = \sum_{i=1}^{n} p_i 1\{x_i \leq x\}$$

where $1\{x_i \leq x\}$ is an indicator variable equal to 1 if the condition $x_i \leq x$ is satisfied; 0 otherwise.

Discrete random variables

- A **random variable** is a function from the sample space $\mathcal{S}$ to a real number.
- A random variable is **discrete** if it takes countably many distinct values, $X \in \{x_1, ..., x_n\}$, where the distance between values is not always meaningful.
- Example: Rolling a die, $X \in \{1, 2, 3, 4, 5, 6\}$, where X is the face value.
- **Probability Mass Function:** The **pmf** f of random variable X evaluated at x gives the probability that X equals the discrete value x,

$$p = f(x) = \Pr(X = x)$$

- **Cumulative Distribution Function**: The **cdf** F of random variable X evaluated at x gives the probability that X equals a value at least as large as x,

$$F(x) = \Pr(X \leq x) = \sum_{i=1}^{n} p_i 1\{x_i \leq x\}$$

where $1\{x_i \leq x\}$ is an indicator variable equal to 1 if the condition $x_i \leq x$ is satisfied; 0 otherwise.

Continuous

Continuous random variables

- A **continuous** random variable is one which takes an uncountably infinite number of possible values, each with vanishingly small probability, where the distance between values is usually meaningful.
- Examples: A real value taken between 0 and 1; a random point on a plane; measurement of lengths, weights, temperatures.
- The **cdf** of a continuous random variable is continuous and differentiable — its **pdf** may have jumps, but commonly used distributions can be represented by a continuous, differentiable probability density function.
- **Probability Density Function:** The derivative of the Cumulative Distribution function:

$$f(x) = \frac{dF}{dx}(x)$$
$$F(x) = \int_{-\infty}^{x} f(t)dt$$

- The lower bound of the support of the random variable could be strictly greater than $-\infty$ of course, a common case being the positive real numbers $\int_0^x f(t)dt$.

Continuous random variables

- A **continuous** random variable is one which takes an uncountably infinite number of possible values, each with vanishingly small probability, where the distance between values is usually meaningful.
- Examples: A real value taken between 0 and 1; a random point on a plane; measurement of lengths, weights, temperatures.
- The **cdf** of a continuous random variable is continuous and differentiable — its **pdf** may have jumps, but commonly used distributions can be represented by a continuous, differentiable probability density function.
- **Probability Density Function:** The derivative of the Cumulative Distribution function:

$$f(x) = \frac{dF}{dx}(x)$$

$$F(x) = \int_{-\infty}^{x} f(t)dt$$

- The lower bound of the support of the random variable could be strictly greater than $-\infty$ of course, a common case being the positive real numbers $\int_0^x f(t)dt$.

Continuous random variables

- A **continuous** random variable is one which takes an uncountably infinite number of possible values, each with vanishingly small probability, where the distance between values is usually meaningful.
- Examples: A real value taken between 0 and 1; a random point on a plane; measurement of lengths, weights, temperatures.
- The **cdf** of a continuous random variable is continuous and differentiable — its **pdf** may have jumps, but commonly used distributions can be represented by a continuous, differentiable probability density function.
- **Probability Density Function:** The derivative of the Cumulative Distribution function:

$$f(x) = \frac{dF}{dx}(x)$$

$$F(x) = \int_{-\infty}^{x} f(t)dt$$

- The lower bound of the support of the random variable could be strictly greater than $-\infty$ of course, a common case being the positive real numbers $\int_0^x f(t)dt$.

Continuous random variables

- A **continuous** random variable is one which takes an uncountably infinite number of possible values, each with vanishingly small probability, where the distance between values is usually meaningful.
- Examples: A real value taken between 0 and 1; a random point on a plane; measurement of lengths, weights, temperatures.
- The **cdf** of a continuous random variable is continuous and differentiable — its **pdf** may have jumps, but commonly used distributions can be represented by a continuous, differentiable probability density function.
- **Probability Density Function:** The derivative of the Cumulative Distribution function:

$$f(x) = \frac{dF}{dx}(x)$$
$$F(x) = \int_{-\infty}^{x} f(t)dt$$

- The lower bound of the support of the random variable could be strictly greater than $-\infty$ of course, a common case being the positive real numbers $\int_0^x f(t)dt$.

Continuous random variables

- A **continuous** random variable is one which takes an uncountably infinite number of possible values, each with vanishingly small probability, where the distance between values is usually meaningful.
- Examples: A real value taken between 0 and 1; a random point on a plane; measurement of lengths, weights, temperatures.
- The **cdf** of a continuous random variable is continuous and differentiable — its **pdf** may have jumps, but commonly used distributions can be represented by a continuous, differentiable probability density function.
- **Probability Density Function:** The derivative of the Cumulative Distribution function:

$$f(x) = \frac{dF}{dx}(x)$$
$$F(x) = \int_{-\infty}^{x} f(t)dt$$

- The lower bound of the support of the random variable could be strictly greater than $-\infty$ of course, a common case being the positive real numbers $\int_0^x f(t)dt$.

Probability

Conditional Probability

- **Joint Probability:** Probability of event A and event B occurring together,

$$\Pr(A \cap B) = \Pr(A|B) \cdot \Pr(B)$$

- **Conditional Probability:** Probability of event A given that event B has occurred,

$$\Pr(A|B) = \Pr(A \cap B) / \Pr(B)$$

- **Marginal Probability:** Unconditional Probability of event A irrespective of whether event B has occurred or not – "an absolute" determined in the universe.
- A is **independent** of B iff

$$\Pr(A \cap B) = \Pr(A) \cdot \Pr(B)$$

- A **independent** of B is sometimes denoted $A \perp\!\!\!\perp B$.
- $A \perp\!\!\!\perp B$ implies

$$\Pr(A|B) = \Pr(A) \text{ and } \Pr(B|A) = \Pr(B)$$

- If A and B are independent, knowing whether B occurred carries no information about A.

Conditional Probability

- **Joint Probability:** Probability of event A and event B occurring together,

$$\Pr(A \cap B) = \Pr(A|B) \cdot \Pr(B)$$

- **Conditional Probability:** Probability of event A given that event B has occurred,

$$\Pr(A|B) = \Pr(A \cap B) / \Pr(B)$$

- **Marginal Probability:** Unconditional Probability of event A irrespective of whether event B has occurred or not – "an absolute" determined in the universe.
- A is **independent** of B iff

$$\Pr(A \cap B) = \Pr(A) \cdot \Pr(B)$$

- A **independent** of B is sometimes denoted $A \perp\!\!\!\perp B$.
- $A \perp\!\!\!\perp B$ implies

$$\Pr(A|B) = \Pr(A) \text{ and } \Pr(B|A) = \Pr(B)$$

- If A and B are independent, knowing whether B occurred carries no information about A.

Conditional Probability

- **Joint Probability:** Probability of event A and event B occurring together,

$$\Pr(A \cap B) = \Pr(A|B) \cdot \Pr(B)$$

- **Conditional Probability:** Probability of event A given that event B has occurred,

$$\Pr(A|B) = \Pr(A \cap B) / \Pr(B)$$

- **Marginal Probability:** Unconditional Probability of event A irrespective of whether event B has occurred or not – "an absolute" determined in the universe.

- A is **independent** of B iff

$$\Pr(A \cap B) = \Pr(A) \cdot \Pr(B)$$

- A **independent** of B is sometimes denoted $A \perp\!\!\!\perp B$.
- $A \perp\!\!\!\perp B$ implies

$$\Pr(A|B) = \Pr(A) \text{ and } \Pr(B|A) = \Pr(B)$$

- If A and B are independent, knowing whether B occurred carries no information about A.

Conditional Probability

- **Joint Probability:** Probability of event A and event B occurring together,

$$\Pr(A \cap B) = \Pr(A|B) \cdot \Pr(B)$$

- **Conditional Probability:** Probability of event A given that event B has occurred,

$$\Pr(A|B) = \Pr(A \cap B) / \Pr(B)$$

- **Marginal Probability:** Unconditional Probability of event A irrespective of whether event B has occurred or not – "an absolute" determined in the universe.
- A is **independent** of B iff

$$\Pr(A \cap B) = \Pr(A) \cdot \Pr(B)$$

- A **independent** of B is sometimes denoted $A \perp\!\!\!\perp B$.
- $A \perp\!\!\!\perp B$ implies

$$\Pr(A|B) = \Pr(A) \text{ and } \Pr(B|A) = \Pr(B)$$

- If A and B are independent, knowing whether B occurred carries no information about A.

Conditional Probability

- **Joint Probability:** Probability of event A and event B occurring together,

$$\Pr(A \cap B) = \Pr(A|B) \cdot \Pr(B)$$

- **Conditional Probability:** Probability of event A given that event B has occurred,

$$\Pr(A|B) = \Pr(A \cap B) / \Pr(B)$$

- **Marginal Probability:** Unconditional Probability of event A irrespective of whether event B has occurred or not – "an absolute" determined in the universe.
- A is **independent** of B iff

$$\Pr(A \cap B) = \Pr(A) \cdot \Pr(B)$$

- A **independent** of B is sometimes denoted $A \perp\!\!\!\perp B$.
- $A \perp\!\!\!\perp B$ implies

$$\Pr(A|B) = \Pr(A) \text{ and } \Pr(B|A) = \Pr(B)$$

- If A and B are independent, knowing whether B occurred carries no information about A.

Conditional Probability

- **Joint Probability:** Probability of event A and event B occurring together,

$$\Pr(A \cap B) = \Pr(A|B) \cdot \Pr(B)$$

- **Conditional Probability:** Probability of event A given that event B has occurred,

$$\Pr(A|B) = \Pr(A \cap B) / \Pr(B)$$

- **Marginal Probability:** Unconditional Probability of event A irrespective of whether event B has occurred or not – "an absolute" determined in the universe.
- A is **independent** of B iff

$$\Pr(A \cap B) = \Pr(A) \cdot \Pr(B)$$

- A **independent** of B is sometimes denoted $A \perp\!\!\!\perp B$.
- $A \perp\!\!\!\perp B$ implies

$$\Pr(A|B) = \Pr(A) \text{ and } \Pr(B|A) = \Pr(B)$$

- If A and B are independent, knowing whether B occurred carries no information about A.

Conditional Probability

- **Joint Probability:** Probability of event A and event B occurring together,

$$\Pr(A \cap B) = \Pr(A|B) \cdot \Pr(B)$$

- **Conditional Probability:** Probability of event A given that event B has occurred,

$$\Pr(A|B) = \Pr(A \cap B) / \Pr(B)$$

- **Marginal Probability:** Unconditional Probability of event A irrespective of whether event B has occurred or not – "an absolute" determined in the universe.
- A is **independent** of B iff

$$\Pr(A \cap B) = \Pr(A) \cdot \Pr(B)$$

- A **independent** of B is sometimes denoted $A \perp\!\!\!\perp B$.
- $A \perp\!\!\!\perp B$ implies

$$\Pr(A|B) = \Pr(A) \text{ and } \Pr(B|A) = \Pr(B)$$

- If A and B are independent, knowing whether B occurred carries no information about A.

Bayes’ Rule

Bayes' Rule

- **Bayes' Rule:**

$$\boxed{\Pr(A|B) = \frac{\Pr(B|A)\cdot\Pr(A)}{\Pr(B)}}$$

- **Example:** Interpreting the results of screening tests. The test is not perfect — false positives and false negatives randomly occur.
- A positive test is therefore only a presumption of sickness, not an absolute certainty.
- Let $\Pr(\text{sick})$ be the unconditional probability of being sick. Bayes' rule gives the probability of being sick conditional on testing positive, $\Pr(\text{sick}|\text{positive})$:

$$\Pr(\text{sick}|\text{positive}) = \frac{\Pr(\text{positive}|\text{sick})\cdot\Pr(\text{sick})}{\Pr(\text{positive})}$$

- $\Pr(\text{sick}|\text{positive})$ is larger than $\Pr(\text{sick})$
- $\Pr(\text{sick}|\text{positive})$ and $\Pr(\text{positive}|\text{sick})$ are commonly confused, but they can be very different.

Bayes' Rule

- **Bayes' Rule:**

$$\boxed{\Pr(A|B) = \frac{\Pr(B|A) \cdot \Pr(A)}{\Pr(B)}}$$

- **Example:** Interpreting the results of screening tests. The test is not perfect — false positives and false negatives randomly occur.
- A positive test is therefore only a presumption of sickness, not an absolute certainty.
- Let $\Pr(\text{sick})$ be the unconditional probability of being sick. Bayes' rule gives the probability of being sick conditional on testing positive, $\Pr(\text{sick}|\text{positive})$:

$$\Pr(\text{sick}|\text{positive}) = \frac{\Pr(\text{positive}|\text{sick}) \cdot \Pr(\text{sick})}{\Pr(\text{positive})}$$

- $\Pr(\text{sick}|\text{positive})$ is larger than $\Pr(\text{sick})$
- $\Pr(\text{sick}|\text{positive})$ and $\Pr(\text{positive}|\text{sick})$ are commonly confused, but they can be very different.

Bayes' Rule

- **Bayes' Rule:**

$$\boxed{\Pr(A|B) = \frac{\Pr(B|A) \cdot \Pr(A)}{\Pr(B)}}$$

- **Example:** Interpreting the results of screening tests. The test is not perfect — false positives and false negatives randomly occur.
- A positive test is therefore only a presumption of sickness, not an absolute certainty.
- Let $\Pr(\text{sick})$ be the unconditional probability of being sick. Bayes' rule gives the probability of being sick conditional on testing positive, $\Pr(\text{sick}|\text{positive})$:

$$\Pr(\text{sick}|\text{positive}) = \frac{\Pr(\text{positive}|\text{sick}) \cdot \Pr(\text{sick})}{\Pr(\text{positive})}$$

- $\Pr(\text{sick}|\text{positive})$ is larger than $\Pr(\text{sick})$
- $\Pr(\text{sick}|\text{positive})$ and $\Pr(\text{positive}|\text{sick})$ are commonly confused, but they can be very different.

Bayes' Rule

- **Bayes' Rule:**

$$\boxed{\Pr(A|B) = \frac{\Pr(B|A) \cdot \Pr(A)}{\Pr(B)}}$$

- **Example:** Interpreting the results of screening tests. The test is not perfect — false positives and false negatives randomly occur.
- A positive test is therefore only a presumption of sickness, not an absolute certainty.
- Let $\Pr(\text{sick})$ be the unconditional probability of being sick. Bayes' rule gives the probability of being sick conditional on testing positive, $\Pr(\text{sick}|\text{positive})$:

$$\Pr(\text{sick}|\text{positive}) = \frac{\Pr(\text{positive}|\text{sick}) \cdot \Pr(\text{sick})}{\Pr(\text{positive})}$$

- $\Pr(\text{sick}|\text{positive})$ is larger than $\Pr(\text{sick})$
- $\Pr(\text{sick}|\text{positive})$ and $\Pr(\text{positive}|\text{sick})$ are commonly confused, but they can be very different.

Bayes' Rule

- **Bayes' Rule:**

$$\boxed{\Pr(A|B) = \frac{\Pr(B|A) \cdot \Pr(A)}{\Pr(B)}}$$

- **Example:** Interpreting the results of screening tests. The test is not perfect — false positives and false negatives randomly occur.
- A positive test is therefore only a presumption of sickness, not an absolute certainty.
- Let $\Pr(\text{sick})$ be the unconditional probability of being sick. Bayes' rule gives the probability of being sick conditional on testing positive, $\Pr(\text{sick}|\text{positive})$:

$$\Pr(\text{sick}|\text{positive}) = \frac{\Pr(\text{positive}|\text{sick}) \cdot \Pr(\text{sick})}{\Pr(\text{positive})}$$

- $\Pr(\text{sick}|\text{positive})$ is larger than $\Pr(\text{sick})$
- $\Pr(\text{sick}|\text{positive})$ and $\Pr(\text{positive}|\text{sick})$ are commonly confused, but they can be very different.

Bayes' Rule

- **Bayes' Rule:**

$$\boxed{\Pr(A|B) = \frac{\Pr(B|A) \cdot \Pr(A)}{\Pr(B)}}$$

- **Example:** Interpreting the results of screening tests. The test is not perfect — false positives and false negatives randomly occur.
- A positive test is therefore only a presumption of sickness, not an absolute certainty.
- Let $\Pr(\text{sick})$ be the unconditional probability of being sick. Bayes' rule gives the probability of being sick conditional on testing positive, $\Pr(\text{sick}|\text{positive})$:

$$\Pr(\text{sick}|\text{positive}) = \frac{\Pr(\text{positive}|\text{sick}) \cdot \Pr(\text{sick})}{\Pr(\text{positive})}$$

- $\Pr(\text{sick}|\text{positive})$ is larger than $\Pr(\text{sick})$
- $\Pr(\text{sick}|\text{positive})$ and $\Pr(\text{positive}|\text{sick})$ are commonly confused, but they can be very different.

