Expected Value, Variance, Distributions

Hayk Aprikyan, Hayk Tarkhanyan

Expected Value

Suppose you are playing roulette, with numbers 1 to 38 on it. You bet a number and spin it.

Expected Value

Suppose you are playing roulette, with numbers 1 to 38 on it. You bet a number and spin it. Say you have picked the number 8 and bet $1.

Expected Value

Suppose you are playing roulette, with numbers 1 to 38 on it. You bet a number and spin it. Say you have picked the number 8 and bet \$1.

- If it falls on 8, you win \$36,

Expected Value

Suppose you are playing roulette, with numbers 1 to 38 on it. You bet a number and spin it. Say you have picked the number 8 and bet \$1.

- If it falls on 8, you win \$36,
- Otherwise, you lose \$1.

Expected Value

Suppose you are playing roulette, with numbers 1 to 38 on it. You bet a number and spin it. Say you have picked the number 8 and bet \$1.

- If it falls on 8, you win \$36,
- Otherwise, you lose \$1.

Expected Value

Suppose you are playing roulette, with numbers 1 to 38 on it. You bet a number and spin it. Say you have picked the number 8 and bet \$1.

- If it falls on 8, you win \$36,
- Otherwise, you lose \$1.

Would you play this game?

Expected Value

Suppose you are playing roulette, with numbers 1 to 38 on it. You bet a number and spin it. Say you have picked the number 8 and bet $1.

- If it falls on 8, you win $36,
- Otherwise, you lose $1.

Would you play this game? What if instead of $36, you won $150 if it fell on 8?

Expected Value

Suppose you are playing roulette, with numbers 1 to 38 on it. You bet a number and spin it. Say you have picked the number 8 and bet \$1.

- If it falls on 8, you win \$36,
- Otherwise, you lose \$1.

Would you play this game? What if instead of \$36, you won \$150 if it fell on 8? How about \$37.01?

Expected Value

Since the chance of winning is only $\frac{1}{38}$, if you play it a couple of thousands times (say 38000), then you can expect to win about ~ 1000 times and lose ~ 37000 times. Your net revenue would then be:

$$1000 \cdot (+37.01) + 37000 \cdot (-1) > 0 \qquad \Rightarrow \qquad \$\$$$

Expected Value

Since the chance of winning is only $\frac{1}{38}$, if you play it a couple of thousands times (say 38000), then you can expect to win about ~ 1000 times and lose ~ 37000 times. Your net revenue would then be:

$$1000 \cdot (+37.01) + 37000 \cdot (-1) > 0 \qquad \Rightarrow \qquad \$\$\$$$

Expected Value

Since the chance of winning is only $\frac{1}{38}$, if you play it a couple of thousands times (say 38000), then you can expect to win about ~ 1000 times and lose ~ 37000 times. Your net revenue would then be:

$$1000 \cdot (+37.01) + 37000 \cdot (-1) > 0 \qquad \Rightarrow \qquad \$\$\$$$

Similarly, if you were again waiting for a bus and someone asked you "When will it come?", what would you say would be the average number of minutes the bus is expected to arrive.

Expected Value

Since the chance of winning is only $\frac{1}{38}$, if you play it a couple of thousands times (say 38000), then you can expect to win about ~ 1000 times and lose ~ 37000 times. Your net revenue would then be:

$$1000 \cdot (+37.01) + 37000 \cdot (-1) > 0 \qquad \Rightarrow \qquad \$\$\$$$

Similarly, if you were again waiting for a bus and someone asked you "When will it come?", what would you say would be the average number of minutes the bus is expected to arrive.

These examples motivate the notion of the **mean** or **expected value** of a random variable.

