Mathematics for Machine Learning

Vazgen Mikayelyan

September 1, 2020

σ-algebra

Definition

Let Ω *be some set and* $\mathcal{F}$ *be a set of some subsets of* Ω.

σ-algebra

Definition

Let Ω be some set and $\mathcal{F}$ be a set of some subsets of Ω. $\mathcal{F}$ is called a σ-algebra if it satisfies the following three properties:

σ-algebra

Definition

Let Ω be some set and $\mathcal{F}$ be a set of some subsets of Ω. $\mathcal{F}$ is called a σ-algebra if it satisfies the following three properties:

- $\Omega \in \mathcal{F}$,

σ-algebra

Definition

Let Ω be some set and $\mathcal{F}$ be a set of some subsets of Ω. $\mathcal{F}$ is called a σ-algebra if it satisfies the following three properties:

- $\Omega \in \mathcal{F}$,
- *if* $A \in \mathcal{F}$, *then* $\Omega \setminus A \in \mathcal{F}$,

σ-algebra

Definition

Let Ω be some set and $\mathcal{F}$ be a set of some subsets of Ω. $\mathcal{F}$ is called a σ-algebra if it satisfies the following three properties:

- $\Omega \in \mathcal{F}$,
- *if* $A \in \mathcal{F}$, *then* $\Omega \setminus A \in \mathcal{F}$,
- *if* $A_n \in \mathcal{F}$, $n \in \mathbb{N}$, *then* $\bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$.

Probability (Measure) Definition

Probability (Measure) Definition

Probability Measure Definition

A function $\mathbb{P} : \mathcal{F} \to \mathbb{R}$ is called a **Probability Measure** on $(\Omega, \mathcal{F})$, if it satisfies the following axioms:

Probability (Measure) Definition

Probability Measure Definition

A function $\mathbb{P} : \mathcal{F} \to \mathbb{R}$ is called a **Probability Measure** on $(\Omega, \mathcal{F})$, if it satisfies the following axioms:

P1. For any $A \in \mathcal{F}$, $\mathbb{P}(A) \geq 0$;

Probability (Measure) Definition

Probability Measure Definition

A function $\mathbb{P} : \mathcal{F} \to \mathbb{R}$ is called a **Probability Measure** on $(\Omega, \mathcal{F})$, if it satisfies the following axioms:

P1. For any $A \in \mathcal{F}$, $\mathbb{P}(A) \geq 0$;

P2. $\mathbb{P}(\Omega) = 1$;

Probability (Measure) Definition

Probability Measure Definition

A function $\mathbb{P} : \mathcal{F} \to \mathbb{R}$ is called a **Probability Measure** on $(\Omega, \mathcal{F})$, if it satisfies the following axioms:

P1. For any $A \in \mathcal{F}$, $\mathbb{P}(A) \geq 0$;

P2. $\mathbb{P}(\Omega) = 1$;

P3. For any sequence of pairwise mutually exclusive (disjoint) events $A_n \in \mathcal{F}$, i.e., for any sequence $A_n \in \mathcal{F}$ with $A_i \cap A_j = \varnothing$ for $i \neq j$, we have

$$\mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mathbb{P}(A_n).$$

Probability (Measure) Definition

Probability Measure is very similar (and shares the properties of) any other Measure -

- Cardinality (no. of elements),
- Length (in 1D),
- Area (in 2D),
- Volume (in 3D and moreD).

Probability (Measure) Definition

Probability Measure is very similar (and shares the properties of) any other Measure -

- Cardinality (no. of elements),
- Length (in 1D),
- Area (in 2D),
- Volume (in 3D and moreD).

The difference is only that the Probability of the Sample Space is 1, $\mathbb{P}(\Omega) = 1$.

Properties of the Probability Measure

1. $\mathbb{P}(\varnothing) = 0$;

Properties of the Probability Measure

1. $\mathbb{P}(\varnothing) = 0$;
2. if $A, B \in \mathcal{F}$ are mutually exclusive events, i.e., if $A \cap B = \varnothing$, then

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B);$$

Properties of the Probability Measure

1. $\mathbb{P}(\varnothing) = 0$;
2. if $A, B \in \mathcal{F}$ are mutually exclusive events, i.e., if $A \cap B = \varnothing$, then

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B);$$

3. for any event $A \in \mathcal{F}$,

$$\mathbb{P}(\overline{A}) = 1 - \mathbb{P}(A);$$

Here $\overline{A} = A^c = \Omega \setminus A$.