Bivariate Distributions

Bivariate Distributions: Discrete Case

- **Discrete Bivariate Distribution:**
  - **Joint pmf:**

$$f_{X,Y}(x,y) = \Pr(X = x, Y = y)$$

  - **Marginal pmf:**

$$f_X(x) = \Pr(X = x) = \sum_{y \in \mathcal{Y}} \Pr(X = x, Y = y)$$

  - **Conditional pmf:**

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$$

Bivariate Distributions: Discrete Case

- **Discrete Bivariate Distribution:**
  - **Joint pmf:**

$$f_{X,Y}(x, y) = \Pr(X = x, Y = y)$$

  - **Marginal pmf:**

$$f_X(x) = \Pr(X = x) = \sum_{y \in \mathcal{Y}} \Pr(X = x, Y = y)$$

  - **Conditional pmf:**

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x, y)}{f_Y(y)}$$

Bivariate Distributions: Discrete Case

- **Discrete Bivariate Distribution:**
  - **Joint pmf:**

$$f_{X,Y}(x,y) = \Pr(X = x, Y = y)$$

  - **Marginal pmf:**

$$f_X(x) = \Pr(X = x) = \sum_{y \in \mathcal{Y}} \Pr(X = x, Y = y)$$

  - **Conditional pmf:**

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$$

Bivariate Distributions: Discrete Case

- **Discrete Bivariate Distribution:**
  - **Joint pmf:**

$$f_{X,Y}(x,y) = \Pr(X = x, Y = y)$$

  - **Marginal pmf:**

$$f_X(x) = \Pr(X = x) = \sum_{y \in \mathcal{Y}} \Pr(X = x, Y = y)$$

  - **Conditional pmf:**

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$$

Bivariate Distributions: Continuous Case

- **Continuous Bivariate Distribution:**
  - **Joint cdf:**

$$F_{X,Y}(x,y) = \Pr(X \le x, Y \le y) = \int_{-\infty}^{x}\int_{-\infty}^{y} f_{X,Y}(u,t)\,du\,dt$$

  - **Joint pdf:**

$$f_{X,Y}(x,y) = \frac{\partial^2 F_{X,Y}}{\partial x\,\partial y}(x,y)$$

  - **Marginal cdf:**

$$F_X(x) = \Pr(X \le x) = F_{X,Y}(x,\infty)$$

  - **Marginal pdf:**

$$f_X(x) = \frac{dF_X}{dx}(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dy$$

  - **Conditional pdf:**

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$$

Bivariate Distributions: Continuous Case

- **Continuous Bivariate Distribution:**
  - **Joint cdf:**

$$F_{X,Y}(x,y) = \Pr(X \le x, Y \le y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(u,t)\, du\, dt$$

  - **Joint pdf:**

$$f_{X,Y}(x,y) = \frac{\partial^2 F_{X,Y}}{\partial x\, \partial y}(x,y)$$

  - **Marginal cdf:**

$$F_X(x) = \Pr(X \le x) = F_{X,Y}(x,\infty)$$

  - **Marginal pdf:**

$$f_X(x) = \frac{dF_X}{dx}(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\, dy$$

  - **Conditional pdf:**

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$$

Bivariate Distributions: Continuous Case

- **Continuous Bivariate Distribution:**
  - **Joint cdf:**

$$F_{X,Y}(x,y) = \Pr(X \le x, Y \le y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(u,t)\, du\, dt$$

  - **Joint pdf:**

$$f_{X,Y}(x,y) = \frac{\partial^2 F_{X,Y}}{\partial x\, \partial y}(x,y)$$

  - **Marginal cdf:**

$$F_X(x) = \Pr(X \le x) = F_{X,Y}(x,\infty)$$

  - **Marginal pdf:**

$$f_X(x) = \frac{dF_X}{dx}(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\, dy$$

  - **Conditional pdf:**

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$$

Bivariate Distributions: Continuous Case

- **Continuous Bivariate Distribution:**
  - **Joint cdf:**

$$F_{X,Y}(x,y) = \Pr(X \leq x, Y \leq y) = \int_{-\infty}^{x}\int_{-\infty}^{y} f_{X,Y}(u,t)\,du\,dt$$

  - **Joint pdf:**

$$f_{X,Y}(x,y) = \frac{\partial^2 F_{X,Y}}{\partial x\,\partial y}(x,y)$$

  - **Marginal cdf:**

$$F_X(x) = \Pr(X \leq x) = F_{X,Y}(x,\infty)$$

  - **Marginal pdf:**

$$f_X(x) = \frac{dF_X}{dx}(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dy$$

  - **Conditional pdf:**

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$$

Bivariate Distributions: Continuous Case

- **Continuous Bivariate Distribution:**
  - **Joint cdf:**

$$F_{X,Y}(x,y) = \Pr(X \leq x, Y \leq y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(u,t)\, du\, dt$$

  - **Joint pdf:**

$$f_{X,Y}(x,y) = \frac{\partial^2 F_{X,Y}}{\partial x\, \partial y}(x,y)$$

  - **Marginal cdf:**

$$F_X(x) = \Pr(X \leq x) = F_{X,Y}(x,\infty)$$

  - **Marginal pdf:**

$$f_X(x) = \frac{dF_X}{dx}(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\, dy$$

  - **Conditional pdf:**

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$$

Bivariate Distributions: Continuous Case

- **Continuous Bivariate Distribution:**
  - **Joint cdf:**

$$F_{X,Y}(x,y) = \Pr(X \leq x, Y \leq y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(u,t)\, du\, dt$$

  - **Joint pdf:**

$$f_{X,Y}(x,y) = \frac{\partial^2 F_{X,Y}}{\partial x\, \partial y}(x,y)$$

  - **Marginal cdf:**

$$F_X(x) = \Pr(X \leq x) = F_{X,Y}(x,\infty)$$

  - **Marginal pdf:**

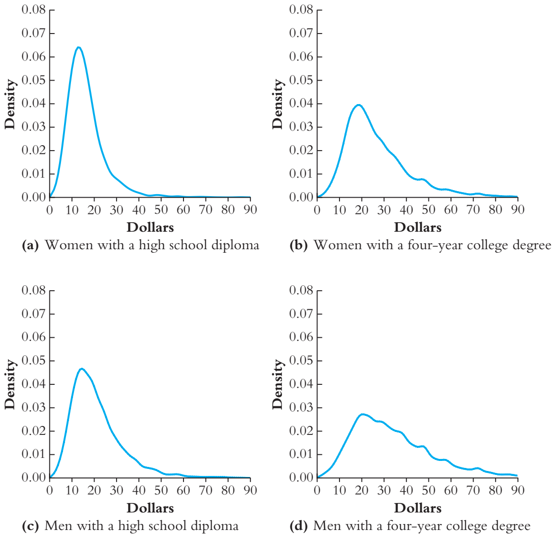
$$f_X(x) = \frac{dF_X}{dx}(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\, dy$$

  - **Conditional pdf:**

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$$

Conditional Distributions

Conditional Distributions: Average Hourly Earnings



Conditional Distribution of Average Hourly Earnings of U.S. Full-Time Workers in 2015, Given Education Level and Sex.

Conditional Distributions: Average Hourly Earnings

			Percentile			
	Mean	Standard Deviation	25%	50% (median)	75%	90%
(a) Women with a high school diploma	$16.28	$8.91	$10.99	$14.42	$19.23	$25.64
(b) Women with a four-year college degree	27.23	16.18	16.83	23.56	33.65	47.60
(c) Men with a high school diploma	21.22	11.96	13.22	19.12	26.10	36.06
(d) Men with a four-year college degree	35.10	20.36	20.67	30.92	44.71	60.90

Average hourly earnings are the sum of annual pre-tax wages, salaries, tips, and bonuses divided by the number of hours worked annually.

Conditional Distribution of Average Hourly Earnings of U.S. Full-Time Workers in 2015, Given Education Level and Sex.

Independence

Independence

- Two random variables X, Y are independent if

$$f_{X,Y}(x, y) = f_X(x) f_Y(y) \text{ for all } x, y$$

- This definition holds for both discrete and continuous variables.
- $X \perp\!\!\!\perp Y$ implies

$$f_{X|Y}(x|y) = f_X(x)$$

Independence

- Two random variables X, Y are independent if

$$f_{X,Y}(x, y) = f_X(x) f_Y(y) \text{ for all } x, y$$

- This definition holds for both discrete and continuous variables.
- $X \perp\!\!\!\perp Y$ implies

$$f_{X|Y}(x|y) = f_X(x)$$

Independence

- Two random variables X, Y are independent if

$$f_{X,Y}(x, y) = f_X(x) f_Y(y) \text{ for all } x, y$$

- This definition holds for both discrete and continuous variables.
- $X \perp\!\!\!\perp Y$ implies

$$f_{X|Y}(x|y) = f_X(x)$$

Expectations

Expectation

- **Expectation of Discrete Random Variable:**

$$\boxed{\mathrm{E}[X] = \sum_{i=1}^{n} x_i \, f_X(x_i)}$$

- **Expectation of Continuous Random Variable:**

$$\boxed{\mathrm{E}[X] = \int_{-\infty}^{\infty} x \, f_X(x) \, dx}$$

- The expectation of a random variable is also known as “expected value”, “first moment,” and more casually as “average.”

Expectation

- **Expectation of Discrete Random Variable:**

$$\boxed{\mathrm{E}[X] = \sum_{i=1}^{n} x_i \, f_X(x_i)}$$

- **Expectation of Continuous Random Variable:**

$$\boxed{\mathrm{E}[X] = \int_{-\infty}^{\infty} x \, f_X(x) \, dx}$$

- The expectation of a random variable is also known as "expected value", "first moment," and more casually as "average."

Expectation

- **Expectation of Discrete Random Variable:**

$$\boxed{\mathrm{E}[X] = \sum_{i=1}^{n} x_i \, f_X(x_i)}$$

- **Expectation of Continuous Random Variable:**

$$\boxed{\mathrm{E}[X] = \int_{-\infty}^{\infty} x \, f_X(x) \, dx}$$

- The expectation of a random variable is also known as “expected value”, “first moment,” and more casually as “average.”

Expectation

- **Properties of Expectations:**
- The expectation operator is linear:

$$\mathrm{E}[aX + bY] = a\,\mathrm{E}[X] + b\,\mathrm{E}[Y]$$

- The expectation of a composition is:

$$\mathrm{E}[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x) dx$$

for some real-valued function g with appropriate support.

Expectation

- **Properties of Expectations:**
- The expectation operator is linear:

$$\mathrm{E}[aX + bY] = a\,\mathrm{E}[X] + b\,\mathrm{E}[Y]$$

- The expectation of a composition is:

$$\mathrm{E}[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x) dx$$

for some real-valued function g with appropriate support.

Expectation

- **Properties of Expectations:**
- The expectation operator is linear:

$$\mathrm{E}[aX + bY] = a\,\mathrm{E}[X] + b\,\mathrm{E}[Y]$$

- The expectation of a composition is:

$$\mathrm{E}[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x) dx$$

for some real-valued function g with appropriate support.

Variance

Variance

- The expectation and variance are special cases of the "moments" of a random variable. The expectation is the first moment — The variance is the second "central" moment. Moments of order k are defined as follows:
- **kth moment of X:**

$$\mathrm{E}[X^k]$$

- **kth central moment of X:**

$$\mathrm{E}\left[(X - \mathrm{E}[X])^k\right]$$

- **Variance as Second Central Moment:**

$$\begin{aligned}\mathrm{var}(X) &= \mathrm{E}\left[(X - \mathrm{E}[X])^2\right] \\ &= \mathrm{E}[X^2] - 2\,\mathrm{E}[X \cdot \mathrm{E}[X]] + \mathrm{E}(\mathrm{E}[X])^2 \\ &= \mathrm{E}[X^2] - \mathrm{E}[X]^2\end{aligned}$$

Variance

- The expectation and variance are special cases of the "moments" of a random variable. The expectation is the first moment — The variance is the second "central" moment. Moments of order k are defined as follows:
- ***k*th moment** of X:

$$\mathrm{E}[X^k]$$

- kth central moment of X:

$$\mathrm{E}\left[(X - \mathrm{E}[X])^k\right]$$

- Variance as Second Central Moment:

$$\begin{aligned}\mathrm{var}(X) &= \mathrm{E}\left[(X - \mathrm{E}[X])^2\right] \\ &= \mathrm{E}[X^2] - 2\,\mathrm{E}[X \cdot \mathrm{E}[X]] + \mathrm{E}(\mathrm{E}[X])^2 \\ &= \mathrm{E}[X^2] - \mathrm{E}[X]^2\end{aligned}$$

Variance

- The expectation and variance are special cases of the “moments” of a random variable. The expectation is the first moment — The variance is the second “central” moment. Moments of order k are defined as follows:
- ***k*th moment** of X:

$$\mathrm{E}[X^k]$$

- ***k*th central moment** of X:

$$\mathrm{E}\left[(X - \mathrm{E}[X])^k\right]$$

- **Variance as Second Central Moment:**

$$\begin{aligned}\mathrm{var}(X) &= \mathrm{E}\left[(X - \mathrm{E}[X])^2\right] \\ &= \mathrm{E}[X^2] - 2\,\mathrm{E}[X \cdot \mathrm{E}[X]] + \mathrm{E}(\mathrm{E}[X])^2 \\ &= \mathrm{E}[X^2] - \mathrm{E}[X]^2\end{aligned}$$

Variance

- The expectation and variance are special cases of the "moments" of a random variable. The expectation is the first moment — The variance is the second "central" moment. Moments of order k are defined as follows:
- ***k*th moment** of X:

$$\mathrm{E}[X^k]$$

- ***k*th central moment** of X:

$$\mathrm{E}\left[(X - \mathrm{E}[X])^k\right]$$

- **Variance as Second Central Moment:**

$$\begin{aligned}\mathrm{var}(X) &= \mathrm{E}\left[(X - \mathrm{E}[X])^2\right] \\ &= \mathrm{E}[X^2] - 2\,\mathrm{E}[X \cdot \mathrm{E}[X]] + \mathrm{E}(\mathrm{E}[X])^2 \\ &= \mathrm{E}[X^2] - \mathrm{E}[X]^2\end{aligned}$$

Standard Deviation

- **Standard deviation:**

$$\boxed{\sigma(X) = \sqrt{\mathrm{var}(X)}}$$

- The standard deviation is expressed in the same units as the expected value, whereas the variance is expressed in squared-units, which may not be so easily interpreted.
- While the expectation operator is linear, the variance is quadratic in the following sense:

$$\mathrm{var}(a + bX) = b^2\,\mathrm{var}(X)$$

- The standard deviation satisfies:

$$\sigma(a + bX) = b\,\sigma(X)$$

  This is not the same as “linearity” since the constant term a has vanished.

Standard Deviation

- **Standard deviation:**

$$\boxed{\sigma(X) = \sqrt{\operatorname{var}(X)}}$$

- The standard deviation is expressed in the same units as the expected value, whereas the variance is expressed in squared-units, which may not be so easily interpreted.
- While the expectation operator is linear, the variance is quadratic in the following sense:

$$\operatorname{var}(a + bX) = b^2 \operatorname{var}(X)$$

- The standard deviation satisfies:

$$\sigma(a + bX) = b\,\sigma(X)$$

This is not the same as "linearity" since the constant term a has vanished.

Standard Deviation

- **Standard deviation:**

$$\boxed{\sigma(X) = \sqrt{\mathrm{var}(X)}}$$

- The standard deviation is expressed in the same units as the expected value, whereas the variance is expressed in squared-units, which may not be so easily interpreted.
- While the expectation operator is linear, the variance is quadratic in the following sense:

$$\mathrm{var}(a + bX) = b^2 \,\mathrm{var}(X)$$

- The standard deviation satisfies:

$$\sigma(a + bX) = b\,\sigma(X)$$

This is not the same as “linearity” since the constant term a has vanished.

Standard Deviation

- **Standard deviation:**

$$\boxed{\sigma(X) = \sqrt{\operatorname{var}(X)}}$$

- The standard deviation is expressed in the same units as the expected value, whereas the variance is expressed in squared-units, which may not be so easily interpreted.
- While the expectation operator is linear, the variance is quadratic in the following sense:

$$\operatorname{var}(a + bX) = b^2 \operatorname{var}(X)$$

- The standard deviation satisfies:

$$\sigma(a + bX) = b\,\sigma(X)$$

  This is not the same as "linearity" since the constant term a has vanished.