Expected Value

Definition

If X is a discrete random variable, its **expected value** is defined by

$$\mathbb{E}(X) = \sum_{x_i} x_i \cdot \mathbb{P}(X = x_i)$$

Expected Value

Definition

If X is a discrete random variable, its **expected value** is defined by

$$\mathbb{E}(X) = \sum_{x_i} x_i \cdot \mathbb{P}(X = x_i)$$

Definition

If X is a continuous random variable, its **expected value** is defined by

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} x \cdot f_X(x)\, dx$$

Expected Value

Definition

If X is a discrete random variable, its **expected value** is defined by

$$\mathbb{E}(X) = \sum_{x_i} x_i \cdot \mathbb{P}(X = x_i)$$

Definition

If X is a continuous random variable, its **expected value** is defined by

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} x \cdot f_X(x)\, dx$$

In words, the expected value is the weighted average of all its possible values, each of the values being weighted by its probability.

Expected Value

We have seen that if X is a random variable, so is X^2, X^3 or $\sin X$.

Expected Value

We have seen that if X is a random variable, so is X^2, X^3 or $\sin X$. In general,

Theorem

If X is a random variable and $g(x)$ is a continuous function, then $g(X)$ is a random variable too.

Expected Value

We have seen that if X is a random variable, so is X^2, X^3 or $\sin X$. In general,

Theorem

If X is a random variable and $g(x)$ is a continuous function, then $g(X)$ is a random variable too.

Theorem

If X is a continuous random variable, then for any continuous function g,

$$\mathbb{E}(g(X)) = \int_{-\infty}^{\infty} g(x) \cdot f_X(x)\, dx$$

If X is a discrete random variable, then for any continuous function g,

$$\mathbb{E}(g(X)) = \sum_{x_i} g(x_i) \cdot \mathbb{P}(X = x_i)$$

Expected Value

Properties

1. If a is a constant,
$$\mathbb{E}(aX) = a \cdot \mathbb{E}(X)$$

Expected Value

Properties

1. If a is a constant,
$$\mathbb{E}(aX) = a \cdot \mathbb{E}(X)$$
2. If b is a constant,
$$\mathbb{E}(X + b) = \mathbb{E}(X) + b$$

Expected Value

Properties

1. If a is a constant,
$$\mathbb{E}(aX) = a \cdot \mathbb{E}(X)$$
2. If b is a constant,
$$\mathbb{E}(X + b) = \mathbb{E}(X) + b$$
3. $\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y)$

Expected Value

Properties

1. If a is a constant,
$$\mathbb{E}(aX) = a \cdot \mathbb{E}(X)$$
2. If b is a constant,
$$\mathbb{E}(X + b) = \mathbb{E}(X) + b$$
3. $\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y)$
4. $\mathbb{E}(X_1 + X_2 + \cdots + X_n) = \mathbb{E}(X_1) + \mathbb{E}(X_2) + \cdots + \mathbb{E}(X_n)$

Expected Value

Properties

1. If a is a constant,
$$\mathbb{E}(aX) = a \cdot \mathbb{E}(X)$$
2. If b is a constant,
$$\mathbb{E}(X + b) = \mathbb{E}(X) + b$$
3. $\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y)$
4. $\mathbb{E}(X_1 + X_2 + \cdots + X_n) = \mathbb{E}(X_1) + \mathbb{E}(X_2) + \cdots + \mathbb{E}(X_n)$
5. If $X \geq Y$, then $\mathbb{E}(X) \geq \mathbb{E}(Y)$

Expected Value

Properties

1. If a is a constant,
$$\mathbb{E}(aX) = a \cdot \mathbb{E}(X)$$
2. If b is a constant,
$$\mathbb{E}(X + b) = \mathbb{E}(X) + b$$
3. $\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y)$
4. $\mathbb{E}(X_1 + X_2 + \cdots + X_n) = \mathbb{E}(X_1) + \mathbb{E}(X_2) + \cdots + \mathbb{E}(X_n)$
5. If $X \geq Y$, then $\mathbb{E}(X) \geq \mathbb{E}(Y)$

Expected Value

Properties

1. If a is a constant,
$$\mathbb{E}(aX) = a \cdot \mathbb{E}(X)$$
2. If b is a constant,
$$\mathbb{E}(X + b) = \mathbb{E}(X) + b$$
3. $\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y)$
4. $\mathbb{E}(X_1 + X_2 + \cdots + X_n) = \mathbb{E}(X_1) + \mathbb{E}(X_2) + \cdots + \mathbb{E}(X_n)$
5. If $X \geq Y$, then $\mathbb{E}(X) \geq \mathbb{E}(Y)$

Theorem

If X and Y are independent random variables, then

$$\mathbb{E}(XY) = \mathbb{E}(X) \cdot \mathbb{E}(Y)$$

The converse is not necessarily true.