Properties of the Probability Measure

4. If $A_1, A_2, ..., A_n \in \mathcal{F}$ are pairwise disjoint (mutually exclusive), i.e., if $A_i \cap A_j = \varnothing$ for $i \neq j$, then

$$\mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} \mathbb{P}(A_i);$$

Properties of the Probability Measure

4. If $A_1, A_2, ..., A_n \in \mathcal{F}$ are pairwise disjoint (mutually exclusive), i.e., if $A_i \cap A_j = \varnothing$ for $i \neq j$, then

$$\mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} \mathbb{P}(A_i);$$

5. for any events $A, B \in \mathcal{F}$ (not necessarily disjoint),

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B);$$

Classical Probability Models: Discrete Sample Spaces

Discrete = Finite or Countably Infinite

Classical Probability Models: Discrete Sample Spaces

Discrete = Finite or Countably Infinite

To give Probability Models, Probability Spaces, we need to give:

- The Sample Space Ω;

Classical Probability Models: Discrete Sample Spaces

Discrete = Finite or Countably Infinite

To give Probability Models, Probability Spaces, we need to give:

- The Sample Space Ω;
- The set of Events $\mathcal{F}$;

Classical Probability Models: Discrete Sample Spaces

Discrete = Finite or Countably Infinite

To give Probability Models, Probability Spaces, we need to give:

- The Sample Space Ω;
- The set of Events $\mathcal{F}$;
- The Probability Measure $\mathbb{P}$.

Classical Probability Models: Finite Sample Spaces

Assume the Sample Space Ω is finite:

- Our Sample Space is $\Omega = \{\omega_1, \omega_2, ..., \omega_n\}$.

Classical Probability Models: Finite Sample Spaces

Assume the Sample Space Ω is finite:

- Our Sample Space is $\Omega = \{\omega_1, \omega_2, ..., \omega_n\}$.
- Every subset of Ω is an Event, i.e., $\mathcal{F} = 2^\Omega$, the power set of Ω.

Classical Probability Models: Finite Sample Spaces

Assume the Sample Space Ω is finite:

- Our Sample Space is $\Omega = \{\omega_1, \omega_2, ..., \omega_n\}$.
- Every subset of Ω is an Event, i.e., $\mathcal{F} = 2^\Omega$, the power set of Ω.
- We take any real numbers $p_1, p_2, ..., p_n$ with

$$p_1 \geq 0, ..., p_n \geq 0, \quad p_1 + p_2 + ... + p_n = 1,$$

and define

$$\mathbb{P}(\{\omega_1\}) = p_1, \quad \mathbb{P}(\{\omega_2\}) = p_2, \quad ..., \quad \mathbb{P}(\{\omega_n\}) = p_n.$$

Classical Probability Models: Finite Sample Spaces

We write this in a more convenient table form:

Outcome	ω_1	ω_2	$\dots$	ω_n
$\mathbb{P}(\{\omega_k\})$	p_1	p_2	$\dots$	p_n

Classical Probability Models: Finite Sample Spaces

We write this in a more convenient table form:

Outcome	ω_1	ω_2	...	ω_n
$\mathbb{P}(\{\omega_k\})$	p_1	p_2	...	p_n

We define, for any event A,

$$\mathbb{P}(A) = \sum_{\omega_i \in A} p_i,$$

and also add $\mathbb{P}(\varnothing) = 0$.

Conditional Probability

Definition: Conditional Probability

Assume that $(\Omega, \mathcal{F}, \mathbb{P})$ is a Probability Space and A, B are two events such that $\mathbb{P}(B) \neq 0$. The conditional probability of A given B (or the probability of A under the condition of B) is defined to be

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}.$$

Conditional Probability

Definition: Conditional Probability

Assume that $(\Omega, \mathcal{F}, \mathbb{P})$ is a Probability Space and A, B are two events such that $\mathbb{P}(B) \neq 0$. The conditional probability of A given B (or the probability of A under the condition of B) is defined to be

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}.$$

Example

Suppose that somebody rolls two fair dice. Compute the probability that the value of the first one is 2, given the information that their sum is no greater than 5.