Covariance

Covariance

- **Covariance:**

$$\boxed{\operatorname{cov}(X,Y) = \mathrm{E}\left[(X - \mathrm{E}[X])\,(Y - \mathrm{E}[Y])\right]}$$

- **Properties:**

$$\begin{aligned}\operatorname{cov}(X,Y) &= \mathrm{E}[XY] - \mathrm{E}[X]\,\mathrm{E}[Y] \\ &= \mathrm{E}[(X - \mathrm{E}[X])\,Y] \\ &= \mathrm{E}[X\,(Y - \mathrm{E}[Y])]\end{aligned}$$

$$\operatorname{cov}(a + bX + cY, \alpha + \beta X + \gamma Y) = b\beta \operatorname{var}(X) + c\gamma \operatorname{var}(Y) + (b\gamma + c\beta)\operatorname{cov}(X,Y)$$

- **Bivariate Variance:**

$$\begin{aligned}\operatorname{var}(X) &= \operatorname{cov}(X,X) \\ \operatorname{var}(a + bX + cY) &= b^2 \operatorname{var}(X) + c^2 \operatorname{var}(Y) + 2bc\operatorname{cov}(X,Y)\end{aligned}$$

$$\operatorname{var}\left(\sum_{i=1}^{N} b_i X_i\right) = \sum_{i=1}^{N}\left(\sum_{j=1}^{N} b_i b_j \operatorname{cov}(X_i, X_j)\right)$$

Covariance

- **Covariance:**

$$\boxed{\operatorname{cov}(X,Y) = \mathrm{E}\left[(X - \mathrm{E}[X])\,(Y - \mathrm{E}[Y])\right]}$$

- **Properties:**

$$\begin{aligned}
\operatorname{cov}(X,Y) &= \mathrm{E}[XY] - \mathrm{E}[X]\,\mathrm{E}[Y] \\
&= \mathrm{E}[(X - \mathrm{E}[X])\,Y] \\
&= \mathrm{E}[X\,(Y - \mathrm{E}[Y])]
\end{aligned}$$

$$\operatorname{cov}(a + bX + cY, \alpha + \beta X + \gamma Y) = b\beta \operatorname{var}(X) + c\gamma \operatorname{var}(Y) + (b\gamma + c\beta)\operatorname{cov}(X,Y)$$

- **Bivariate Variance:**

$$\operatorname{var}(X) = \operatorname{cov}(X,X)$$

$$\operatorname{var}(a + bX + cY) = b^2 \operatorname{var}(X) + c^2 \operatorname{var}(Y) + 2bc\operatorname{cov}(X,Y)$$

$$\operatorname{var}\left(\sum_{i=1}^{N} b_i X_i\right) = \sum_{i=1}^{N}\left(\sum_{j=1}^{N} b_i b_j \operatorname{cov}(X_i, X_j)\right)$$

Covariance

- **Covariance:**

$$\boxed{\operatorname{cov}(X,Y) = \mathrm{E}\left[(X - \mathrm{E}[X])\,(Y - \mathrm{E}[Y])\right]}$$

- **Properties:**

$$\begin{aligned}\operatorname{cov}(X,Y) &= \mathrm{E}[XY] - \mathrm{E}[X]\,\mathrm{E}[Y] \\ &= \mathrm{E}[(X - \mathrm{E}[X])\,Y] \\ &= \mathrm{E}[X\,(Y - \mathrm{E}[Y])]\end{aligned}$$

$$\operatorname{cov}(a + bX + cY, \alpha + \beta X + \gamma Y) = b\beta \operatorname{var}(X) + c\gamma \operatorname{var}(Y) + (b\gamma + c\beta) \operatorname{cov}(X,Y)$$

- **Bivariate Variance:**

$$\begin{aligned}\operatorname{var}(X) &= \operatorname{cov}(X,X) \\ \operatorname{var}(a + bX + cY) &= b^2 \operatorname{var}(X) + c^2 \operatorname{var}(Y) + 2bc \operatorname{cov}(X,Y) \\ \operatorname{var}\left(\sum_{i=1}^{N} b_i X_i\right) &= \sum_{i=1}^{N}\left(\sum_{j=1}^{N} b_i b_j \operatorname{cov}(X_i, X_j)\right)\end{aligned}$$

Variance of Sums

Variance of Sums

- Variance of sums:

$$
\begin{aligned}
\operatorname{var}(a+bX+cY) &= \mathrm{E}\left[(a+bX+cY-\mathrm{E}[a+bX+cY])^2\right] \\
&= \mathrm{E}\left[(a+bX+cY-(a+b\,\mathrm{E}[X]+c\,\mathrm{E}[Y]))^2\right] \\
&= \mathrm{E}\left[(b(X-\mathrm{E}[X])+c(Y-\mathrm{E}[Y]))^2\right] \\
&= \mathrm{E}\left[b^2(X-\mathrm{E}[X])^2+c^2(Y-\mathrm{E}[Y])^2\right. \\
&\quad \left. +\,2bc(X-\mathrm{E}[X])(Y-\mathrm{E}[Y])\right] \\
&= b^2\,\mathrm{E}[(X-\mathrm{E}[X])^2]+c^2\,\mathrm{E}[(Y-\mathrm{E}[Y])^2] \\
&\quad +\ 2bc\,\mathrm{E}(X-\mathrm{E}[X])(Y-\mathrm{E}[Y]) \\
&= b^2\operatorname{var}(X)+c^2\operatorname{var}(Y)+2bc\operatorname{cov}(X,Y)
\end{aligned}
$$

Correlation

Correlation

- **Correlation:**

$$\boxed{\operatorname{corr}(X,Y) = \frac{\operatorname{cov}(X,Y)}{\sqrt{\operatorname{var}(X)\cdot\operatorname{var}(Y)}}}$$

- **Cauchy-Schwartz Inequality:**

$$|\operatorname{cov}(X,Y)| \leq \sqrt{\operatorname{var}(X)\cdot\operatorname{var}(Y)}$$

- **Properties**

$$\operatorname{corr}(X,Y) = \operatorname{corr}(Y,X)$$
$$-1 \leq \operatorname{corr}(X,Y) \leq 1$$
$$\operatorname{corr}(X,Y) = +1 \quad \text{if} \quad Y = a + bX, \quad \text{with} \quad b > 0$$
$$\operatorname{corr}(X,Y) = -1 \quad \text{if} \quad Y = a + bX, \quad \text{with} \quad b < 0$$
$$\mathrm{E}[X|Y] = \mathrm{E}[X] \implies \operatorname{corr}(X,Y) = 0$$
$$\mathrm{E}[Y|X] = \mathrm{E}[Y] \implies \operatorname{corr}(X,Y) = 0$$

Correlation

- **Correlation:**

$$\boxed{\operatorname{corr}(X,Y) = \frac{\operatorname{cov}(X,Y)}{\sqrt{\operatorname{var}(X)\cdot\operatorname{var}(Y)}}}$$

- **Cauchy-Schwartz Inequality:**

$$|\operatorname{cov}(X,Y)| \leq \sqrt{\operatorname{var}(X)\cdot\operatorname{var}(Y)}$$

- **Properties**

$$\operatorname{corr}(X,Y) = \operatorname{corr}(Y,X)$$

$$-1 \leq \operatorname{corr}(X,Y) \leq 1$$

$$\operatorname{corr}(X,Y) = +1 \quad \text{if} \quad Y = a + bX, \quad \text{with} \quad b > 0$$

$$\operatorname{corr}(X,Y) = -1 \quad \text{if} \quad Y = a + bX, \quad \text{with} \quad b < 0$$

$$\mathrm{E}[X|Y] = \mathrm{E}[X] \implies \operatorname{corr}(X,Y) = 0$$

$$\mathrm{E}[Y|X] = \mathrm{E}[Y] \implies \operatorname{corr}(X,Y) = 0$$

Correlation

- **Correlation:**

$$\boxed{\operatorname{corr}(X,Y) = \frac{\operatorname{cov}(X,Y)}{\sqrt{\operatorname{var}(X)\cdot\operatorname{var}(Y)}}}$$

- **Cauchy-Schwartz Inequality:**

$$|\operatorname{cov}(X,Y)| \le \sqrt{\operatorname{var}(X)\cdot\operatorname{var}(Y)}$$

- **Properties**

$$\begin{aligned}
&\operatorname{corr}(X,Y) = \operatorname{corr}(Y,X)\\
&-1 \le \operatorname{corr}(X,Y) \le 1\\
&\operatorname{corr}(X,Y) = +1 \quad \text{if} \quad Y = a + bX, \quad \text{with} \quad b > 0\\
&\operatorname{corr}(X,Y) = -1 \quad \text{if} \quad Y = a + bX, \quad \text{with} \quad b < 0\\
&\mathrm{E}[X|Y] = \mathrm{E}[X] \implies \operatorname{corr}(X,Y) = 0\\
&\mathrm{E}[Y|X] = \mathrm{E}[Y] \implies \operatorname{corr}(X,Y) = 0
\end{aligned}$$

Conditional Expectation

Conditional Expectation

- **Conditional Expectation of Discrete Random Variable:**

$$\mathrm{E}[Y|X=x] = \sum_{i=1}^{n} y_i \Pr(y_i|X=x)$$

- **Conditional Expectation of Continuous Random Variable:**

$$\mathrm{E}[Y|X=x] = \int_{-\infty}^{\infty} y f_{Y|X}(y|X=x) dy$$

Conditional Expectation

- **Conditional Expectation of Discrete Random Variable:**

$$\mathrm{E}[Y|X=x] = \sum_{i=1}^{n} y_i \Pr(y_i|X=x)$$

- **Conditional Expectation of Continuous Random Variable:**

$$\mathrm{E}[Y|X=x] = \int_{-\infty}^{\infty} y f_{Y|X}(y|X=x) dy$$

Iterated Expectations

Law of Iterated Expectation

- **Law of Iterated Expectation:**

$$\mathrm{E}[Y] = \mathrm{E}[\mathrm{E}(Y|X)]$$

- The expected value of Y can be calculated from the probability distribution of $Y|X$ and X.
- If X has sample space $x_1, x_2, \ldots, x_n$, the law can be written explicitly as

$$\mathrm{E}[Y] = \sum_{i=1}^{n} \mathrm{E}[Y|X = x_i] \Pr(X = x_i)$$

- You choose an integer X at random between 1 and 3. Then you choose an integer Y at random between 1 and $X = x$. Calculate the expected value of Y.

$$\begin{aligned}
\mathrm{E}[Y] = \mathrm{E}[\mathrm{E}(Y|X)] &= \sum_{i=1}^{3} \mathrm{E}[Y|X = x_i] \Pr(X = x_i) = \frac{1}{3} \sum_{i=1}^{3} \mathrm{E}[Y|X = x_i] \\
&= \frac{1}{3} \cdot \left(\frac{1}{1} + \frac{1+2}{2} + \frac{1+2+3}{3} \right) \\
&= 1.5
\end{aligned}$$

Law of Iterated Expectation

- **Law of Iterated Expectation:**

$$\mathrm{E}[Y] = \mathrm{E}[\mathrm{E}(Y|X)]$$

- The expected value of Y can be calculated from the probability distribution of $Y|X$ and X.
- If X has sample space $x_1, x_2, \ldots, x_n$, the law can be written explicitly as

$$\mathrm{E}[Y] = \sum_{i=1}^{n} \mathrm{E}[Y|X = x_i] \Pr(X = x_i)$$

- You choose an integer X at random between 1 and 3. Then you choose an integer Y at random between 1 and $X = x$. Calculate the expected value of Y.

$$\begin{aligned}
\mathrm{E}[Y] = \mathrm{E}[\mathrm{E}(Y|X)] = \sum_{i=1}^{3} \mathrm{E}[Y|X = x_i] \Pr(X = x_i) = \frac{1}{3}\sum_{i=1}^{3} \mathrm{E}[Y|X = x_i] \\
= \frac{1}{3} \cdot \left(\frac{1}{1} + \frac{1+2}{2} + \frac{1+2+3}{3}\right) \\
= 1.5
\end{aligned}$$

Law of Iterated Expectation

- **Law of Iterated Expectation:**

$$\mathrm{E}[Y] = \mathrm{E}[\mathrm{E}(Y|X)]$$

- The expected value of Y can be calculated from the probability distribution of $Y|X$ and X.
- If X has sample space $x_1, x_2, \ldots, x_n$, the law can be written explicitly as

$$\mathrm{E}[Y] = \sum_{i=1}^{n} \mathrm{E}[Y|X = x_i] \Pr(X = x_i)$$

- You choose an integer X at random between 1 and 3. Then you choose an integer Y at random between 1 and $X = x$. Calculate the expected value of Y.

$$\begin{aligned}\mathrm{E}[Y] = \mathrm{E}[\mathrm{E}(Y|X)] &= \sum_{i=1}^{3} \mathrm{E}[Y|X = x_i] \Pr(X = x_i) = \frac{1}{3} \sum_{i=1}^{3} \mathrm{E}[Y|X = x_i] \\ &= \frac{1}{3} \cdot \left(\frac{1}{1} + \frac{1+2}{2} + \frac{1+2+3}{3} \right) \\ &= 1.5\end{aligned}$$

Law of Iterated Expectation

- **Law of Iterated Expectation:**

$$E[Y] = E[E(Y|X)]$$

- The expected value of Y can be calculated from the probability distribution of $Y|X$ and X.
- If X has sample space $x_1, x_2, \ldots, x_n$, the law can be written explicitly as

$$E[Y] = \sum_{i=1}^{n} E[Y|X = x_i] \Pr(X = x_i)$$

- You choose an integer X at random between 1 and 3. Then you choose an integer Y at random between 1 and $X = x$. Calculate the expected value of Y.

$$\begin{aligned} E[Y] &= E[E(Y|X)] = \sum_{i=1}^{3} E[Y|X = x_i] \Pr(X = x_i) = \frac{1}{3}\sum_{i=1}^{3} E[Y|X = x_i] \\ &= \frac{1}{3} \cdot \left(\frac{1}{1} + \frac{1+2}{2} + \frac{1+2+3}{3}\right) \\ &= 1.5 \end{aligned}$$

Conditional Variance

Conditional Variance

- **Conditional Variance:**
$$\mathrm{var}(Y|X) = \mathrm{E}\left[(Y - \mathrm{E}[Y|X])^2|X\right]$$
- **Law of Total Variance:**
$$\mathrm{var}(Y) = \mathrm{E}\left[\mathrm{var}(Y|X)\right] + \mathrm{var}\left(\mathrm{E}[Y|X]\right)$$
where the second term $\mathrm{var}\left(\mathrm{E}[Y|X]\right)$ captures the explained part of the variance and the first term $\mathrm{E}\left[\mathrm{var}(Y|X)\right]$ the unexplained part.
- The Analysis of Variance (ANOVA) is based on the Law of Total Variance. In ANOVA, the observed variance in a particular variable is partitioned into components attributable to different sources of variation.
- The law of total variance is also known as the law of iterated variances.

Conditional Variance

- **Conditional Variance:**

$$\operatorname{var}(Y|X) = \mathrm{E}\left[(Y - \mathrm{E}[Y|X])^2|X\right]$$

- **Law of Total Variance:**

$$\operatorname{var}(Y) = \mathrm{E}\left[\operatorname{var}(Y|X)\right] + \operatorname{var}\left(\mathrm{E}[Y|X]\right)$$

  where the second term $\operatorname{var}\left(\mathrm{E}[Y|X]\right)$ captures the explained part of the variance and the first term $\mathrm{E}\left[\operatorname{var}(Y|X)\right]$ the unexplained part.

- The Analysis of Variance (ANOVA) is based on the Law of Total Variance. In ANOVA, the observed variance in a particular variable is partitioned into components attributable to different sources of variation.

- The law of total variance is also known as the law of iterated variances.

Conditional Variance

- **Conditional Variance:**

$$\operatorname{var}(Y|X) = \mathrm{E}\left[(Y - \mathrm{E}[Y|X])^2|X\right]$$

- **Law of Total Variance:**

$$\operatorname{var}(Y) = \mathrm{E}\left[\operatorname{var}(Y|X)\right] + \operatorname{var}\left(\mathrm{E}[Y|X]\right)$$

  where the second term $\operatorname{var}\left(\mathrm{E}[Y|X]\right)$ captures the explained part of the variance and the first term $\mathrm{E}\left[\operatorname{var}(Y|X)\right]$ the unexplained part.

- The Analysis of Variance (ANOVA) is based on the Law of Total Variance. In ANOVA, the observed variance in a particular variable is partitioned into components attributable to different sources of variation.

- The law of total variance is also known as the law of iterated variances.

Conditional Variance

- **Conditional Variance:**

$$\mathrm{var}(Y|X) = \mathrm{E}\left[(Y - \mathrm{E}[Y|X])^2|X\right]$$

- **Law of Total Variance:**

$$\mathrm{var}(Y) = \mathrm{E}\left[\mathrm{var}(Y|X)\right] + \mathrm{var}\left(\mathrm{E}[Y|X]\right)$$

  where the second term $\mathrm{var}\left(\mathrm{E}[Y|X]\right)$ captures the explained part of the variance and the first term $\mathrm{E}\left[\mathrm{var}(Y|X)\right]$ the unexplained part.

- The Analysis of Variance (ANOVA) is based on the Law of Total Variance. In ANOVA, the observed variance in a particular variable is partitioned into components attributable to different sources of variation.

- The law of total variance is also known as the law of iterated variances.