Variance

Now assume you are offered to play one of these two games:

- You toss a coin and win \$1 if it is Heads, otherwise you lose \$1,
- You toss a coin and win \$10000 if it is Heads, otherwise you lose \$10000.

Variance

Now assume you are offered to play one of these two games:

- You toss a coin and win \$1 if it is Heads, otherwise you lose \$1,
- You toss a coin and win \$10000 if it is Heads, otherwise you lose \$10000.

In both games, your overall winnings is on average \$0. However in the first game the amounts differ from \$0 by a small amount, while in the second game, they differ by such a large amount you might not take a risk of playing at all.

Variance

Now assume you are offered to play one of these two games:

- You toss a coin and win \$1 if it is Heads, otherwise you lose \$1,
- You toss a coin and win \$10000 if it is Heads, otherwise you lose \$10000.

In both games, your overall winnings is on average \$0. However in the first game the amounts differ from \$0 by a small amount, while in the second game, they differ by such a large amount you might not take a risk of playing at all.

In this case, we say that the winnings of the second game have a **higher variance** than those of the first one.

Variance

Definition

The **variance** of a random variable X, denoted by $\mathrm{Var}(X)$, is defined by:

$$\mathrm{Var}(X) = \mathbb{E}\left((X - \mathbb{E}(X))^2\right)$$

Variance

Definition

The **variance** of a random variable X, denoted by $\mathrm{Var}(X)$, is defined by:

$$\mathrm{Var}(X) = \mathbb{E}\big((X - \mathbb{E}(X))^2\big)$$

Opening the brackets and simplifying the expression, we get the following formula:

$$\mathrm{Var}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2$$

Variance

Definition

The **variance** of a random variable X, denoted by $\mathrm{Var}(X)$, is defined by:

$$\mathrm{Var}(X) = \mathbb{E}\big((X - \mathbb{E}(X))^2\big)$$

Opening the brackets and simplifying the expression, we get the following formula:

$$\mathrm{Var}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2$$

Definition

$\sqrt{\mathrm{Var}(X)}$ is called the **standard deviation** of X.

Variance

Definition

The **variance** of a random variable X, denoted by $\mathrm{Var}(X)$, is defined by:

$$\mathrm{Var}(X) = \mathbb{E}\left((X - \mathbb{E}(X))^2\right)$$

Opening the brackets and simplifying the expression, we get the following formula:

$$\mathrm{Var}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2$$

Definition

$\sqrt{\mathrm{Var}(X)}$ is called the **standard deviation** of X.

The standard deviation shows how much, in average, do the values of the random variable deviate from their average ($\mathbb{E}(X)$).

Variance

Example

Let X denote the number on a fair die. Then,

$$\mathbb{E}(X) =$$

Example

Let X denote the number on a fair die. Then,

$$\mathbb{E}(X) = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} = \frac{7}{2} = 3.5$$

Example

Let X denote the number on a fair die. Then,

$$\mathbb{E}(X) = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} = \frac{7}{2} = 3.5$$

To calculate $\text{Var}(X)$, we need $\mathbb{E}(X^2)$:

$$\mathbb{E}(X^2) = 1 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 9 \cdot \frac{1}{6} + 16 \cdot \frac{1}{6} + 25 \cdot \frac{1}{6} + 36 \cdot \frac{1}{6} = \frac{91}{6}$$

Variance

Example

Let X denote the number on a fair die. Then,

$$\mathbb{E}(X) = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} = \frac{7}{2} = 3.5$$