Conditional Probability

The Chain Rule (The multiplication rule)

Assume $B \subset \Omega$ is a fixed event and $\mathbb{P}(B) \neq 0$. Then

$$\mathbb{P}(A \cap B) = P(B)P(A|B).$$

More general

$$\mathbb{P}(E_1 \cap E_2 \cap E_3 \cap \ldots \cap E_n)$$

$$= \mathbb{P}(E_1)\mathbb{P}(E_2|E_1)\mathbb{P}(E_3|E_1 \cap E_2) \ldots \mathbb{P}(E_n|E_1 \cap \ldots \cap E_{n-1})$$

Independence

Natural (contextual) Definition of Independence

Events A and B of the same experiment are called independent if the occurrence of one of them is absolutely uninfluenced by the occurrence of second event. The events will be called dependent if they are not independent.

Independence

Natural (contextual) Definition of Independence

Events A and B of the same experiment are called independent if the occurrence of one of them is absolutely uninfluenced by the occurrence of second event. The events will be called dependent if they are not independent.

Formal (mathematical) Definition of Independence

Events A and B of the same experiment are called independent if

$$\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B),$$

The events will be called dependent if they are not independent.

Independence

Properties of independent events

Let A, B, C be some events of an experiment.

Properties of independent events

Let A, B, C be some events of an experiment.

1. If $\mathbb{P}(A) > 0$ then A and B are independent iff $\mathbb{P}(B|A) = \mathbb{P}(B)$.

Properties of independent events

Let A, B, C be some events of an experiment.

1. If $\mathbb{P}(A) > 0$ then A and B are independent iff $\mathbb{P}(B|A) = \mathbb{P}(B)$.
2. If A and B are independent events, where $\mathbb{P}(A) > 0$ and $\mathbb{P}(B) > 0$, then A and B be cannot be mutually exclusive.

Independence

Properties of independent events

Let A, B, C be some events of an experiment.

1. If $\mathbb{P}(A) > 0$ then A and B are independent iff $\mathbb{P}(B|A) = \mathbb{P}(B)$.
2. If A and B are independent events, where $\mathbb{P}(A) > 0$ and $\mathbb{P}(B) > 0$, then A and B be cannot be mutually exclusive.
3. If A and B are independent events, then so are $\overline{A}$ and B, also A and $\overline{B}$, and also $\overline{A}$ and $\overline{B}$.

Independence

Properties of independent events

Let A, B, C be some events of an experiment.

1. If $\mathbb{P}(A) > 0$ then A and B are independent iff $\mathbb{P}(B|A) = \mathbb{P}(B)$.
2. If A and B are independent events, where $\mathbb{P}(A) > 0$ and $\mathbb{P}(B) > 0$, then A and B be cannot be mutually exclusive.
3. If A and B are independent events, then so are $\overline{A}$ and B, also A and $\overline{B}$, and also $\overline{A}$ and $\overline{B}$.
4. Let B and C are mutually exclusive events. If A and B are independent, and A and C are also independent, then A and $B \cup C$ are independent, too.

Independence

Example

Pick at random a card from a deck of 52 cards. Then we return that card into the deck, and again pick at random a card. What is the probability that the first card will be an ace and the second card will be a number < 7*?*

Example

Pick at random a card from a deck of 52 cards. Then we return that card into the deck, and again pick at random a card. What is the probability that the first card will be an ace and the second card will be a number < 7*?*

Example

Experiment - 2 hunters shoot at a duck simultaneously. The hunters are equally good and hit the target with probability 0.7.
Event A - First hunter hits the target
Event B - Second hunter hits the target
What is the probability that the duck will survive (both hunters failed)?