Data Types

Data Types

- **Experimental Data:** Obtained from experiments designed to assess the causal effect of a treatment on an outcome.
  - Randomized controlled trials: Ideal experimental data for program evaluation.
  - Example: Tennessee STAR project (Student-Teacher Achievement Ratio). A four-year longitudinal study: Over 7,000 students in 79 schools randomly assigned into small/medium/large classes.
- **Observational Data:** A type of data where researchers have no control on how the treatment is allocated.
  - Obtained from surveys or administrative records: e.g. National Education Longitudinal Study.
- **Cross-Sectional Data:** A type of data collected by observing many subjects (such as individuals, firms, countries, or regions) at one point in time.
  - Example: California Test Score data: each district is one observation.
- **Time-Series Data:** A type of data collected at successive points in time.
  - Example: US inflation and unemployment rate data. Successive points in time are usually equally spaced (daily, monthly, quarterly, annual), but may not be.

Data Types

- **Experimental Data:** Obtained from experiments designed to assess the causal effect of a treatment on an outcome.
  - Randomized controlled trials: Ideal experimental data for program evaluation.
  - Example: Tennessee STAR project (Student-Teacher Achievement Ratio). A four-year longitudinal study: Over 7,000 students in 79 schools randomly assigned into small/medium/large classes.
- **Observational Data:** A type of data where researchers have no control on how the treatment is allocated.
- **Cross-Sectional Data:** A type of data collected by observing many subjects (such as individuals, firms, countries, or regions) at one point in time.
- **Time-Series Data:** A type of data collected at successive points in time.

Data Types

- **Experimental Data:** Obtained from experiments designed to assess the causal effect of a treatment on an outcome.
  - Randomized controlled trials: Ideal experimental data for program evaluation.
  - Example: Tennessee STAR project (Student-Teacher Achievement Ratio). A four-year longitudinal study: Over 7,000 students in 79 schools randomly assigned into small/medium/large classes.
- **Observational Data:** A type of data where researchers have no control on how the treatment is allocated.
- **Cross-Sectional Data:** A type of data collected by observing many subjects (such as individuals, firms, countries, or regions) at one point in time.
- **Time-Series Data:** A type of data collected at successive points in time.

Data Types

- **Experimental Data:** Obtained from experiments designed to assess the causal effect of a treatment on an outcome.
  - Randomized controlled trials: Ideal experimental data for program evaluation.
  - Example: Tennessee STAR project (Student-Teacher Achievement Ratio). A four-year longitudinal study: Over 7,000 students in 79 schools randomly assigned into small/medium/large classes.
- **Observational Data:** A type of data where researchers have no control on how the treatment is allocated.
  - Obtained from surveys or administrative records, e.g. National Education Longitudinal Study.
- **Cross-Sectional Data:** A type of data collected by observing many subjects (such as individuals, firms, countries, or regions) at one point in time.
  - Example: California Test Score data: each district is one observation.
- **Time-Series Data:** A type of data collected at successive points in time.
  - Example: US inflation and unemployment rate data. Successive points in time are usually equally spaced (daily, monthly, quarterly, annual), but may not be.

Data Types

- **Experimental Data:** Obtained from experiments designed to assess the causal effect of a treatment on an outcome.
  - Randomized controlled trials: Ideal experimental data for program evaluation.
  - Example: Tennessee STAR project (Student-Teacher Achievement Ratio). A four-year longitudinal study: Over 7,000 students in 79 schools randomly assigned into small/medium/large classes.
- **Observational Data:** A type of data where researchers have no control on how the treatment is allocated.
  - Obtained from surveys or administrative records, e.g. National Education Longitudinal Study.
- **Cross-Sectional Data:** A type of data collected by observing many subjects (such as individuals, firms, countries, or regions) at one point in time.
- **Time-Series Data:** A type of data collected at successive points in time.

Data Types

- **Experimental Data:** Obtained from experiments designed to assess the causal effect of a treatment on an outcome.
  - Randomized controlled trials: Ideal experimental data for program evaluation.
  - Example: Tennessee STAR project (Student-Teacher Achievement Ratio). A four-year longitudinal study: Over 7,000 students in 79 schools randomly assigned into small/medium/large classes.
- **Observational Data:** A type of data where researchers have no control on how the treatment is allocated.
  - Obtained from surveys or administrative records, e.g. National Education Longitudinal Study.
- **Cross-Sectional Data:** A type of data collected by observing many subjects (such as individuals, firms, countries, or regions) at one point in time.
  - Example: California Test Score data: each district is one observation.
- **Time-Series Data:** A type of data collected at successive points in time.

Data Types

- **Experimental Data:** Obtained from experiments designed to assess the causal effect of a treatment on an outcome.
  - Randomized controlled trials: Ideal experimental data for program evaluation.
  - Example: Tennessee STAR project (Student-Teacher Achievement Ratio). A four-year longitudinal study: Over 7,000 students in 79 schools randomly assigned into small/medium/large classes.
- **Observational Data:** A type of data where researchers have no control on how the treatment is allocated.
  - Obtained from surveys or administrative records, e.g. National Education Longitudinal Study.
- **Cross-Sectional Data:** A type of data collected by observing many subjects (such as individuals, firms, countries, or regions) at one point in time.
  - Example: California Test Score data: each district is one observation.
- **Time-Series Data:** A type of data collected at successive points in time.

Data Types

- **Experimental Data:** Obtained from experiments designed to assess the causal effect of a treatment on an outcome.
  - Randomized controlled trials: Ideal experimental data for program evaluation.
  - Example: Tennessee STAR project (Student-Teacher Achievement Ratio). A four-year longitudinal study: Over 7,000 students in 79 schools randomly assigned into small/medium/large classes.
- **Observational Data:** A type of data where researchers have no control on how the treatment is allocated.
  - Obtained from surveys or administrative records, e.g. National Education Longitudinal Study.
- **Cross-Sectional Data:** A type of data collected by observing many subjects (such as individuals, firms, countries, or regions) at one point in time.
  - Example: California Test Score data: each district is one observation.
- **Time-Series Data:** A type of data collected at successive points in time.
  - Example: US inflation and unemployment rate data. Successive points in time are usually equally spaced (daily, monthly, quarterly, annual), but may not be.

Data Types

- **Experimental Data:** Obtained from experiments designed to assess the causal effect of a treatment on an outcome.
  - Randomized controlled trials: Ideal experimental data for program evaluation.
  - Example: Tennessee STAR project (Student-Teacher Achievement Ratio). A four-year longitudinal study: Over 7,000 students in 79 schools randomly assigned into small/medium/large classes.
- **Observational Data:** A type of data where researchers have no control on how the treatment is allocated.
  - Obtained from surveys or administrative records, e.g. National Education Longitudinal Study.
- **Cross-Sectional Data:** A type of data collected by observing many subjects (such as individuals, firms, countries, or regions) at one point in time.
  - Example: California Test Score data: each district is one observation.
- **Time-Series Data:** A type of data collected at successive points in time.
  - Example: US inflation and unemployment rate data. Successive points in time are usually equally spaced (daily, monthly, quarterly, annual), but may not be.

Data Types

- **Longitudinal Data:** A research design that involves repeated observations of the same variables over periods of time.
  - Example: Panel Study of Income Dynamics (PSID). It may or may not be randomized. Often used to monitor individuals across several periods of their lives.
- **Cohort Studies:** One type of longitudinal study which sample a cohort — a group of people who share a defining characteristic, who experienced a common event, such as birth or graduation.
  - Example: Millenium Cohort Study (evaluate the long-term health effects of military service).
- **Panel Data:** A subset of longitudinal data where observations are for the same subjects each time. A balanced panel is a dataset in which each panel member is observed every year.
  - Example: National Longitudinal Survey of Youth (NLSY).

Data Types

- **Longitudinal Data:** A research design that involves repeated observations of the same variables over periods of time.
  - Example: Panel Study of Income Dynamics (PSID). It may or may not be randomized. Often used to monitor individuals across several periods of their lives.
- **Cohort Studies:** One type of longitudinal study which sample a cohort — a group of people who share a defining characteristic, who experienced a common event, such as birth or graduation.
  - Example: Millenium Cohort Study (evaluate the long-term health effects of military service).
- **Panel Data:** A subset of longitudinal data where observations are for the same subjects each time. A balanced panel is a dataset in which each panel member is observed every year.
  - Example: National Longitudinal Survey of Youth (NLSY)

Data Types

- **Longitudinal Data:** A research design that involves repeated observations of the same variables over periods of time.
  - Example: Panel Study of Income Dynamics (PSID). It may or may not be randomized. Often used to monitor individuals across several periods of their lives.
- **Cohort Studies:** One type of longitudinal study which sample a cohort — a group of people who share a defining characteristic, who experienced a common event, such as birth or graduation.
  - Example: Millenium Cohort Study (evaluate the long-term health effects of military service).
- **Panel Data:** A subset of longitudinal data where observations are for the same subjects each time. A balanced panel is a dataset in which each panel member is observed every year.
  - Example: National Longitudinal Survey of Youth (NLSY).

Data Types

- **Longitudinal Data:** A research design that involves repeated observations of the same variables over periods of time.
  - Example: Panel Study of Income Dynamics (PSID). It may or may not be randomized. Often used to monitor individuals across several periods of their lives.
- **Cohort Studies:** One type of longitudinal study which sample a cohort — a group of people who share a defining characteristic, who experienced a common event, such as birth or graduation.
  - Example: Millenium Cohort Study (evaluate the long-term health effects of military service).
- **Panel Data:** A subset of longitudinal data where observations are for the same subjects each time. A balanced panel is a dataset in which each panel member is observed every year.
  - Example: National Longitudinal Survey of Youth (NLSY).

Data Types

- **Longitudinal Data:** A research design that involves repeated observations of the same variables over periods of time.
  - Example: Panel Study of Income Dynamics (PSID). It may or may not be randomized. Often used to monitor individuals across several periods of their lives.
- **Cohort Studies:** One type of longitudinal study which sample a cohort — a group of people who share a defining characteristic, who experienced a common event, such as birth or graduation.
  - Example: Millenium Cohort Study (evaluate the long-term health effects of military service).
- **Panel Data:** A subset of longitudinal data where observations are for the same subjects each time. A balanced panel is a dataset in which each panel member is observed every year.
  - Example: National Longitudinal Survey of Youth (NLSY).

Data Types

- **Longitudinal Data:** A research design that involves repeated observations of the same variables over periods of time.
  - Example: Panel Study of Income Dynamics (PSID). It may or may not be randomized. Often used to monitor individuals across several periods of their lives.
- **Cohort Studies:** One type of longitudinal study which sample a cohort — a group of people who share a defining characteristic, who experienced a common event, such as birth or graduation.
  - Example: Millenium Cohort Study (evaluate the long-term health effects of military service).
- **Panel Data:** A subset of longitudinal data where observations are for the same subjects each time. A balanced panel is a dataset in which each panel member is observed every year.
  - Example: National Longitudinal Survey of Youth (NLSY).

Centrality

Measures of Central Tendency

- **Population mean:** μ_Y
- A population mean may or may not exist.
- **Sample mean:** $\overline{Y}$
- The sample mean is constructed from observed realizations of the random variable Y, so for a sample of size n,

$$\overline{Y} = \frac{y_1 + y_2 + \ldots + y_n}{n}$$

  where y_i denotes the measured sample values. The bar is a short-hand for "sample mean."
- The central tendency is also referred to as "location."
- The expected value of Y is also called the mean of Y.

$$\mathrm{E}[Y] = \sum_{i=1}^{n} y_i \Pr(y_i) = \frac{1}{n}\sum_{i=1}^{n} y_i$$

- Other measures of central tendency: **Median, Mode, Truncated Mean, Geometric Mean, Harmonic Mean**. The geometric mean is particularly useful for time series, e.g. to compute the compounded annual growth rate.

Measures of Central Tendency

- **Population mean:** μ_Y
- A population mean may or may not exist.
- **Sample mean:** $\overline{Y}$
- The sample mean is constructed from observed realizations of the random variable Y, so for a sample of size n,

$$\overline{Y} = \frac{y_1 + y_2 + \ldots + y_n}{n}$$

  where y_i denotes the measured sample values. The bar is a short-hand for "sample mean."
- The central tendency is also referred to as "location."
- The expected value of Y is also called the mean of Y.

$$\mathrm{E}[Y] = \sum_{i=1}^{n} y_i \Pr(y_i) = \frac{1}{n} \sum_{i=1}^{n} y_i$$

- Other measures of central tendency: **Median**, **Mode**, **Truncated Mean**, **Geometric Mean**, **Harmonic Mean**. The geometric mean is particularly useful for time series, e.g. to compute the compounded annual growth rate.

Measures of Central Tendency

- **Population mean:** μ_Y
- A population mean may or may not exist.
- **Sample mean:** $\overline{Y}$
- The sample mean is constructed from observed realizations of the random variable Y, so for a sample of size n,

$$\overline{Y} = \frac{y_1 + y_2 + \ldots + y_n}{n}$$

  where y_i denotes the measured sample values. The bar is a short-hand for "sample mean."
- The central tendency is also referred to as "location."
- The expected value of Y is also called the mean of Y.

$$\mathrm{E}[Y] = \sum_{i=1}^{n} y_i \Pr(y_i) = \frac{1}{n}\sum_{i=1}^{n} y_i$$

- Other measures of central tendency: **Median**, **Mode**, **Truncated Mean**, **Geometric Mean**, **Harmonic Mean**. The geometric mean is particularly useful for time series, e.g. to compute the compounded annual growth rate.

Measures of Central Tendency

- **Population mean:** μ_Y
- A population mean may or may not exist.
- **Sample mean:** $\overline{Y}$
- The sample mean is constructed from observed realizations of the random variable Y, so for a sample of size n,

$$\overline{Y} = \frac{y_1 + y_2 + \ldots + y_n}{n}$$

  where y_i denotes the measured sample values. The bar is a short-hand for "sample mean."
- The central tendency is also referred to as "location."
- The expected value of Y is also called the mean of Y.

$$\mathrm{E}[Y] = \sum_{i=1}^{n} y_i \Pr(y_i) = \frac{1}{n} \sum_{i=1}^{n} y_i$$

- Other measures of central tendency: **Median, Mode, Truncated Mean, Geometric Mean, Harmonic Mean.** The geometric mean is particularly useful for time series, e.g. to compute the compounded annual growth rate.

Measures of Central Tendency

- **Population mean:** μ_Y
- A population mean may or may not exist.
- **Sample mean:** $\overline{Y}$
- The sample mean is constructed from observed realizations of the random variable Y, so for a sample of size n,

$$\overline{Y} = \frac{y_1 + y_2 + \ldots + y_n}{n}$$

  where y_i denotes the measured sample values. The bar is a short-hand for "sample mean."
- The central tendency is also referred to as "location."
- The expected value of Y is also called the mean of Y.

$$\mathrm{E}[Y] = \sum_{i=1}^{n} y_i \Pr(y_i) = \frac{1}{n} \sum_{i=1}^{n} y_i$$

- Other measures of central tendency: **Median, Mode, Truncated Mean, Geometric Mean, Harmonic Mean.** The geometric mean is particularly useful for time series, e.g. to compute the compounded annual growth rate.

Measures of Central Tendency

- **Population mean:** μ_Y
- A population mean may or may not exist.
- **Sample mean:** $\overline{Y}$
- The sample mean is constructed from observed realizations of the random variable Y, so for a sample of size n,

$$\overline{Y} = \frac{y_1 + y_2 + \ldots + y_n}{n}$$

  where y_i denotes the measured sample values. The bar is a short-hand for "sample mean."
- The central tendency is also referred to as "location."
- The expected value of Y is also called the mean of Y.

$$\mathrm{E}[Y] = \sum_{i=1}^{n} y_i \Pr(y_i) = \frac{1}{n}\sum_{i=1}^{n} y_i$$

- Other measures of central tendency: **Median, Mode, Truncated Mean, Geometric Mean, Harmonic Mean.** The geometric mean is particularly useful for time series, e.g. to compute the compounded annual growth rate.

Measures of Central Tendency

- **Population mean:** μ_Y
- A population mean may or may not exist.
- **Sample mean:** $\overline{Y}$
- The sample mean is constructed from observed realizations of the random variable Y, so for a sample of size n,

$$\overline{Y} = \frac{y_1 + y_2 + \ldots + y_n}{n}$$

where y_i denotes the measured sample values. The bar is a short-hand for "sample mean."
- The central tendency is also referred to as "location."
- The expected value of Y is also called the mean of Y.

$$\mathrm{E}[Y] = \sum_{i=1}^{n} y_i \Pr(y_i) = \frac{1}{n} \sum_{i=1}^{n} y_i$$

- Other measures of central tendency: **Median**, **Mode**, **Truncated Mean**, **Geometric Mean**, **Harmonic Mean**. The geometric mean is particularly useful for time series, e.g. to compute the compounded annual growth rate.