To calculate $\mathrm{Var}(X)$, we need $\mathbb{E}(X^2)$:

$$\mathbb{E}(X^2) = 1 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 9 \cdot \frac{1}{6} + 16 \cdot \frac{1}{6} + 25 \cdot \frac{1}{6} + 36 \cdot \frac{1}{6} = \frac{91}{6}$$

$$\mathrm{Var}(X) = \mathbb{E}(X^2) - \big(\mathbb{E}(X)\big)^2 = \frac{91}{6} - \left(\frac{7}{2}\right)^2 = \frac{91}{6} - \frac{49}{4} \approx 2.92$$

Variance

Properties

1. $\mathrm{Var}(X) \geq 0$,

Variance

Properties

1. $\mathrm{Var}(X) \geq 0$,
2. If X is constant, $\mathrm{Var}(X) = 0$,

Variance

Properties

1. $\mathrm{Var}(X) \geq 0$,
2. If X is constant, $\mathrm{Var}(X) = 0$,
3. $\mathrm{Var}(aX) = a^2 \cdot \mathrm{Var}(X)$ for any $a \in \mathbb{R}$,

Variance

Properties

1. $\mathrm{Var}(X) \geq 0$,
2. If X is constant, $\mathrm{Var}(X) = 0$,
3. $\mathrm{Var}(aX) = a^2 \cdot \mathrm{Var}(X)$ for any $a \in \mathbb{R}$,
4. $\mathrm{Var}(X + Y) \neq \mathrm{Var}(X) + \mathrm{Var}(Y)$, instead:

$$\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2(\mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y))$$

Variance

Properties

1. $\mathrm{Var}(X) \geq 0$,
2. If X is constant, $\mathrm{Var}(X) = 0$,
3. $\mathrm{Var}(aX) = a^2 \cdot \mathrm{Var}(X)$ for any $a \in \mathbb{R}$,
4. $\mathrm{Var}(X + Y) \neq \mathrm{Var}(X) + \mathrm{Var}(Y)$, instead:

$$\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2(\mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y))$$

5. If X and Y are independent,

$$\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$$

Variance

Properties

1. $\mathrm{Var}(X) \geq 0$,
2. If X is constant, $\mathrm{Var}(X) = 0$,
3. $\mathrm{Var}(aX) = a^2 \cdot \mathrm{Var}(X)$ for any $a \in \mathbb{R}$,
4. $\mathrm{Var}(X + Y) \neq \mathrm{Var}(X) + \mathrm{Var}(Y)$, instead:

$$\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2(\mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y))$$

5. If X and Y are independent,

$$\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$$

Variance

Properties

1. $\text{Var}(X) \geq 0$,
2. If X is constant, $\text{Var}(X) = 0$,
3. $\text{Var}(aX) = a^2 \cdot \text{Var}(X)$ for any $a \in \mathbb{R}$,
4. $\text{Var}(X + Y) \neq \text{Var}(X) + \text{Var}(Y)$, instead:

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2(\mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y))$$

5. If X and Y are independent,

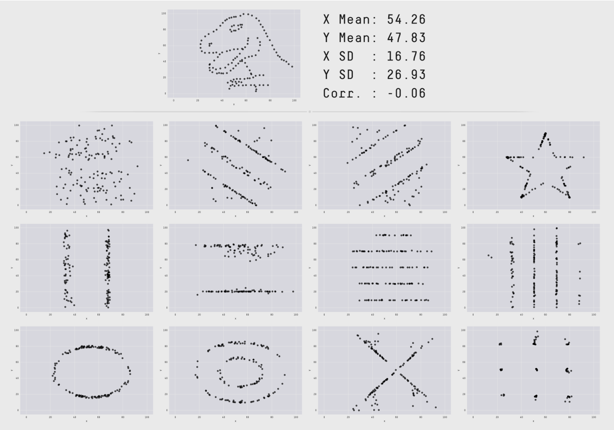
$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$$

Why do you think the 4th point makes sense?

Variance

Warning

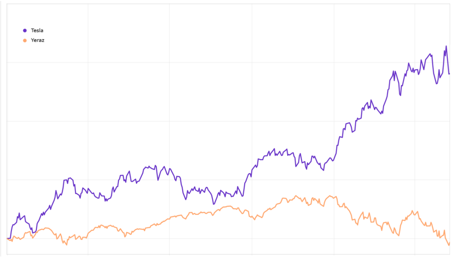
Expected value and variance are very useful to describe random variables, **but they are not everything!** They do not replace CDF/PDF/PMF!