Total Probability Formula, TPF

Theorem

Assume A is some event, and B is another event of the same experiment such that $\mathbb{P}(B) > 0$. Then the following formula is valid:

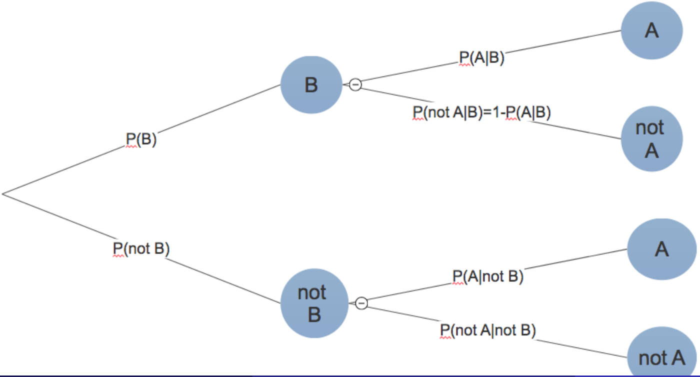
$$\mathbb{P}(A) = \mathbb{P}(B)\mathbb{P}(A|B) + \mathbb{P}(\overline{B})\mathbb{P}(A|\overline{B}).$$

Total Probability Formula, TPF

Theorem

Assume A is some event, and B is another event of the same experiment such that $\mathbb{P}(B) > 0$. Then the following formula is valid:

$$\mathbb{P}(A) = \mathbb{P}(B)\mathbb{P}(A|B) + \mathbb{P}(\overline{B})\mathbb{P}(A|\overline{B}).$$



Total Probability Formula, TPF

Example

6 green balls and 4 red balls (identical apart from color) are placed in a box. A child is asked to select a ball at random. Then the second child is asked to draw another ball from the box. What is the probability that the second child chooses a red ball?

General Total Probability Formula

Theorem

Let $B_1, B_2, \ldots, B_n$ *be some events of the same experiment such that they are pairwise disjoint, they contain any outcome that belongs to event A, i.e.* $A \subset \bigcup_{k=1}^{n} B_k$. *Then*

$$\mathbb{P}(A) = \sum_{k=1}^{n} \mathbb{P}(B_k)\mathbb{P}(A|B_k).$$

General Total Probability Formula

Theorem

Let $B_1, B_2, \ldots, B_n$ *be some events of the same experiment such that they are pairwise disjoint, they contain any outcome that belongs to event A, i.e.* $A \subset \bigcup_{k=1}^{n} B_k$*. Then*

$$\mathbb{P}(A) = \sum_{k=1}^{n} \mathbb{P}(B_k)\mathbb{P}(A|B_k).$$

Example

Assume we are rolling a die. If 1 is shown, we are tossing once a coin. If 2 is shown on the die, we are tossing a coin 2 times, and so on, if 6 is shown on the die top face, we are tossing a coin 6 times. What is the probability that we will have at least one head in this experiment?

Bayes' Formula (Bayes' Rule)

Bayes' Formula (Bayes' Rule)

$$\mathbb{P}(B|A) = \frac{\mathbb{P}(B)\mathbb{P}(A|B)}{\mathbb{P}(A)}.$$

Bayes' Formula (Bayes' Rule)

$$\mathbb{P}(B|A) = \frac{\mathbb{P}(B)\mathbb{P}(A|B)}{\mathbb{P}(A)}.$$

Probability, Statistics and Machine Learning experts give the following names to the terms in the Bayes Rule:

- $\mathbb{P}(B)$ - Prior Probability;
- $\mathbb{P}(A|B)$ - Likelihood;
- $\mathbb{P}(B|A)$ - Posterior Probability;
- $\mathbb{P}(A)$ - Evidence.

Bayes' Formula (Bayes' Rule)

$$\mathbb{P}(B|A) = \frac{\mathbb{P}(B)\mathbb{P}(A|B)}{\mathbb{P}(A)}.$$

Probability, Statistics and Machine Learning experts give the following names to the terms in the Bayes Rule:

- $\mathbb{P}(B)$ - Prior Probability;
- $\mathbb{P}(A|B)$ - Likelihood;
- $\mathbb{P}(B|A)$ - Posterior Probability;
- $\mathbb{P}(A)$ - Evidence.