Dispersion

- **Population standard deviation:** σ_Y.
- A population STD may or may not exist. For a population of size N,

$$\sigma_Y = \sqrt{\frac{\sum_{i=1}^{N}(Y_i - \overline{Y})^2}{N}}$$

- **Sample standard deviation:** $\hat{\sigma}_Y$ or s_Y.
- Measures how far away, on average, a random observation Y_i is from the sample mean $\overline{Y}$,

$$\hat{\sigma}_Y = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \overline{Y})^2}{n-1}}$$

  The hat is a short-hand for "estimated." Division by $n-1$ (rather than n) makes the estimate unbiased — for large n, it makes little difference.
- Other measures of dispersion: **Variance, Range, Interquartile Range (IQR), Median Absolute Deviation (MAD), Average Absolute Deviation (AAD).**
- The MAD is more robust to outliers than the STD.

Measures of Dispersion

- **Population standard deviation:** σ_Y.
- A population STD may or may not exist. For a population of size N,

$$\sigma_Y = \sqrt{\frac{\sum_{i=1}^{N}(Y_i - \overline{Y})^2}{N}}$$

- **Sample standard deviation:** $\hat{\sigma}_Y$ or s_Y.
- Measures how far away, on average, a random observation Y_i is from the sample mean $\overline{Y}$,

$$\hat{\sigma}_Y = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \overline{Y})^2}{n-1}}$$

  The hat is a short-hand for "estimated." Division by $n-1$ (rather than n) makes the estimate unbiased — for large n, it makes little difference.
- Other measures of dispersion: **Variance, Range, Interquartile Range (IQR), Median Absolute Deviation (MAD), Average Absolute Deviation (AAD).**
- The MAD is more robust to outliers than the STD.

Measures of Dispersion

- **Population standard deviation:** σ_Y.
- A population STD may or may not exist. For a population of size N,

$$\sigma_Y = \sqrt{\frac{\sum_{i=1}^{N}(Y_i - \overline{Y})^2}{N}}$$

- **Sample standard deviation:** $\hat{\sigma}_Y$ or s_Y.
- Measures how far away, on average, a random observation Y_i is from the sample mean $\overline{Y}$,

$$\hat{\sigma}_Y = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \overline{Y})^2}{n-1}}$$

  The hat is a short-hand for "estimated." Division by $n-1$ (rather than n) makes the estimate unbiased — for large n, it makes little difference.
- Other measures of dispersion: **Variance, Range, Interquartile Range (IQR), Median Absolute Deviation (MAD), Average Absolute Deviation (AAD).**
- The MAD is more robust to outliers than the STD.

Measures of Dispersion

- **Population standard deviation:** σ_Y.
- A population STD may or may not exist. For a population of size N,

$$\sigma_Y = \sqrt{\frac{\sum_{i=1}^{N}(Y_i - \overline{Y})^2}{N}}$$

- **Sample standard deviation:** $\hat{\sigma}_Y$ or s_Y.
- Measures how far away, on average, a random observation Y_i is from the sample mean $\overline{Y}$,

$$\hat{\sigma}_Y = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \overline{Y})^2}{n-1}}$$

  The hat is a short-hand for "estimated." Division by $n-1$ (rather than n) makes the estimate unbiased — for large n, it makes little difference.
- Other measures of dispersion: **Variance, Range, Interquartile Range (IQR), Median Absolute Deviation (MAD), Average Absolute Deviation (AAD).**
- The MAD is more robust to outliers than the STD.

Measures of Dispersion

- **Population standard deviation:** σ_Y.
- A population STD may or may not exist. For a population of size N,

$$\sigma_Y = \sqrt{\frac{\sum_{i=1}^{N}(Y_i - \overline{Y})^2}{N}}$$

- **Sample standard deviation:** $\hat{\sigma}_Y$ or s_Y.
- Measures how far away, on average, a random observation Y_i is from the sample mean $\overline{Y}$,

$$\hat{\sigma}_Y = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \overline{Y})^2}{n-1}}$$

  The hat is a short-hand for "estimated." Division by $n-1$ (rather than n) makes the estimate unbiased — for large n, it makes little difference.
- Other measures of dispersion: **Variance**, **Range**, **Interquartile Range (IQR)**, **Median Absolute Deviation (MAD)**, **Average Absolute Deviation (AAD)**.
- The MAD is more robust to outliers than the STD.

Measures of Dispersion

- **Population standard deviation:** σ_Y.
- A population STD may or may not exist. For a population of size N,

$$\sigma_Y = \sqrt{\frac{\sum_{i=1}^{N}(Y_i - \overline{Y})^2}{N}}$$

- **Sample standard deviation:** $\hat{\sigma}_Y$ or s_Y.
- Measures how far away, on average, a random observation Y_i is from the sample mean $\overline{Y}$,

$$\hat{\sigma}_Y = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \overline{Y})^2}{n-1}}$$

  The hat is a short-hand for "estimated." Division by $n-1$ (rather than n) makes the estimate unbiased — for large n, it makes little difference.
- Other measures of dispersion: **Variance**, **Range**, **Interquartile Range (IQR)**, **Median Absolute Deviation (MAD)**, **Average Absolute Deviation (AAD)**.
- The MAD is more robust to outliers than the STD.

Normal Distribution

Normal Distribution

- **Normal Distribution:** A continuous distribution with a symmetric bell-shaped probability density function. The normal density with mean μ and variance σ^2 is symmetric around μ and has 95% of its probability between $\mu - 1.96\sigma$ and $\mu + 1.96\sigma$. A normally distributed random variable is uniquely defined by its mean and variance and is denoted

$$Y \sim N(\mu, \sigma^2)$$

  The two parameters μ and σ^2 are sufficient to completely describe any normal distribution.

- **Standard Normal Distribution:** The special case with mean zero and unit variance, $N(0, 1)$. The standard normal cumulative distribution is often denoted Φ, that is for a fixed value z,

$$\Pr(Z \leq z) = \Phi(z)$$

- All normal distributions can be converted to the standard normal by standardization,

$$Y \sim N(\mu, \sigma^2) \implies Z = \frac{Y - \mu}{\sigma} \sim N(0, 1)$$

Normal Distribution

- **Normal Distribution:** A continuous distribution with a symmetric bell-shaped probability density function. The normal density with mean μ and variance σ^2 is symmetric around μ and has 95% of its probability between $\mu - 1.96\sigma$ and $\mu + 1.96\sigma$. A normally distributed random variable is uniquely defined by its mean and variance and is denoted

$$Y \sim N(\mu, \sigma^2)$$

  The two parameters μ and σ^2 are sufficient to completely describe any normal distribution.

- **Standard Normal Distribution:** The special case with mean zero and unit variance, $N(0, 1)$. The standard normal cumulative distribution is often denoted Φ, that is for a fixed value z,

$$\Pr(Z \leq z) = \Phi(z)$$

- All normal distributions can be converted to the standard normal by standardization,

$$Y \sim N(\mu, \sigma^2) \implies Z = \frac{Y - \mu}{\sigma} \sim N(0, 1)$$

Normal Distribution

- **Normal Distribution:** A continuous distribution with a symmetric bell-shaped probability density function. The normal density with mean μ and variance σ^2 is symmetric around μ and has 95% of its probability between $\mu - 1.96\sigma$ and $\mu + 1.96\sigma$. A normally distributed random variable is uniquely defined by its mean and variance and is denoted

$$Y \sim N(\mu, \sigma^2)$$

  The two parameters μ and σ^2 are sufficient to completely describe any normal distribution.

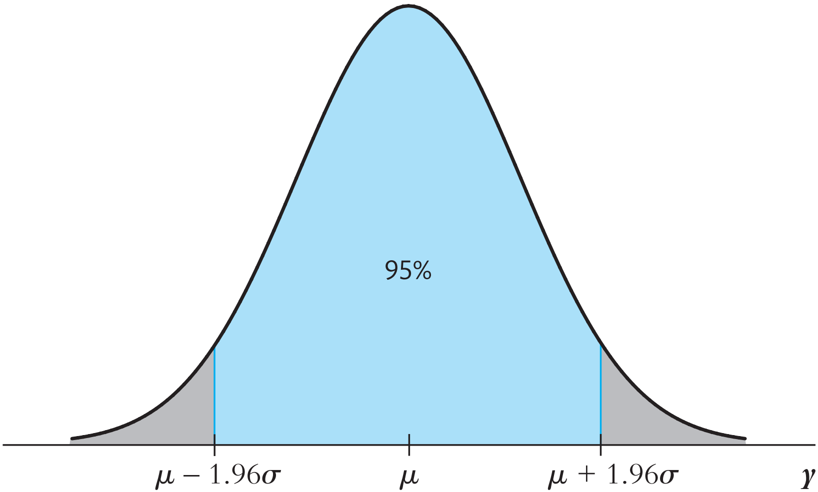
- **Standard Normal Distribution:** The special case with mean zero and unit variance, $N(0, 1)$. The standard normal cumulative distribution is often denoted Φ, that is for a fixed value z,

$$\Pr(Z \leq z) = \Phi(z)$$

- All normal distributions can be converted to the standard normal by standardization,

$$Y \sim N(\mu, \sigma^2) \implies Z = \frac{Y - \mu}{\sigma} \sim N(0, 1)$$

Normal Distribution: Thin Tails



The Normal distribution has "thin tails": very little probability lies in the tails, and so few outliers are expected to occur.

Chi-Square Distribution

Chi-Square Distribution

- **Chi-Square Distribution:** Distribution of the sum of m squared independent standard normal random variables. This distribution is parameterized by the "degrees of freedom" m.
- Named after the Greek letter "chi" χ.
- Let Z_1, Z_2, Z_3 be independent standard normal random variables. Then $Z_1^2 + Z_2^2 + Z_3^2$ has a chi-squared distribution with 3 degrees of freedom.

$$Z_1, Z_2, Z_3 \sim iid\, N(0,1) \implies Z_1^2 + Z_2^2 + Z_3^2 \sim \chi_3^2$$

- The 95th percentile of the χ_3^2 distribution is 7.81, so

$$\Pr(Z_1^2 + Z_2^2 + Z_3^2 \leq 7.81) = 0.95$$

Chi-Square Distribution

- **Chi-Square Distribution:** Distribution of the sum of m squared independent standard normal random variables. This distribution is parameterized by the “degrees of freedom” m.
- Named after the Greek letter “chi” χ.
- Let Z_1, Z_2, Z_3 be independent standard normal random variables. Then $Z_1^2 + Z_2^2 + Z_3^2$ has a chi-squared distribution with 3 degrees of freedom.

$$Z_1, Z_2, Z_3 \sim iid\, N(0,1) \implies Z_1^2 + Z_2^2 + Z_3^2 \sim \chi_3^2$$

- The 95th percentile of the χ_3^2 distribution is 7.81, so

$$\Pr(Z_1^2 + Z_2^2 + Z_3^2 \leq 7.81) = 0.95$$

Chi-Square Distribution

- **Chi-Square Distribution:** Distribution of the sum of m squared independent standard normal random variables. This distribution is parameterized by the "degrees of freedom" m.
- Named after the Greek letter "chi" χ.
- Let Z_1, Z_2, Z_3 be independent standard normal random variables. Then $Z_1^2 + Z_2^2 + Z_3^2$ has a chi-squared distribution with 3 degrees of freedom.

$$Z_1, Z_2, Z_3 \sim iid\, N(0,1) \implies Z_1^2 + Z_2^2 + Z_3^2 \sim \chi_3^2$$

- The 95th percentile of the χ_3^2 distribution is 7.81, so

$$\Pr(Z_1^2 + Z_2^2 + Z_3^2 \leq 7.81) = 0.95$$

Chi-Square Distribution

- **Chi-Square Distribution:** Distribution of the sum of m squared independent standard normal random variables. This distribution is parameterized by the "degrees of freedom" m.
- Named after the Greek letter "chi" χ.
- Let Z_1, Z_2, Z_3 be independent standard normal random variables. Then $Z_1^2 + Z_2^2 + Z_3^2$ has a chi-squared distribution with 3 degrees of freedom.

$$Z_1, Z_2, Z_3 \sim iid\, N(0,1) \implies Z_1^2 + Z_2^2 + Z_3^2 \sim \chi_3^2$$

- The 95th percentile of the χ_3^2 distribution is 7.81, so

$$\Pr(Z_1^2 + Z_2^2 + Z_3^2 \leq 7.81) = 0.95$$

Student-t Distribution

Student-t Distribution

- **Student-t Distribution:** The distribution of the ratio of a standard normal random variable to the square root of an independently distributed chi-squared random variable!
- Named after Statistician William Sealy Gosset, who used the pseudonym "Student."
- The Student-t distribution is the theoretical distribution of a standardized normal distribution when the standard deviation used in the standardization procedure is estimated from the sample data. The distribution is parameterized by the "degrees of freedom".
- Let Z be a standard normal random variable, let W be a chi-squared random variable with m degrees of freedom, with Z and W independently distributed, then

$$Z \sim N(0,1), \quad W \sim \chi^2(m), \quad Z \perp W \implies \frac{Z}{\sqrt{W/m}} \sim t(m)$$

- The Student-t distribution has a bell shape similar to that of the normal distribution, but with "fatter" tails. As the degrees of freedom are increased, the Student-t distribution tends to the standard normal distribution. For values of m larger than 20 there is little practical difference between the two distributions.

Student-t Distribution

- **Student-t Distribution:** The distribution of the ratio of a standard normal random variable to the square root of an independently distributed chi-squared random variable!
- Named after Statistician William Sealy Gosset, who used the pseudonym "Student."
- The Student-t distribution is the theoretical distribution of a standardized normal distribution when the standard deviation used in the standardization procedure is estimated from the sample data. The distribution is parameterized by the "degrees of freedom".
- Let Z be a standard normal random variable, let W be a chi-squared random variable with m degrees of freedom, with Z and W independently distributed, then

$$Z \sim N(0,1), \quad W \sim \chi^2(m), \quad Z \perp W \implies \frac{Z}{\sqrt{W/m}} \sim t(m)$$

- The Student-t distribution has a bell shape similar to that of the normal distribution, but with "fatter" tails. As the degrees of freedom are increased, the Student-t distribution tends to the standard normal distribution. For values of m larger than 20 there is little practical difference between the two distributions.

Student-t Distribution

- **Student-t Distribution:** The distribution of the ratio of a standard normal random variable to the square root of an independently distributed chi-squared random variable!
- Named after Statistician William Sealy Gosset, who used the pseudonym "Student."
- The Student-t distribution is the theoretical distribution of a standardized normal distribution when the standard deviation used in the standardization procedure is estimated from the sample data. The distribution is parameterized by the "degrees of freedom".
- Let Z be a standard normal random variable, let W be a chi-squared random variable with m degrees of freedom, with Z and W independently distributed, then

$$Z \sim N(0,1), \quad W \sim \chi^2(m), \quad Z \perp W \implies \frac{Z}{\sqrt{W/m}} \sim t(m)$$

- The Student-t distribution has a bell shape similar to that of the normal distribution, but with "fatter" tails. As the degrees of freedom are increased, the Student-t distribution tends to the standard normal distribution. For values of m larger than 20 there is little practical difference between the two distributions.

Student-t Distribution

- **Student-t Distribution:** The distribution of the ratio of a standard normal random variable to the square root of an independently distributed chi-squared random variable!
- Named after Statistician William Sealy Gosset, who used the pseudonym "Student."
- The Student-t distribution is the theoretical distribution of a standardized normal distribution when the standard deviation used in the standardization procedure is estimated from the sample data. The distribution is parameterized by the "degrees of freedom".
- Let Z be a standard normal random variable, let W be a chi-squared random variable with m degrees of freedom, with Z and W independently distributed, then

$$Z \sim N(0,1), \quad W \sim \chi^2(m), \quad Z \perp\!\!\!\perp W \implies \frac{Z}{\sqrt{W/m}} \sim t(m)$$

- The Student-t distribution has a bell shape similar to that of the normal distribution, but with "fatter" tails. As the degrees of freedom are increased, the Student-t distribution tends to the standard normal distribution. For values of m larger than 20 there is little practical difference between the two distributions.

Student-t Distribution

- **Student-t Distribution:** The distribution of the ratio of a standard normal random variable to the square root of an independently distributed chi-squared random variable!
- Named after Statistician William Sealy Gosset, who used the pseudonym "Student."
- The Student-t distribution is the theoretical distribution of a standardized normal distribution when the standard deviation used in the standardization procedure is estimated from the sample data. The distribution is parameterized by the "degrees of freedom".
- Let Z be a standard normal random variable, let W be a chi-squared random variable with m degrees of freedom, with Z and W independently distributed, then

$$Z \sim N(0,1), \quad W \sim \chi^2(m), \quad Z \perp\!\!\!\perp W \implies \frac{Z}{\sqrt{W/m}} \sim t(m)$$

- The Student-t distribution has a bell shape similar to that of the normal distribution, but with "fatter" tails. As the degrees of freedom are increased, the Student-t distribution tends to the standard normal distribution. For values of m larger than 20 there is little practical difference between the two distributions.