[source]

Covariance

Suppose X is the stock price of Tesla, Y is the stock price of Yeraz, and you have some Yeraz stocks.



Covariance

Suppose X is the stock price of Tesla, Y is the stock price of Yeraz, and you have some Yeraz stocks.

Due to some reasons, the stock of Tesla starts to decrease. Naturally, you are interested in how could that affect your Yeraz stocks, i.e.

Question

How would a change of X affect Y?

Covariance

Suppose X is the stock price of Tesla, Y is the stock price of Yeraz, and you have some Yeraz stocks.

Due to some reasons, the stock of Tesla starts to decrease. Naturally, you are interested in how could that affect your Yeraz stocks, i.e.

Question

How would a change of X affect Y?

More specifically, if X goes up by 1 unit, how much would Y change?

Covariance

Definition

The **covariance** between random variables X and Y is defined by

$$\mathrm{Cov}(X, Y) = \mathbb{E}\big((X - \mathbb{E}(X))(Y - \mathbb{E}(Y))\big)$$

Covariance

Definition

The **covariance** between random variables X and Y is defined by

$$\text{Cov}(X, Y) = \mathbb{E}\big((X - \mathbb{E}(X))(Y - \mathbb{E}(Y))\big)$$

Note that the formula for covariance can be simplified to

$$\text{Cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$$

Definition

The **covariance** between random variables X and Y is defined by

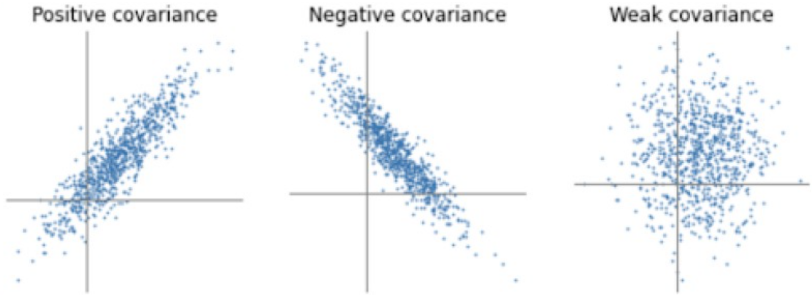
$$\mathrm{Cov}(X, Y) = \mathbb{E}\big((X - \mathbb{E}(X))(Y - \mathbb{E}(Y))\big)$$

Note that the formula for covariance can be simplified to

$$\mathrm{Cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$$

Covariance shows how much the linear growth of one RV is related to the linear growth of the other RV. It is very similar to the concept of dot product of the two vectors.

Covariance



Positive covariance
Negative covariance
Weak covariance

Covariance

Properties

1. $\text{Cov}(X, Y) = \text{Cov}(Y, X)$,

Covariance

Properties

1. $\text{Cov}(X, Y) = \text{Cov}(Y, X)$,
2. $\text{Cov}(X, X) = \text{Var}(X)$,

Covariance

Properties

1. $\mathrm{Cov}(X, Y) = \mathrm{Cov}(Y, X)$,
2. $\mathrm{Cov}(X, X) = \mathrm{Var}(X)$,
3. $\mathrm{Cov}(a \cdot X, Y) = a \cdot \mathrm{Cov}(X, Y)$,

Covariance

Properties

1. $\text{Cov}(X, Y) = \text{Cov}(Y, X)$,
2. $\text{Cov}(X, X) = \text{Var}(X)$,
3. $\text{Cov}(a \cdot X, Y) = a \cdot \text{Cov}(X, Y)$,
4. $\text{Cov}\left(X_1 + X_2, Y_1 + Y_2\right) =$
   $\text{Cov}(X_1, Y_1) + \text{Cov}(X_1, Y_2) + \text{Cov}(X_2, Y_1) + \text{Cov}(X_2, Y_2)$,