$\mathbb{P}(B)$ measures the probability of B without any side information, before knowing that some other event happened. This is our Prior Probability. The probability $\mathbb{P}(B|A)$ is the update of our previous measure under the information that A happened, we update the probability after knowing that A happened - this is our Posterior Probability.

Bayes' Formula (Bayes' Rule)

Example

In some country, the risk to develop a lung cancer is 0.1%*. We know that* 20% *of total population are smokers, and the chance to develop a lung cancer for smokers is* 0.4%.
Problem 1. (by TPF): What is the probability that a randomly chosen non-smoker will develop a lung cancer?

Bayes' Formula (Bayes' Rule)

Example

In some country, the risk to develop a lung cancer is 0.1%*. We know that* 20% *of total population are smokers, and the chance to develop a lung cancer for smokers is* 0.4%.
Problem 1. (by TPF): What is the probability that a randomly chosen non-smoker will develop a lung cancer?
Problem 2. (by Bayes' Formula): Assume we choose a person randomly, and he has a lung cancer. What is the probability that he is a Smoker?

Bayes' Formula (Bayes' Rule)

If $\mathbb{P}(A) > 0$ and $0 < \mathbb{P}(B) < 1$ then

$$\mathbb{P}(B|A) = \frac{\mathbb{P}(B)\mathbb{P}(A|B)}{\mathbb{P}(B)\mathbb{P}(A|B) + \mathbb{P}(\overline{B})\mathbb{P}(A|\overline{B})}.$$

Bayes' Formula (Bayes' Rule)

If $\mathbb{P}(A) > 0$ and $0 < \mathbb{P}(B) < 1$ then

$$\mathbb{P}(B|A) = \frac{\mathbb{P}(B)\mathbb{P}(A|B)}{\mathbb{P}(B)\mathbb{P}(A|B) + \mathbb{P}(\overline{B})\mathbb{P}(A|\overline{B})}.$$

Bayes' Formula: Umbrella

If it is raining in the morning there is a 90% chance that I will bring my umbrella. If it is not raining in the morning there is only a 20% chance of me taking my umbrella. On any given morning the probability of rain is 0.1. If you see me with an umbrella, what is the probability that it was raining that morning?

Bayes' General Formula

Theorem

Assume we have an event A and mutually exclusive events (hypotheses) $B_1, B_2, \ldots, B_n$ *such that* $\mathbb{P}(A) > 0, \mathbb{P}(B_k) > 0, k = 1, 2, \ldots, n$, *and* $A \subset \bigcup_{k=1}^{n} B_k$. *Then, for any* $i \in \{1, 2, \ldots, n\}$, *the following formula takes place:*

$$\mathbb{P}(B_i|A) = \frac{\mathbb{P}(B_i)\mathbb{P}(A|B_i)}{\sum_{k=1}^{n} \mathbb{P}(B_k)\mathbb{P}(A|B_k)}.$$

Bayes' General Formula

Asian Suppliers

A computer manufacturer receives hard disks from three different supplier countries. Malaysia supplies 40% of the disks, Bangladesh supplies 25%, while Laos supplies the rest. Disks from Malaysia have a defective rate of 4%, those from Bangladesh 3%, while Laos have a 5% rate. A disk is checked and found defective. What is the probability that it was supplied from Laos?

Bayes' Formula

Medical Testing

Assume a person decides to take a medical test for some disease. The probability that the test gives a correct answer is 98%. This means that if a person has that disease, then the test gives positive result in 98% cases. And if a person did not have that disease, then the test gives negative result in 98% cases. It is known that only 0.1% of total population has that disease. Now, a person receives the test results, and the test shows a positive result. What is the probability that the person actually has that disease?

Bayes' Formula

Medical Testing

Assume a person decides to take a medical test for some disease. The probability that the test gives a correct answer is 98%. This means that if a person has that disease, then the test gives positive result in 98% cases. And if a person did not have that disease, then the test gives negative result in 98% cases. It is known that only 0.1% of total population has that disease. Now, a person receives the test results, and the test shows a positive result. What is the probability that the person actually has that disease?