Fisher’s F Distribution

Fisher's F Distribution

- **F Distribution:** The distribution of the ratio of two independently distributed chi-squared random variables. Used to conduct F-tests, particularly in the context of Analysis of Variance (ANOVA) and when comparing the fits of different linear regression models, e.g. Chow test.
- Named after Statistician Ronald Aylmer Fisher.
- Let V be a chi-squared random variable with n degrees of freedom, let W be a chi-squared random variable with m degrees of freedom, with V and W independently distributed, then

$$W \sim \chi^2(m), \quad V \sim \chi^2(n), \quad W \perp V \implies F = \frac{W/m}{V/n} \sim F(m,n)$$

- Special case:

$$F(m, n \to \infty) \to \chi^2(m)$$

- The 95th percentile of the $F(3,30)$ distribution is 2.92; The 95th percentile of the $F(3,90)$ distribution is 2.71. In the limit, as $n \to \infty$, the 95th percentile tends to 2.60.

Fisher's F Distribution

- **F Distribution:** The distribution of the ratio of two independently distributed chi-squared random variables. Used to conduct F-tests, particularly in the context of Analysis of Variance (ANOVA) and when comparing the fits of different linear regression models, e.g. Chow test.
- Named after Statistician Ronald Aylmer Fisher.
- Let V be a chi-squared random variable with n degrees of freedom, let W be a chi-squared random variable with m degrees of freedom, with V and W independently distributed, then

$$W \sim \chi^2(m), \quad V \sim \chi^2(n), \quad W \perp V \implies F = \frac{W/m}{V/n} \sim F(m, n)$$

- Special case:

$$F(m, n \to \infty) \to \chi^2(m)$$

- The 95th percentile of the $F(3, 30)$ distribution is 2.92; The 95th percentile of the $F(3, 90)$ distribution is 2.71. In the limit, as $n \to \infty$, the 95th percentile tends to 2.60.

Fisher's F Distribution

- **F Distribution:** The distribution of the ratio of two independently distributed chi-squared random variables. Used to conduct F-tests, particularly in the context of Analysis of Variance (ANOVA) and when comparing the fits of different linear regression models, e.g. Chow test.
- Named after Statistician Ronald Aylmer Fisher.
- Let V be a chi-squared random variable with n degrees of freedom, let W be a chi-squared random variable with m degrees of freedom, with V and W independently distributed, then

$$W \sim \chi^2(m), \quad V \sim \chi^2(n), \quad W \perp\!\!\!\perp V \implies F = \frac{W/m}{V/n} \sim F(m,n)$$

- Special case:

$$F(m, n \to \infty) \to \chi^2(m)$$

- The 95th percentile of the $F(3,30)$ distribution is 2.92; The 95th percentile of the $F(3,90)$ distribution is 2.71. In the limit, as $n \to \infty$, the 95th percentile tends to 2.60.

Fisher's F Distribution

- **F Distribution:** The distribution of the ratio of two independently distributed chi-squared random variables. Used to conduct F-tests, particularly in the context of Analysis of Variance (ANOVA) and when comparing the fits of different linear regression models, e.g. Chow test.
- Named after Statistician Ronald Aylmer Fisher.
- Let V be a chi-squared random variable with n degrees of freedom, let W be a chi-squared random variable with m degrees of freedom, with V and W independently distributed, then

$$W \sim \chi^2(m), \quad V \sim \chi^2(n), \quad W \perp\!\!\!\perp V \implies F = \frac{W/m}{V/n} \sim F(m,n)$$

- Special case:

$$F(m, n \to \infty) \to \chi^2(m)$$

- The 95th percentile of the $F(3,30)$ distribution is 2.92; The 95th percentile of the $F(3,90)$ distribution is 2.71. In the limit, as $n \to \infty$, the 95th percentile tends to 2.60.

Fisher's F Distribution

- **F Distribution:** The distribution of the ratio of two independently distributed chi-squared random variables. Used to conduct F-tests, particularly in the context of Analysis of Variance (ANOVA) and when comparing the fits of different linear regression models, e.g. Chow test.
- Named after Statistician Ronald Aylmer Fisher.
- Let V be a chi-squared random variable with n degrees of freedom, let W be a chi-squared random variable with m degrees of freedom, with V and W independently distributed, then

$$W \sim \chi^2(m), \quad V \sim \chi^2(n), \quad W \perp\!\!\!\perp V \implies F = \frac{W/m}{V/n} \sim F(m,n)$$

- Special case:

$$F(m, n \to \infty) \to \chi^2(m)$$

- The 95th percentile of the $F(3,30)$ distribution is 2.92; The 95th percentile of the $F(3,90)$ distribution is 2.71. In the limit, as $n \to \infty$, the 95th percentile tends to 2.60.

Outliers

Outliers

- **Outlier:** A data point that differs significantly from other observations.
- Outliers are expected to occur in random data, but distributions differ in how frequently outliers are observed and how far they lie from the center of the distribution.
- The probability of an outlier drawn from a normal distribution is small — not zero, but small.
- **Stock Market Crash:** On "Black Monday" (October 19, 1987), the Dow Jones Industrial Average fell 22.6%. By comparison, the standard deviation of daily percentage price changes on the DJIA had been about one-percent over the previous twenty years!
- A drop of 22.6% was a negative return of 21 standard deviations $22.6/1.08$. If daily percentage price changes were normally distributed, then the probability of a change of at least 21 standard deviations would be

$$\Pr(|Z| \geq 21) = 2\Phi(-21) \approx 6.6 \times 10^{-98}$$

- Stock price percentage changes have a distribution with heavier tails than the normal distribution!

Outliers

- **Outlier:** A data point that differs significantly from other observations.
- Outliers are expected to occur in random data, but distributions differ in how frequently outliers are observed and how far they lie from the center of the distribution.
- The probability of an outlier drawn from a normal distribution is small — not zero, but small.
- **Stock Market Crash:** On "Black Monday" (October 19, 1987), the Dow Jones Industrial Average fell 22.6%. By comparison, the standard deviation of daily percentage price changes on the DJIA had been about one-percent over the previous twenty years!
- A drop of 22.6% was a negative return of 21 standard deviations $22.6/1.08$. If daily percentage price changes were normally distributed, then the probability of a change of at least 21 standard deviations would be

$$\Pr(|Z| \geq 21) = 2\Phi(-21) \approx 6.6 \times 10^{-98}$$

- Stock price percentage changes have a distribution with heavier tails than the normal distribution!

Outliers

- **Outlier:** A data point that differs significantly from other observations.
- Outliers are expected to occur in random data, but distributions differ in how frequently outliers are observed and how far they lie from the center of the distribution.
- The probability of an outlier drawn from a normal distribution is small — not zero, but small.
- **Stock Market Crash:** On "Black Monday" (October 19, 1987), the Dow Jones Industrial Average fell 22.6%. By comparison, the standard deviation of daily percentage price changes on the DJIA had been about one-percent over the previous twenty years!
- A drop of 22.6% was a negative return of 21 standard deviations $22.6/1.08$. If daily percentage price changes were normally distributed, then the probability of a change of at least 21 standard deviations would be

$$\Pr(|Z| \geq 21) = 2\Phi(-21) \approx 6.6 \times 10^{-98}$$

- Stock price percentage changes have a distribution with heavier tails than the normal distribution!

Outliers

- **Outlier:** A data point that differs significantly from other observations.
- Outliers are expected to occur in random data, but distributions differ in how frequently outliers are observed and how far they lie from the center of the distribution.
- The probability of an outlier drawn from a normal distribution is small — not zero, but small.
- **Stock Market Crash:** On "Black Monday" (October 19, 1987), the Dow Jones Industrial Average fell 22.6%. By comparison, the standard deviation of daily percentage price changes on the DJIA had been about one-percent over the previous twenty years!
- A drop of 22.6% was a negative return of 21 standard deviations $22.6/1.08$. If daily percentage price changes were normally distributed, then the probability of a change of at least 21 standard deviations would be

$$\Pr(|Z| \geq 21) = 2\Phi(-21) \approx 6.6 \times 10^{-98}$$

- Stock price percentage changes have a distribution with heavier tails than the normal distribution!

Outliers

- **Outlier:** A data point that differs significantly from other observations.
- Outliers are expected to occur in random data, but distributions differ in how frequently outliers are observed and how far they lie from the center of the distribution.
- The probability of an outlier drawn from a normal distribution is small — not zero, but small.
- **Stock Market Crash:** On “Black Monday” (October 19, 1987), the Dow Jones Industrial Average fell 22.6%. By comparison, the standard deviation of daily percentage price changes on the DJIA had been about one-percent over the previous twenty years!
- A drop of 22.6% was a negative return of 21 standard deviations $22.6/1.08$. If daily percentage price changes were normally distributed, then the probability of a change of at least 21 standard deviations would be

$$\Pr(|Z| \geq 21) = 2\Phi(-21) \approx 6.6 \times 10^{-98}$$

- Stock price percentage changes have a distribution with heavier tails than the normal distribution!

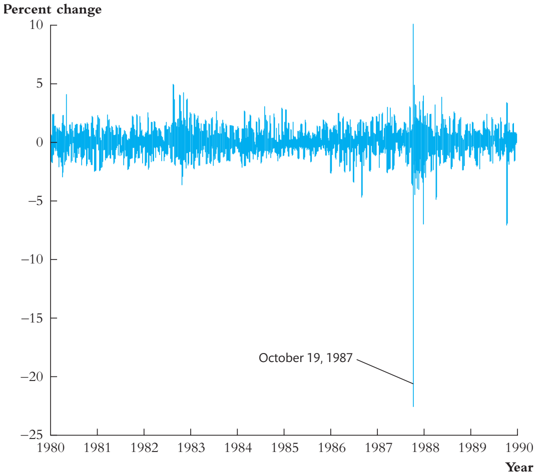
Outliers

- **Outlier:** A data point that differs significantly from other observations.
- Outliers are expected to occur in random data, but distributions differ in how frequently outliers are observed and how far they lie from the center of the distribution.
- The probability of an outlier drawn from a normal distribution is small — not zero, but small.
- **Stock Market Crash:** On "Black Monday" (October 19, 1987), the Dow Jones Industrial Average fell 22.6%. By comparison, the standard deviation of daily percentage price changes on the DJIA had been about one-percent over the previous twenty years!
- A drop of 22.6% was a negative return of 21 standard deviations $22.6/1.08$. If daily percentage price changes were normally distributed, then the probability of a change of at least 21 standard deviations would be

$$\Pr(|Z| \geq 21) = 2\Phi(-21) \approx 6.6 \times 10^{-98}$$

- Stock price percentage changes have a distribution with heavier tails than the normal distribution!

Outliers: The Stock Market



From January 1980 through September 2017, the average percentage daily change of “the Dow” index was 0.04% and its standard deviation was 1.08%. On October 19, 1987—“Black Monday”—the Dow fell 22.6%, or 21 standard deviations.

Daily Percentage Changes in the Dow Jones Industrial Average in the 1980s.

Outliers: The Stock Market

Date	Percentage Change (x)	Standardized Change $z = (x - \mu)/\sigma$	Normal Probability of a Change at Least This Large $\Pr(\lvert Z \rvert \geq \lvert z \rvert) = 2\Phi(-\lvert z \rvert)$
October 19, 1987	–22.6	–21.0	6.6×10^{-98}
October 13, 2008	11.1	10.2	1.5×10^{-24}
October 28, 2008	10.9	10.0	1.0×10^{-23}
October 21, 1987	10.1	9.4	7.7×10^{-21}
October 26, 1987	–8.0	–7.5	7.2×10^{-14}
October 15, 2008	–7.9	–7.3	2.3×10^{-13}
December 01, 2008	–7.7	–7.2	7.4×10^{-13}
October 09, 2008	–7.3	–6.8	8.5×10^{-12}
October 27, 1997	–7.2	–6.7	2.2×10^{-11}
September 17, 2001	–7.1	–6.6	3.1×10^{-11}

Ten Largest Daily Percentage Changes in the Dow Jones Industrial Average, January 1980-September 2017, with the Normal Probability of a Change at Least as Large.

Sampling Distribution

Random Sampling

- **Simple Random Sampling:** A fixed number of objects are selected from a population, with each member of the population equally likely to be included in the sample.
- Let n observations in a sample of size n be denoted $Y_1, \ldots, Y_n$. Because these random variables are independently drawn from the same population using the same selection procedure, they are said to be *independently and identically distributed (i.i.d).*
- **Sampling Distribution:** The *sample mean* of the n observations is:

$$\overline{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$$

- **Law of Large Numbers:** If $Y_1, \ldots, Y_n$ are independently and identically distributed from a common population with mean μ_Y, and if large outliers are unlikely (the distribution has finite variance), then the sample mean converges in probability to the population mean,

$$\overline{Y} \xrightarrow{p} \mu_Y$$

- If $\overline{Y}$ converges "in probability" to μ_Y, then $\overline{Y}$ is said to be "consistent" for μ_Y. It means that as the sample size n increases, the sample mean $\overline{Y}$ lies inside any arbitrary interval around μ_Y with probability 1.

Random Sampling

- **Simple Random Sampling:** A fixed number of objects are selected from a population, with each member of the population equally likely to be included in the sample.
- Let n observations in a sample of size n be denoted $Y_1, \ldots, Y_n$. Because these random variables are independently drawn from the same population using the same selection procedure, they are said to be *independently and identically distributed (i.i.d).*
- **Sampling Distribution:** The *sample mean* of the n observations is:

$$\overline{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$$

- **Law of Large Numbers:** If $Y_1, \ldots, Y_n$ are independently and identically distributed from a common population with mean μ_Y, and if large outliers are unlikely (the distribution has finite variance), then the sample mean converges in probability to the population mean,

$$\overline{Y} \xrightarrow{p} \mu_Y$$

- If $\overline{Y}$ converges "in probability" to μ_Y, then $\overline{Y}$ is said to be "consistent" for μ_Y. It means that as the sample size n increases, the sample mean $\overline{Y}$ lies inside any arbitrary interval around μ_Y with probability 1.

Random Sampling

- **Simple Random Sampling:** A fixed number of objects are selected from a population, with each member of the population equally likely to be included in the sample.
- Let n observations in a sample of size n be denoted $Y_1, \ldots, Y_n$. Because these random variables are independently drawn from the same population using the same selection procedure, they are said to be *independently and identically distributed (i.i.d)*.
- **Sampling Distribution:** The *sample mean* of the n observations is:

$$\overline{Y} = \frac{1}{n} \sum_{i=1}^{n} Y_i$$

- **Law of Large Numbers:** If $Y_1, \ldots, Y_n$ are independently and identically distributed from a common population with mean μ_Y, and if large outliers are unlikely (the distribution has finite variance), then the sample mean converges in probability to the population mean,

$$\overline{Y} \xrightarrow{p} \mu_Y$$

- If $\overline{Y}$ converges "in probability" to μ_Y, then $\overline{Y}$ is said to be "consistent" for μ_Y. It means that as the sample size n increases, the sample mean $\overline{Y}$ lies inside any arbitrary interval around μ_Y with probability 1.

Random Sampling

- **Simple Random Sampling:** A fixed number of objects are selected from a population, with each member of the population equally likely to be included in the sample.
- Let n observations in a sample of size n be denoted $Y_1, \ldots, Y_n$. Because these random variables are independently drawn from the same population using the same selection procedure, they are said to be *independently and identically distributed (i.i.d)*.
- **Sampling Distribution:** The *sample mean* of the n observations is:

$$\overline{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$$

- **Law of Large Numbers:** If $Y_1, \ldots, Y_n$ are independently and identically distributed from a common population with mean μ_Y, and if large outliers are unlikely (the distribution has finite variance), then the sample mean converges in probability to the population mean,

$$\overline{Y} \xrightarrow{p} \mu_Y$$

- If $\overline{Y}$ converges “in probability” to μ_Y, then $\overline{Y}$ is said to be “consistent” for μ_Y. It means that as the sample size n increases, the sample mean $\overline{Y}$ lies inside any arbitrary interval around μ_Y with probability 1.

Random Sampling

- **Simple Random Sampling:** A fixed number of objects are selected from a population, with each member of the population equally likely to be included in the sample.
- Let n observations in a sample of size n be denoted $Y_1, \ldots, Y_n$. Because these random variables are independently drawn from the same population using the same selection procedure, they are said to be *independently and identically distributed (i.i.d)*.
- **Sampling Distribution:** The *sample mean* of the n observations is:

$$\overline{Y} = \frac{1}{n} \sum_{i=1}^{n} Y_i$$

- **Law of Large Numbers:** If $Y_1, \ldots, Y_n$ are independently and identically distributed from a common population with mean μ_Y, and if large outliers are unlikely (the distribution has finite variance), then the sample mean converges in probability to the population mean,

$$\overline{Y} \xrightarrow{p} \mu_Y$$

- If $\overline{Y}$ converges "in probability" to μ_Y, then $\overline{Y}$ is said to be "consistent" for μ_Y. It means that as the sample size n increases, the sample mean $\overline{Y}$ lies inside any arbitrary interval around μ_Y with probability 1.