Covariance

Properties

1. $\text{Cov}(X, Y) = \text{Cov}(Y, X)$,
2. $\text{Cov}(X, X) = \text{Var}(X)$,
3. $\text{Cov}(a \cdot X, Y) = a \cdot \text{Cov}(X, Y)$,
4. $\text{Cov}\left(X_1 + X_2, Y_1 + Y_2\right) =$
   $\text{Cov}(X_1, Y_1) + \text{Cov}(X_1, Y_2) + \text{Cov}(X_2, Y_1) + \text{Cov}(X_2, Y_2)$,
5. $\text{Var}\left(X_1 + X_2\right) = \text{Var}(X_1) + \text{Var}(X_2) + 2 \cdot \text{Cov}(X_1, X_2)$,

Covariance

Properties

1. $\mathrm{Cov}(X, Y) = \mathrm{Cov}(Y, X)$,
2. $\mathrm{Cov}(X, X) = \mathrm{Var}(X)$,
3. $\mathrm{Cov}(a \cdot X, Y) = a \cdot \mathrm{Cov}(X, Y)$,
4. $\mathrm{Cov}\left(X_1 + X_2, Y_1 + Y_2\right) =$
   $\mathrm{Cov}(X_1, Y_1) + \mathrm{Cov}(X_1, Y_2) + \mathrm{Cov}(X_2, Y_1) + \mathrm{Cov}(X_2, Y_2)$,
5. $\mathrm{Var}\left(X_1 + X_2\right) = \mathrm{Var}(X_1) + \mathrm{Var}(X_2) + 2 \cdot \mathrm{Cov}(X_1, X_2)$,
6. If X and Y are independent, $\mathrm{Cov}(X, Y) = 0$ (converse isn't true).

Covariance

Properties

1. $\text{Cov}(X, Y) = \text{Cov}(Y, X)$,
2. $\text{Cov}(X, X) = \text{Var}(X)$,
3. $\text{Cov}(a \cdot X, Y) = a \cdot \text{Cov}(X, Y)$,
4. $\text{Cov}\left(X_1 + X_2, Y_1 + Y_2\right) =$
   $\text{Cov}(X_1, Y_1) + \text{Cov}(X_1, Y_2) + \text{Cov}(X_2, Y_1) + \text{Cov}(X_2, Y_2)$,
5. $\text{Var}\left(X_1 + X_2\right) = \text{Var}(X_1) + \text{Var}(X_2) + 2 \cdot \text{Cov}(X_1, X_2)$,
6. If X and Y are independent, $\text{Cov}(X, Y) = 0$ (converse isn't true).

Covariance

Properties

1. $\text{Cov}(X, Y) = \text{Cov}(Y, X)$,
2. $\text{Cov}(X, X) = \text{Var}(X)$,
3. $\text{Cov}(a \cdot X, Y) = a \cdot \text{Cov}(X, Y)$,
4. $\text{Cov}\left(X_1 + X_2, Y_1 + Y_2\right) =$
   $\text{Cov}(X_1, Y_1) + \text{Cov}(X_1, Y_2) + \text{Cov}(X_2, Y_1) + \text{Cov}(X_2, Y_2)$,
5. $\text{Var}\left(X_1 + X_2\right) = \text{Var}(X_1) + \text{Var}(X_2) + 2 \cdot \text{Cov}(X_1, X_2)$,
6. If X and Y are independent, $\text{Cov}(X, Y) = 0$ (converse isn't true).

$\text{Cov}(X, Y)$ can be *any* number (positive/negative, large/small, zero, etc). What if we want a normalized, universal method to measure the relatedness level of two random variables?