Have a look at videos
https://www.youtube.com/watch?v=BrK7X_XlGB8
https://www.youtube.com/watch?v=R13BD8qKeTg

Random Variables

Definition

Let $(\Omega, \mathcal{F}, \mathbb{P})$ *be a probability space. If the function* $X : \Omega \to \mathbb{R}$ *satisfies to the condition*

$$\{\omega \in \Omega : X(\omega) \leq x\} \in \mathcal{F}$$

for all $x \in \mathbb{R}$*, then it is called a* **Random Variable.**

Random Variables

Definition

Let $(\Omega, \mathcal{F}, \mathbb{P})$ *be a probability space. If the function* $X : \Omega \to \mathbb{R}$ *satisfies to the condition*

$$\{\omega \in \Omega : X(\omega) \leq x\} \in \mathcal{F}$$

for all $x \in \mathbb{R}$*, then it is called a* **Random Variable.**

Discrete Random Variables

A random variable X is called **discrete** if the set of the values of X,

$$Range(X) = \{X(w) : w \in \Omega\}$$

is either finite, or countably infinite.

Random Variables

Probability Mass Function

For a discrete random variable X, we define the **probability mass function (PMF)** $p_X(x)$ of X by

$$p_X(x) = \mathbb{P}(\{w \in \Omega : X(w) = x\})$$

Random Variables

Probability Mass Function

For a discrete random variable X, we define the **probability mass function (PMF)** $p_X(x)$ of X by

$$p_X(x) = \mathbb{P}(\{w \in \Omega : X(w) = x\}) =: \mathbb{P}(X = x).$$

Probability Mass Function

For a discrete random variable X, we define the **probability mass function (PMF)** $p_X(x)$ of X by

$$p_X(x) = \mathbb{P}(\{w \in \Omega : X(w) = x\}) =: \mathbb{P}(X = x).$$

If $Range(X) = \{x_1, x_2, \ldots, x_k, \ldots\}$ then we usually denote

$$p_k = p_X(x_k) = \mathbb{P}(X = x_k).$$

Random Variables

Probability Mass Function

For a discrete random variable X, we define the **probability mass function (PMF)** $p_X(x)$ of X by

$$p_X(x) = \mathbb{P}(\{w \in \Omega : X(w) = x\}) =: \mathbb{P}(X = x).$$

If $Range(X) = \{x_1, x_2, \ldots, x_k, \ldots\}$ then we usually denote

$$p_k = p_X(x_k) = \mathbb{P}(X = x_k).$$

PMF presented in a tabular form.

X	x_1	x_2	...	x_n	...
$\mathbb{P}(X = x)$	p_1	p_2	...	p_n	...

Random Variables

Probability Mass Function

For a discrete random variable X, we define the **probability mass function (PMF)** $p_X(x)$ of X by

$$p_X(x) = \mathbb{P}(\{w \in \Omega : X(w) = x\}) =: \mathbb{P}(X = x).$$

If $Range(X) = \{x_1, x_2, \ldots, x_k, \ldots\}$ then we usually denote

$$p_k = p_X(x_k) = \mathbb{P}(X = x_k).$$

PMF presented in a tabular form.

X	x_1	x_2	...	x_n	$\ldots$
$\mathbb{P}(X = x)$	p_1	p_2	...	p_n	$\ldots$

Clearly $\sum p_n = 1$.

CDF of a Random Variable

Let Ω be a sample space of an experiment and $X : \Omega \to \mathbb{R}$ be a random variable. Then the function $F_X : \mathbb{R} \to [0, 1]$ defined by the formula

$$F_X(x) = \mathbb{P}(X \leq x)$$

is called the **cumulative distribution function (CDF)** of the r.v. X.

CDF of a Random Variable

Let Ω be a sample space of an experiment and $X : \Omega \to \mathbb{R}$ be a random variable. Then the function $F_X : \mathbb{R} \to [0, 1]$ defined by the formula

$$F_X(x) = \mathbb{P}(X \leq x)$$

is called the **cumulative distribution function (CDF)** of the r.v. X.

CDF Properties

For any r.v. X, its CDF, F_X , satisfies the following properties:

- $0 \leq F_X(x) \leq 1$, for any $x \in \mathbb{R}$.