Random Sampling

- Because the sample is drawn at random, the sample mean $\overline{Y}$ is a random variable. Its distribution is called the "sampling distribution" and satisfies:

$$\mathrm{E}[\overline{Y}] = \frac{1}{n}\sum_{i=1}^{n}\mathrm{E}[Y_i] = \mu_Y$$

$$\begin{aligned}
\mathrm{var}(\overline{Y}) &= \mathrm{var}\left(\frac{1}{n}\sum_{i=1}^{n}Y_i\right) \\
&= \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{var}(Y_i) + \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n}\mathrm{cov}(Y_i,Y_j) \\
&= \frac{1}{n^2}\,n\,\mathrm{var}(Y) \\
&= \frac{\sigma_Y^2}{n}
\end{aligned}$$

Sampling Distribution

- **Sampling from a Normal Distribution:**

$$Y_1, \ldots, Y_n \sim iid\, N\left(\mu_Y, \sigma_Y^2\right) \implies \overline{Y} \sim \left(\mu_Y, \frac{\sigma_Y^2}{n}\right)$$

- **Sampling from Any Distribution:**

$$Y_1, \ldots, Y_n \sim iid\, f\left(\mu_Y, \sigma_Y^2\right) \implies \overline{Y} \sim \left(\mu_Y, \frac{\sigma_Y^2}{n}\right) \quad \text{as } n \to \infty$$

- This result is known as the “Central Limit Theorem.”
- Simulations are a great way to visualize the theorem.

Sampling Distribution

- **Sampling from a Normal Distribution:**

$$Y_1, \ldots, Y_n \sim iid\, N\left(\mu_Y, \sigma_Y^2\right) \implies \overline{Y} \sim \left(\mu_Y, \frac{\sigma_Y^2}{n}\right)$$

- **Sampling from Any Distribution:**

$$Y_1, \ldots, Y_n \sim iid\, f\left(\mu_Y, \sigma_Y^2\right) \implies \overline{Y} \sim \left(\mu_Y, \frac{\sigma_Y^2}{n}\right) \quad \text{as } n \to \infty$$

- This result is known as the “Central Limit Theorem.”
- Simulations are a great way to visualize the theorem.

- **Sampling from a Normal Distribution:**

$$Y_1, \ldots, Y_n \sim iid\, N\left(\mu_Y, \sigma_Y^2\right) \implies \overline{Y} \sim \left(\mu_Y, \frac{\sigma_Y^2}{n}\right)$$

- **Sampling from Any Distribution:**

$$Y_1, \ldots, Y_n \sim iid\, f\left(\mu_Y, \sigma_Y^2\right) \implies \overline{Y} \sim \left(\mu_Y, \frac{\sigma_Y^2}{n}\right) \quad \text{as } n \to \infty$$

- This result is known as the "Central Limit Theorem."
- Simulations are a great way to visualize the theorem.

Sampling Distribution

- **Sampling from a Normal Distribution:**

$$Y_1, \ldots, Y_n \sim iid\, N\left(\mu_Y, \sigma_Y^2\right) \implies \overline{Y} \sim \left(\mu_Y, \frac{\sigma_Y^2}{n}\right)$$

- **Sampling from Any Distribution:**

$$Y_1, \ldots, Y_n \sim iid\, f\left(\mu_Y, \sigma_Y^2\right) \implies \overline{Y} \sim \left(\mu_Y, \frac{\sigma_Y^2}{n}\right) \quad \text{as } n \to \infty$$

- This result is known as the "Central Limit Theorem."
- Simulations are a great way to visualize the theorem.

Central Limit Theorem

Central Limit Theorem

- **Central Limit Theorem (CLT)**: Under general conditions, the distribution of $\overline{Y}$ for large n is well approximated by a normal distribution.
- Let $Y_1, \ldots, Y_n$ be i.i.d. random variables drawn from a population with mean μ_Y and finite variance σ_Y^2. The mean and variance of the sample mean $\overline{Y}$ are

$$\mathrm{E}[\overline{Y}] = \mu_Y$$

$$\sigma_{\overline{Y}}^2 = \frac{\sigma_Y^2}{n}$$

- For large sample size n, the distribution of the sample mean is approximately

$$\overline{Y} \sim N\left(\mu_Y, \frac{\sigma_Y^2}{n}\right)$$

- For large samples, only the mean and the variance are needed to describe the sampling distribution. In practice, the population parameters are typically unknown and the realized observations are used to estimate them.

Central Limit Theorem

- **Central Limit Theorem (CLT)**: Under general conditions, the distribution of $\overline{Y}$ for large n is well approximated by a normal distribution.
- Let $Y_1, \ldots, Y_n$ be i.i.d. random variables drawn from a population with mean μ_Y and finite variance σ_Y^2. The mean and variance of the sample mean $\overline{Y}$ are

$$\mathrm{E}[\overline{Y}] = \mu_Y$$

$$\sigma_{\overline{Y}}^2 = \frac{\sigma_Y^2}{n}$$

- For large sample size n, the distribution of the sample mean is approximately

$$\overline{Y} \sim N\left(\mu_Y, \frac{\sigma_Y^2}{n}\right)$$

- For large samples, only the mean and the variance are needed to describe the sampling distribution. In practice, the population parameters are typically unknown and the realized observations are used to estimate them.

Central Limit Theorem

- **Central Limit Theorem (CLT)**: Under general conditions, the distribution of $\overline{Y}$ for large n is well approximated by a normal distribution.
- Let $Y_1, \ldots, Y_n$ be i.i.d. random variables drawn from a population with mean μ_Y and finite variance σ_Y^2. The mean and variance of the sample mean $\overline{Y}$ are

$$\mathrm{E}[\overline{Y}] = \mu_Y$$

$$\sigma_{\overline{Y}}^2 = \frac{\sigma_Y^2}{n}$$

- For large sample size n, the distribution of the sample mean is approximately

$$\overline{Y} \sim N\left(\mu_Y, \frac{\sigma_Y^2}{n}\right)$$

- For large samples, only the mean and the variance are needed to describe the sampling distribution. In practice, the population parameters are typically unknown and the realized observations are used to estimate them.

Central Limit Theorem

- **Central Limit Theorem (CLT)**: Under general conditions, the distribution of $\overline{Y}$ for large n is well approximated by a normal distribution.
- Let $Y_1, \ldots, Y_n$ be i.i.d. random variables drawn from a population with mean μ_Y and finite variance σ_Y^2. The mean and variance of the sample mean $\overline{Y}$ are

$$\mathrm{E}[\overline{Y}] = \mu_Y$$
$$\sigma_{\overline{Y}}^2 = \frac{\sigma_Y^2}{n}$$

- For large sample size n, the distribution of the sample mean is approximately

$$\overline{Y} \sim N\left(\mu_Y, \frac{\sigma_Y^2}{n}\right)$$

- For large samples, only the mean and the variance are needed to describe the sampling distribution. In practice, the population parameters are typically unknown and the realized observations are used to estimate them.

Summary

Summary

1. The probabilities associated with a random variable are summarized by the cumulative distribution function, the probability distribution function — for discrete random variables — and the probability density function — for continuous random variables.
2. The expected value of a random variable Y, denoted $\mathrm{E}(Y)$, is its probability-weighted average value. The variance of Y is $\sigma_Y^2 = \mathrm{E}[(Y - \mu_Y)^2]$, and the standard deviation of Y is the square root of its variance.
3. The joint probabilities for two random variables, X and Y, are summarized by their joint probability distribution. The conditional probability distribution of Y given $X = x$ is the probability distribution of Y, conditional on X taking on the value x.
4. A normally distributed random variable has a bell-shaped probability density.

Summary

1. The probabilities associated with a random variable are summarized by the cumulative distribution function, the probability distribution function — for discrete random variables — and the probability density function — for continuous random variables.
2. The expected value of a random variable Y, denoted $\mathrm{E}(Y)$, is its probability-weighted average value. The variance of Y is $\sigma_Y^2 = \mathrm{E}[(Y - \mu_Y)^2]$, and the standard deviation of Y is the square root of its variance.
3. The joint probabilities for two random variables, X and Y, are summarized by their joint probability distribution. The conditional probability distribution of Y given $X = x$ is the probability distribution of Y, conditional on X taking on the value x.
4. A normally distributed random variable has a bell-shaped probability density.

Summary

1. The probabilities associated with a random variable are summarized by the cumulative distribution function, the probability distribution function — for discrete random variables — and the probability density function — for continuous random variables.
2. The expected value of a random variable Y, denoted $\mathrm{E}(Y)$, is its probability-weighted average value. The variance of Y is $\sigma_Y^2 = \mathrm{E}[(Y - \mu_Y)^2]$, and the standard deviation of Y is the square root of its variance.
3. The joint probabilities for two random variables, X and Y, are summarized by their joint probability distribution. The conditional probability distribution of Y given $X = x$ is the probability distribution of Y, conditional on X taking on the value x.
4. A normally distributed random variable has a bell-shaped probability density.

Summary

1. The probabilities associated with a random variable are summarized by the cumulative distribution function, the probability distribution function — for discrete random variables — and the probability density function — for continuous random variables.
2. The expected value of a random variable Y, denoted $\mathrm{E}(Y)$, is its probability-weighted average value. The variance of Y is $\sigma_Y^2 = \mathrm{E}[(Y - \mu_Y)^2]$, and the standard deviation of Y is the square root of its variance.
3. The joint probabilities for two random variables, X and Y, are summarized by their joint probability distribution. The conditional probability distribution of Y given $X = x$ is the probability distribution of Y, conditional on X taking on the value x.
4. A normally distributed random variable has a bell-shaped probability density.

Summary

5. Simple random sampling produces n random observations, $Y_1, \ldots, Y_n$, that are independently and identically distributed (i.i.d.).

6. The sample average, $\overline{Y}$, is a random variable with a sampling distribution. If $Y_1, \ldots, Y_n$ are i.i.d., then

Summary

5. Simple random sampling produces n random observations, $Y_1, \ldots, Y_n$, that are independently and identically distributed (i.i.d.).
6. The sample average, $\overline{Y}$, is a random variable with a sampling distribution. If $Y_1, \ldots, Y_n$ are i.i.d., then
   - The sampling distribution of $\overline{Y}$ has mean μ_Y and variance $\sigma^2_{\overline{Y}} = \sigma^2_Y/n$;
   - The law of large numbers states that $\overline{Y}$ converges in probability to μ_Y.
   - By the central limit theorem, $\overline{Y}$ has a normal distribution for large n,

$$\frac{\overline{Y} - \mu_Y}{\sigma_Y/\sqrt{n}} \sim N(0,1)$$

   or equivalently

$$\overline{Y} \sim N(\mu_Y, \sigma^2_Y/n)$$

Summary

5. Simple random sampling produces n random observations, $Y_1, \ldots, Y_n$, that are independently and identically distributed (i.i.d.).
6. The sample average, $\overline{Y}$, is a random variable with a sampling distribution. If $Y_1, \ldots, Y_n$ are i.i.d., then
   - The sampling distribution of $\overline{Y}$ has mean μ_Y and variance $\sigma^2_{\overline{Y}} = \sigma^2_Y/n$;
   - The law of large numbers states that $\overline{Y}$ converges in probability to μ_Y.
   - By the central limit theorem, $\overline{Y}$ has a normal distribution for large n,

$$\frac{\overline{Y} - \mu_Y}{\sigma_Y/\sqrt{n}} \sim N(0,1)$$

   or equivalently

$$\overline{Y} \sim N(\mu_Y, \sigma^2_Y/n)$$

Summary

5. Simple random sampling produces n random observations, $Y_1, \ldots, Y_n$, that are independently and identically distributed (i.i.d.).
6. The sample average, $\overline{Y}$, is a random variable with a sampling distribution. If $Y_1, \ldots, Y_n$ are i.i.d., then
   - The sampling distribution of $\overline{Y}$ has mean μ_Y and variance $\sigma^2_{\overline{Y}} = \sigma^2_Y / n$;
   - The law of large numbers states that $\overline{Y}$ converges in probability to μ_Y.
   - By the central limit theorem, $\overline{Y}$ has a normal distribution for large n,

     $$\frac{\overline{Y} - \mu_Y}{\sigma_Y / \sqrt{n}} \sim N(0, 1)$$

     or equivalently

     $$\overline{Y} \sim N(\mu_Y, \sigma^2_Y / n)$$

Summary

5. Simple random sampling produces n random observations, $Y_1, \ldots, Y_n$, that are independently and identically distributed (i.i.d.).
6. The sample average, $\overline{Y}$, is a random variable with a sampling distribution. If $Y_1, \ldots, Y_n$ are i.i.d., then
   - The sampling distribution of $\overline{Y}$ has mean μ_Y and variance $\sigma^2_{\overline{Y}} = \sigma^2_Y / n$;
   - The law of large numbers states that $\overline{Y}$ converges in probability to μ_Y.
   - By the central limit theorem, $\overline{Y}$ has a normal distribution for large n,

$$\frac{\overline{Y} - \mu_Y}{\sigma_Y / \sqrt{n}} \sim N(0, 1)$$

or equivalently

$$\overline{Y} \sim N(\mu_Y, \sigma^2_Y / n)$$

Problems & Applications

Problems and Applications

Stock & Watson, Introduction (4th), Chapter 2, Exercise 1.

Let Y denote the number of "heads" that occur when two coins are tossed.

1. Derive the probability distribution of Y.
2. Derive the cumulative probability distribution of Y.
3. Derive the mean and variance of Y.

Reading Probability tables AND writing code, compute the following:

Problems and Applications

Stock & Watson, Introduction (4th), Chapter 2, Exercise 1.

Let Y denote the number of "heads" that occur when two coins are tossed.

1. Derive the probability distribution of Y.
2. Derive the cumulative probability distribution of Y.
3. Derive the mean and variance of Y.

Reading Probability tables AND writing code, compute the following:

1. If $Y \sim N(1,4)$, compute $\Pr(Y \leq 3)$.
2. If $Y \sim N(50,25)$, compute $\Pr(40 \leq Y \leq 52)$.
3. If $Y \sim \chi^2_{10}$, compute $\Pr(Y > 18.31)$.
4. If $Y \sim t_{15}$, compute $\Pr(Y > 1.75)$.
5. If $Y \sim F_{7,4}$, compute $\Pr(Y > 2.79)$.
6. If $\mu_Y = 100$ and $\sigma^2_Y = 43$, use the central limit theorem to compute $\Pr(101 \leq \bar{Y} \leq 103)$ for a random sample of size $n = 64$.

Problems and Applications

Stock & Watson, Introduction (4th), Chapter 2, Exercise 1.

Let Y denote the number of “heads” that occur when two coins are tossed.

1. Derive the probability distribution of Y.
2. Derive the cumulative probability distribution of Y.
3. Derive the mean and variance of Y.

Reading Probability tables AND writing code, compute the following:

Problems and Applications

Stock & Watson, Introduction (4th), Chapter 2, Exercise 1.

Let Y denote the number of “heads” that occur when two coins are tossed.

1. Derive the probability distribution of Y.
2. Derive the cumulative probability distribution of Y.
3. Derive the mean and variance of Y.

Reading Probability tables AND writing code, compute the following:

Problems and Applications

Stock & Watson, Introduction (4th), Chapter 2, Exercise 1.

Let Y denote the number of "heads" that occur when two coins are tossed.

1. Derive the probability distribution of Y.
2. Derive the cumulative probability distribution of Y.
3. Derive the mean and variance of Y.

Reading Probability tables AND writing code, compute the following:

1. If $Y \sim N(1, 4)$, compute $\Pr(Y \leq 3)$.
2. If $Y \sim N(50, 25)$, compute $\Pr(40 \leq Y \leq 52)$.
3. If $Y \sim \chi^2_{10}$, compute $\Pr(Y > 18.31)$.
4. If $Y \sim t_{15}$, compute $\Pr(Y > 1.75)$.
5. If $Y \sim F_{7,4}$, compute $\Pr(Y > 2.79)$.
6. If $\mu_Y = 100$ and $\sigma^2_Y = 43$, use the central limit theorem to compute $\Pr(101 \leq \bar{Y} \leq 103)$ for a random sample of size $n = 64$.

Problems and Applications

Stock & Watson, Introduction (4th), Chapter 2, Exercise 1.

Let Y denote the number of "heads" that occur when two coins are tossed.

1. Derive the probability distribution of Y.
2. Derive the cumulative probability distribution of Y.
3. Derive the mean and variance of Y.

Reading Probability tables AND writing code, compute the following:

1. If $Y \sim N(1, 4)$, compute $\Pr(Y \leq 3)$.
2. If $Y \sim N(50, 25)$, compute $\Pr(40 \leq Y \leq 52)$.
3. If $Y \sim \chi^2_{10}$, compute $\Pr(Y > 18.31)$.
4. If $Y \sim t_{15}$, compute $\Pr(Y > 1.75)$.
5. If $Y \sim F_{7,4}$, compute $\Pr(Y > 2.79)$.
6. If $\mu_Y = 100$ and $\sigma^2_Y = 43$, use the central limit theorem to compute $\Pr(101 \leq Y \leq 103)$ for a random sample of size $n = 64$.

Problems and Applications

Stock & Watson, Introduction (4th), Chapter 2, Exercise 1.

Let Y denote the number of "heads" that occur when two coins are tossed.