Correlation

Definition

For two non-constant $(\mathrm{Var}(X), \mathrm{Var}(Y) \neq 0)$ random variables X and Y, the **correlation** (or **Pearson correlation coefficient**) between them is defined as:

$$\rho(X, Y) = \mathrm{Corr}(X, Y) = \frac{\mathrm{Cov}(X, Y)}{\sqrt{\mathrm{Var}(X) \cdot \mathrm{Var}(Y)}}$$

Correlation

Definition

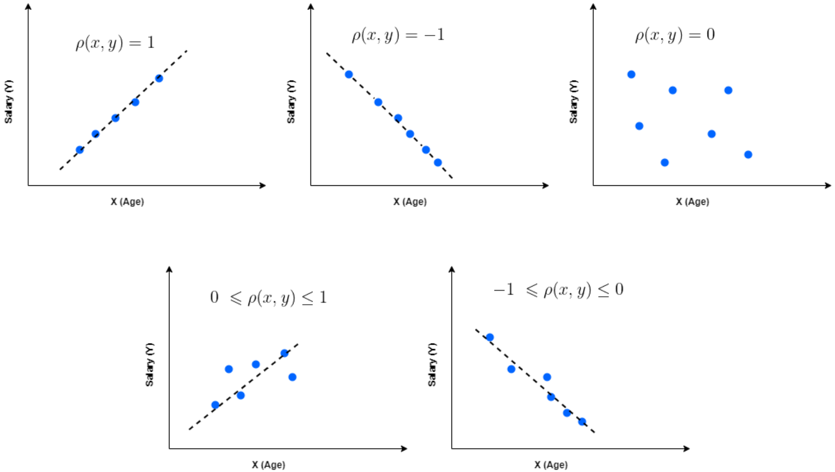
For two non-constant ($\mathrm{Var}(X), \mathrm{Var}(Y) \neq 0$) random variables X and Y, the **correlation** (or **Pearson correlation coefficient**) between them is defined as:

$$\rho(X, Y) = \mathrm{Corr}(X, Y) = \frac{\mathrm{Cov}(X, Y)}{\sqrt{\mathrm{Var}(X) \cdot \mathrm{Var}(Y)}}$$

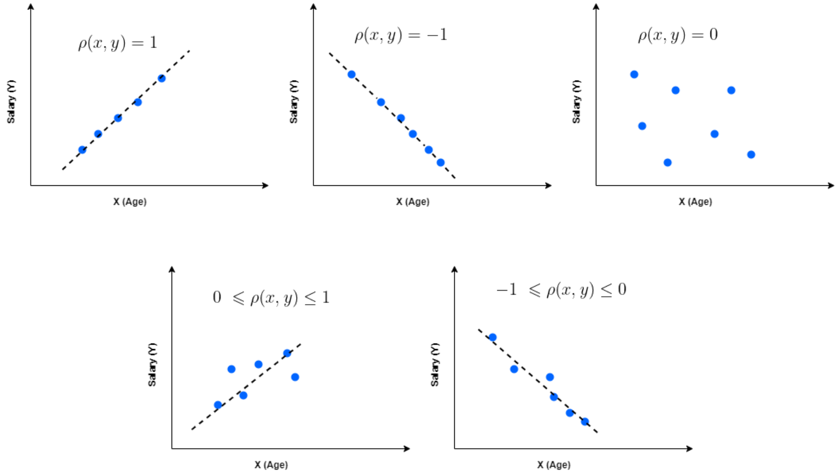
Properties

1. $\rho(X, Y) = \rho(Y, X)$,
2. $-1 \leq \rho(X, Y) \leq 1$,
3. $Y = aX + b$ for some constants a, b if and only if $\rho(X, Y) = \pm 1$.

Correlation



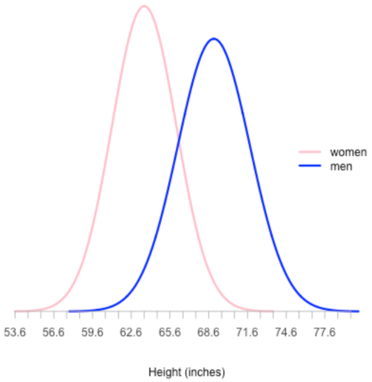
Correlation



- Play with this correlation visualization!