CDF of a Random Variable

Let Ω be a sample space of an experiment and $X : \Omega \to \mathbb{R}$ be a random variable. Then the function $F_X : \mathbb{R} \to [0,1]$ defined by the formula

$$F_X(x) = \mathbb{P}(X \leq x)$$

is called the **cumulative distribution function (CDF)** of the r.v. X.

CDF Properties

For any r.v. X, its CDF, F_X , satisfies the following properties:

- $0 \leq F_X(x) \leq 1$, for any $x \in \mathbb{R}$.
- F_X is an increasing function on $\mathbb{R}$.

Random Variables

CDF of a Random Variable

Let Ω be a sample space of an experiment and $X : \Omega \to \mathbb{R}$ be a random variable. Then the function $F_X : \mathbb{R} \to [0, 1]$ defined by the formula

$$F_X(x) = \mathbb{P}(X \leq x)$$

is called the **cumulative distribution function (CDF)** of the r.v. X.

CDF Properties

For any r.v. X, its CDF, F_X , satisfies the following properties:

- $0 \leq F_X(x) \leq 1$, for any $x \in \mathbb{R}$.
- F_X is an increasing function on $\mathbb{R}$.
- $F_X(-\infty) = 0$ and $F_X(+\infty) = 1$.

Random Variables

CDF of a Random Variable

Let Ω be a sample space of an experiment and $X : \Omega \to \mathbb{R}$ be a random variable. Then the function $F_X : \mathbb{R} \to [0, 1]$ defined by the formula

$$F_X(x) = \mathbb{P}(X \leq x)$$

is called the **cumulative distribution function (CDF)** of the r.v. X.

CDF Properties

For any r.v. X, its CDF, F_X , satisfies the following properties:

- $0 \leq F_X(x) \leq 1$, for any $x \in \mathbb{R}$.
- F_X is an increasing function on $\mathbb{R}$.
- $F_X(-\infty) = 0$ and $F_X(+\infty) = 1$.
- F_X is a right-continuous function, i.e. $F_X(x_0+) = F_X(x_0)$ for any $x_0 \in \mathbb{R}$.

Random Variables

Definition

Two random variables are said to be **identically distributed** *if their distribution functions are equal.*

Random Variables

Definition

Two random variables are said to be **identically distributed** *if their distribution functions are equal.*

Definition

We say that X *is a* **continuous** *RV if there exists a nonnegative function* f*, defined for all real* $x \in \mathbb{R}$*, having the property that*

$$F_X(x) = \mathbb{P}(X \leq x) = \int_{-\infty}^{x} f(t)dt.$$

Definition

Two random variables are said to be **identically distributed** *if their distribution functions are equal.*

Definition

We say that X is a **continuous** *RV if there exists a nonnegative function f, defined for all real $x \in \mathbb{R}$, having the property that*

$$F_X(x) = \mathbb{P}(X \leq x) = \int_{-\infty}^{x} f(t)dt.$$

The function f is called the **probability density function (PDF)** *of the random variable X.*

Example (The RV is X - the number of tosses until the first tails comes up)

The PMF:

Example (The RV is X - the number of tosses until the first tails comes up)

The PMF:

X	1	2	...	n	$\dots$
$\mathbb{P}(X = x)$	$1/2$	$1/4$	...	$1/2^n$	$\dots$

Example (The RV is X - the number of tosses until the first tails comes up)

The PMF:

X	1	2	...	n	$\dots$
$\mathbb{P}(X = x)$	$1/2$	$1/4$	...	$1/2^n$	$\dots$

The CDF formula of X:

Example (The RV is X - the number of tosses until the first tails comes up)

The PMF:

X	1	2	...	n	$\dots$
$\mathbb{P}(X = x)$	$1/2$	$1/4$	...	$1/2^n$	$\dots$

The CDF formula of X:

$$F_X(x) = \begin{cases} 0, & \text{if } x < 1, \\ 1/2, & \text{if } 1 \leq x < 2, \\ 3/4, & \text{if } 2 \leq x < 3, \\ \dots, & \dots \\ (2^k - 1)/2^k, & \text{if } k \leq x < k+1, \\ \dots, & \dots \end{cases}$$