1. Derive the probability distribution of Y.
2. Derive the cumulative probability distribution of Y.
3. Derive the mean and variance of Y.

Reading Probability tables AND writing code, compute the following:

1. If $Y \sim N(1, 4)$, compute $\Pr(Y \leq 3)$.
2. If $Y \sim N(50, 25)$, compute $\Pr(40 \leq Y \leq 52)$.
3. If $Y \sim \chi^2_{10}$, compute $\Pr(Y > 18.31)$.
4. If $Y \sim t_{15}$, compute $\Pr(Y > 1.75)$.
5. If $Y \sim F_{7,4}$, compute $\Pr(Y > 2.79)$.
6. If $\mu_Y = 100$ and $\sigma^2_Y = 43$, use the central limit theorem to compute $\Pr(101 \leq Y \leq 103)$ for a random sample of size $n = 64$.

Problems and Applications

Stock & Watson, Introduction (4th), Chapter 2, Exercise 1.

Let Y denote the number of "heads" that occur when two coins are tossed.

1. Derive the probability distribution of Y.
2. Derive the cumulative probability distribution of Y.
3. Derive the mean and variance of Y.

Reading Probability tables AND writing code, compute the following:

1. If $Y \sim N(1, 4)$, compute $\Pr(Y \leq 3)$.
2. If $Y \sim N(50, 25)$, compute $\Pr(40 \leq Y \leq 52)$.
3. If $Y \sim \chi^2_{10}$, compute $\Pr(Y > 18.31)$.
4. If $Y \sim t_{15}$, compute $\Pr(Y > 1.75)$.
5. If $Y \sim F_{7,4}$, compute $\Pr(Y > 2.79)$.
6. If $\mu_Y = 100$ and $\sigma^2_Y = 43$, use the central limit theorem to compute $\Pr(101 \leq Y \leq 103)$ for a random sample of size $n = 64$.

Problems and Applications

Stock & Watson, Introduction (4th), Chapter 2, Exercise 1.

Let Y denote the number of “heads” that occur when two coins are tossed.

1. Derive the probability distribution of Y.
2. Derive the cumulative probability distribution of Y.
3. Derive the mean and variance of Y.

Reading Probability tables AND writing code, compute the following:

1. If $Y \sim N(1, 4)$, compute $\Pr(Y \leq 3)$.
2. If $Y \sim N(50, 25)$, compute $\Pr(40 \leq Y \leq 52)$.
3. If $Y \sim \chi^2_{10}$, compute $\Pr(Y > 18.31)$.
4. If $Y \sim t_{15}$, compute $\Pr(Y > 1.75)$.
5. If $Y \sim F_{7,4}$, compute $\Pr(Y > 2.79)$.
6. If $\mu_Y = 100$ and $\sigma^2_Y = 43$, use the central limit theorem to compute $\Pr(101 \leq \bar{Y} \leq 103)$ for a random sample of size $n = 64$.

Problems and Applications

Stock & Watson, Introduction (4th), Chapter 2, Exercise 1.

Let Y denote the number of "heads" that occur when two coins are tossed.

1. Derive the probability distribution of Y.
2. Derive the cumulative probability distribution of Y.
3. Derive the mean and variance of Y.

Reading Probability tables AND writing code, compute the following:

1. If $Y \sim N(1, 4)$, compute $\Pr(Y \leq 3)$.
2. If $Y \sim N(50, 25)$, compute $\Pr(40 \leq Y \leq 52)$.
3. If $Y \sim \chi^2_{10}$, compute $\Pr(Y > 18.31)$.
4. If $Y \sim t_{15}$, compute $\Pr(Y > 1.75)$.
5. If $Y \sim F_{7,4}$, compute $\Pr(Y > 2.79)$.
6. If $\mu_Y = 100$ and $\sigma^2_Y = 43$, use the central limit theorem to compute $\Pr(101 \leq Y \leq 103)$ for a random sample of size $n = 64$.

Problems and Applications

Stock & Watson, Introduction (4th), Chapter 2, Exercise 6.

The following table gives the joint probability distribution between employment status and college graduation among those either employed or looking for work (unemployed) in the working-age U.S. population for September 2017.

Employment & College Graduation (Population aged 25 and above, September 2017)

	Unemployed $Y = 0$	Employed $Y = 1$	Total
Non-College Graduates ($X = 0$)	0.026	0.576	0.602
College Graduates ($X = 1$)	0.009	0.389	0.398
Total	0.035	0.965	1.000

1. Compute $E(Y)$.
2. The unemployment rate is the fraction of the labor force that is unemployed. Show that the unemployment rate is given by $1 - E(Y)$.
3. Calculate $E(Y|X = 1)$ and $E(Y|X = 0)$.
4. Calculate the unemployment rate for (i) college graduates and (ii) non-college graduates.
5. A randomly selected member of this population reports being unemployed. What is the probability that this worker is a college graduate? A non–college graduate?
6. Are educational achievement and employment status independent? Explain.

Problems and Applications

Stock & Watson, Introduction (4th), Chapter 2, Exercise 6.

The following table gives the joint probability distribution between employment status and college graduation among those either employed or looking for work (unemployed) in the working-age U.S. population for September 2017.

Employment & College Graduation (Population aged 25 and above, September 2017)

	Unemployed $Y = 0$	Employed $Y = 1$	Total
Non-College Graduates ($X = 0$)	0.026	0.576	0.602
College Graduates ($X = 1$)	0.009	0.389	0.398
Total	0.035	0.965	1.000

1. Compute $E(Y)$.
2. The unemployment rate is the fraction of the labor force that is unemployed. Show that the unemployment rate is given by $1 - E(Y)$.
3. Calculate $E(Y|X = 1)$ and $E(Y|X = 0)$.
4. Calculate the unemployment rate for (i) college graduates and (ii) non-college graduates.
5. A randomly selected member of this population reports being unemployed. What is the probability that this worker is a college graduate? A non–college graduate?
6. Are educational achievement and employment status independent? Explain.

Problems and Applications

Stock & Watson, Introduction (4th), Chapter 2, Exercise 6.

The following table gives the joint probability distribution between employment status and college graduation among those either employed or looking for work (unemployed) in the working-age U.S. population for September 2017.

Employment & College Graduation (Population aged 25 and above, September 2017)

	Unemployed $Y = 0$	Employed $Y = 1$	Total
Non-College Graduates ($X = 0$)	0.026	0.576	0.602
College Graduates ($X = 1$)	0.009	0.389	0.398
Total	0.035	0.965	1.000

1. Compute $E(Y)$.
2. The unemployment rate is the fraction of the labor force that is unemployed. Show that the unemployment rate is given by $1 - E(Y)$.
3. Calculate $E(Y|X = 1)$ and $E(Y|X = 0)$.
4. Calculate the unemployment rate for (i) college graduates and (ii) non-college graduates.
5. A randomly selected member of this population reports being unemployed. What is the probability that this worker is a college graduate? A non–college graduate?
6. Are educational achievement and employment status independent? Explain.

Problems and Applications

Stock & Watson, Introduction (4th), Chapter 2, Exercise 6.

The following table gives the joint probability distribution between employment status and college graduation among those either employed or looking for work (unemployed) in the working-age U.S. population for September 2017.

Employment & College Graduation (Population aged 25 and above, September 2017)

	Unemployed $Y = 0$	Employed $Y = 1$	Total
Non-College Graduates ($X = 0$)	0.026	0.576	0.602
College Graduates ($X = 1$)	0.009	0.389	0.398
Total	0.035	0.965	1.000

1. Compute $E(Y)$.
2. The unemployment rate is the fraction of the labor force that is unemployed. Show that the unemployment rate is given by $1 - E(Y)$.
3. Calculate $E(Y|X = 1)$ and $E(Y|X = 0)$.
4. Calculate the unemployment rate for (i) college graduates and (ii) non-college graduates.
5. A randomly selected member of this population reports being unemployed. What is the probability that this worker is a college graduate? A non–college graduate?
6. Are educational achievement and employment status independent? Explain.

Problems and Applications

Stock & Watson, Introduction (4th), Chapter 2, Exercise 6.

The following table gives the joint probability distribution between employment status and college graduation among those either employed or looking for work (unemployed) in the working-age U.S. population for September 2017.

Employment & College Graduation (Population aged 25 and above, September 2017)

	Unemployed $Y = 0$	Employed $Y = 1$	Total
Non-College Graduates ($X = 0$)	0.026	0.576	0.602
College Graduates ($X = 1$)	0.009	0.389	0.398
Total	0.035	0.965	1.000

1. Compute $E(Y)$.
2. The unemployment rate is the fraction of the labor force that is unemployed. Show that the unemployment rate is given by $1 - E(Y)$.
3. Calculate $E(Y|X = 1)$ and $E(Y|X = 0)$.
4. Calculate the unemployment rate for (i) college graduates and (ii) non-college graduates.
5. A randomly selected member of this population reports being unemployed. What is the probability that this worker is a college graduate? A non–college graduate?
6. Are educational achievement and employment status independent? Explain.

Problems and Applications

Stock & Watson, Introduction (4th), Chapter 2, Exercise 6.

The following table gives the joint probability distribution between employment status and college graduation among those either employed or looking for work (unemployed) in the working-age U.S. population for September 2017.

Employment & College Graduation (Population aged 25 and above, September 2017)

	Unemployed $Y = 0$	Employed $Y = 1$	Total
Non-College Graduates ($X = 0$)	0.026	0.576	0.602
College Graduates ($X = 1$)	0.009	0.389	0.398
Total	0.035	0.965	1.000

1. Compute $E(Y)$.
2. The unemployment rate is the fraction of the labor force that is unemployed. Show that the unemployment rate is given by $1 - E(Y)$.
3. Calculate $E(Y|X = 1)$ and $E(Y|X = 0)$.
4. Calculate the unemployment rate for (i) college graduates and (ii) non-college graduates.
5. A randomly selected member of this population reports being unemployed. What is the probability that this worker is a college graduate? A non–college graduate?
6. Are educational achievement and employment status independent? Explain.

Problems and Applications

Stock & Watson, Introduction (4th), Chapter 2, Exercise 9.

X and Y are discrete random variables with the following joint distribution:

		Value of Y				
		14	22	30	40	65
	1	0.02	0.05	0.10	0.03	0.01
Value of X	5	0.17	0.15	0.05	0.02	0.01
	8	0.02	0.03	0.15	0.10	0.09

That is, $\Pr(X = 1, Y = 14) = 0.02$, and so forth.

1. Calculate the probability distribution, mean, and variance of Y.
2. Calculate the probability distribution, mean, and variance of Y given $X = 8$.
3. Calculate the covariance and correlation between X and Y.

Problems and Applications

Stock & Watson, Introduction (4th), Chapter 2, Exercise 9.

X and Y are discrete random variables with the following joint distribution:

		Value of Y				
		14	22	30	40	65
	1	0.02	0.05	0.10	0.03	0.01
Value of X	5	0.17	0.15	0.05	0.02	0.01
	8	0.02	0.03	0.15	0.10	0.09

That is, $\Pr(X = 1, Y = 14) = 0.02$, and so forth.

1. Calculate the probability distribution, mean, and variance of Y.
2. Calculate the probability distribution, mean, and variance of Y given $X = 8$.
3. Calculate the covariance and correlation between X and Y.

Problems and Applications

Stock & Watson, Introduction (4th), Chapter 2, Exercise 9.

X and Y are discrete random variables with the following joint distribution:

		Value of Y				
		14	22	30	40	65
	1	0.02	0.05	0.10	0.03	0.01
Value of X	5	0.17	0.15	0.05	0.02	0.01
	8	0.02	0.03	0.15	0.10	0.09

That is, $\Pr(X = 1, Y = 14) = 0.02$, and so forth.

1. Calculate the probability distribution, mean, and variance of Y.
2. Calculate the probability distribution, mean, and variance of Y given $X = 8$.
3. Calculate the covariance and correlation between X and Y.

Problems and Applications

1. Guinevere has two children, one of them a girl. What is the probability that the other child is also a girl?
2. You flip a fair coin 10 times. Use the definition of the mathematical expectation for a discrete distribution to calculate the expected number of heads.
3. You flip a biased coin 10 times. The probability of head is 0.9 on a single flip. Calculate the expected number of heads.
4. You roll a die until a six comes up. What is the expected number of rolls?
5. You choose an integer X at random between 1 and 10. Then you choose an integer Y at random between 1 and $X = x$. Calculate the expected value of Y.
6. Six numbers are randomly selected from the discrete set $\{1, 2, \ldots, 100\}$. You bet $\$1$ to draw exactly six given numbers. What is the expected value of the bet if the prize is 1 million? For what prize value does the bet "break even"? (In other words, What prize value gives an expectation of $\$1$, so that you are just as likely to profit as you are to lose)

Problems and Applications

1. Guinevere has two children, one of them a girl. What is the probability that the other child is also a girl?
2. You flip a fair coin 10 times. Use the definition of the mathematical expectation for a discrete distribution to calculate the expected number of heads.
3. You flip a biased coin 10 times. The probability of head is 0.9 on a single flip. Calculate the expected number of heads.
4. You roll a die until a six comes up. What is the expected number of rolls?
5. You choose an integer X at random between 1 and 10. Then you choose an integer Y at random between 1 and $X = x$. Calculate the expected value of Y.
6. Six numbers are randomly selected from the discrete set $\{1, 2, \ldots, 100\}$. You bet $\$1$ to draw exactly six given numbers. What is the expected value of the bet if the prize is 1 million? For what prize value does the bet "break even"? (In other words, What prize value gives an expectation of $\$1$, so that you are just as likely to profit as you are to lose)

Problems and Applications

1. Guinevere has two children, one of them a girl. What is the probability that the other child is also a girl?
2. You flip a fair coin 10 times. Use the definition of the mathematical expectation for a discrete distribution to calculate the expected number of heads.
3. You flip a biased coin 10 times. The probability of head is 0.9 on a single flip. Calculate the expected number of heads.
4. You roll a die until a six comes up. What is the expected number of rolls?
5. You choose an integer X at random between 1 and 10. Then you choose an integer Y at random between 1 and $X = x$. Calculate the expected value of Y.
6. Six numbers are randomly selected from the discrete set $\{1, 2, \dots, 100\}$. You bet $\$1$ to draw exactly six given numbers. What is the expected value of the bet if the prize is 1 million? For what prize value does the bet "break even"? (In other words, What prize value gives an expectation of $\$1$, so that you are just as likely to profit as you are to lose)

Problems and Applications

1. Guinevere has two children, one of them a girl. What is the probability that the other child is also a girl?
2. You flip a fair coin 10 times. Use the definition of the mathematical expectation for a discrete distribution to calculate the expected number of heads.
3. You flip a biased coin 10 times. The probability of head is 0.9 on a single flip. Calculate the expected number of heads.
4. You roll a die until a six comes up. What is the expected number of rolls?
5. You choose an integer X at random between 1 and 10. Then you choose an integer Y at random between 1 and $X = x$. Calculate the expected value of Y.
6. Six numbers are randomly selected from the discrete set $\{1, 2, \ldots, 100\}$. You bet $\$1$ to draw exactly six given numbers. What is the expected value of the bet if the prize is 1 million? For what prize value does the bet “break even”? (In other words, What prize value gives an expectation of $\$1$, so that you are just as likely to profit as you are to lose)

Problems and Applications

1. Guinevere has two children, one of them a girl. What is the probability that the other child is also a girl?
2. You flip a fair coin 10 times. Use the definition of the mathematical expectation for a discrete distribution to calculate the expected number of heads.
3. You flip a biased coin 10 times. The probability of head is 0.9 on a single flip. Calculate the expected number of heads.
4. You roll a die until a six comes up. What is the expected number of rolls?
5. You choose an integer X at random between 1 and 10. Then you choose an integer Y at random between 1 and $X = x$. Calculate the expected value of Y.
6. Six numbers are randomly selected from the discrete set $\{1, 2, \ldots, 100\}$. You bet $\$1$ to draw exactly six given numbers. What is the expected value of the bet if the prize is 1 million? For what prize value does the bet "break even"? (In other words, What prize value gives an expectation of $\$1$, so that you are just as likely to profit as you are to lose)

Problems and Applications

1. Guinevere has two children, one of them a girl. What is the probability that the other child is also a girl?
2. You flip a fair coin 10 times. Use the definition of the mathematical expectation for a discrete distribution to calculate the expected number of heads.
3. You flip a biased coin 10 times. The probability of head is 0.9 on a single flip. Calculate the expected number of heads.
4. You roll a die until a six comes up. What is the expected number of rolls?
5. You choose an integer X at random between 1 and 10. Then you choose an integer Y at random between 1 and $X = x$. Calculate the expected value of Y.
6. Six numbers are randomly selected from the discrete set $\{1, 2, \ldots, 100\}$. You bet $\$1$ to draw exactly six given numbers. What is the expected value of the bet if the prize is 1 million? For what prize value does the bet “break even”? (In other words, What prize value gives an expectation of $\$1$, so that you are just as likely to profit as you are to lose)