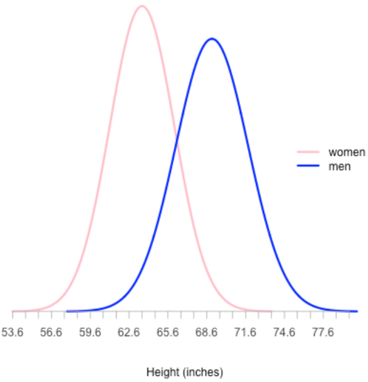
Distributions

Very often in practice, many random variables share similar properties. In particular, the probabilities of their values seem to follow a common pattern, i.e. their CDFs (or PMFs/PDFs) are similar to each other:



Distributions

Very often in practice, many random variables share similar properties. In particular, the probabilities of their values seem to follow a common pattern, i.e. their CDFs (or PMFs/PDFs) are similar to each other:



The way the values of an RV are distributed is called a *distribution*.

Bernoulli Distribution

Consider these situations:

Bernoulli Distribution

Consider these situations:

- You take a pass-fail exam. You either pass or fail.

Bernoulli Distribution

Consider these situations:

- You take a pass-fail exam. You either pass or fail.
- You toss a coin. The outcome is either heads or tails.

Bernoulli Distribution

Consider these situations:

- You take a pass-fail exam. You either pass or fail.
- You toss a coin. The outcome is either heads or tails.
- A child is born. The gender at birth is either male or female.

Bernoulli Distribution

Consider these situations:

- You take a pass-fail exam. You either pass or fail.
- You toss a coin. The outcome is either heads or tails.
- A child is born. The gender at birth is either male or female.

Bernoulli Distribution

Consider these situations:

- You take a pass-fail exam. You either pass or fail.
- You toss a coin. The outcome is either heads or tails.
- A child is born. The gender at birth is either male or female.

Definition

A random variable X is said to be a **Bernoulli** random variable with parameter p, denoted by $X \sim \mathit{Bernoulli}(p)$, if it only takes two values and its PMF is given by:

$$\mathbb{P}(X = x) = \begin{cases} p & \text{for } x = 1 \\ 1 - p & \text{for } x = 0 \\ 0 & \text{otherwise} \end{cases}$$

where $0 < p < 1$.

Bernoulli Distribution

Consider these situations:

- You take a pass-fail exam. You either pass or fail.
- You toss a coin. The outcome is either heads or tails.
- A child is born. The gender at birth is either male or female.

Definition

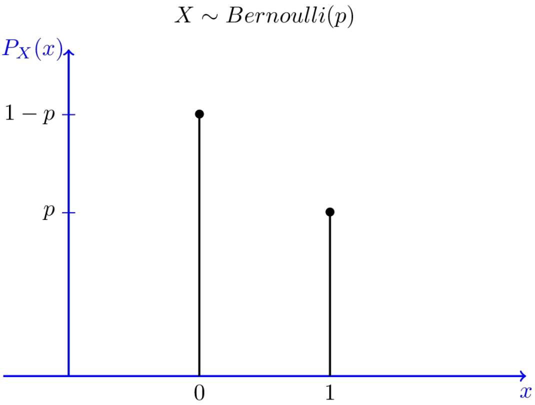
A random variable X is said to be a **Bernoulli** random variable with parameter p, denoted by $X \sim Bernoulli(p)$, if it only takes two values and its PMF is given by:

$$\mathbb{P}(X = x) = \begin{cases} p & \text{for } x = 1 \\ 1 - p & \text{for } x = 0 \\ 0 & \text{otherwise} \end{cases}$$

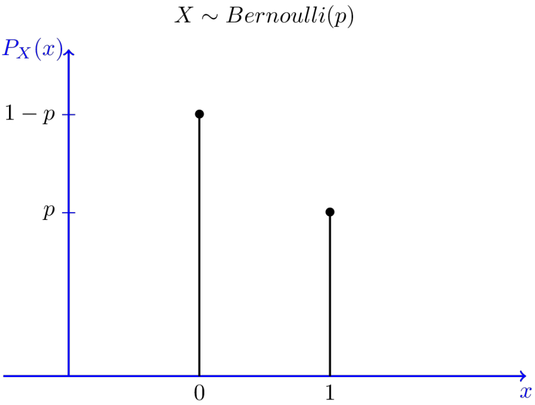
where $0 < p < 1$.

A series of n independent experiments all following Bernoulli distribution $Bernoulli(p)$, is called **Bernoulli trials**.

Bernoulli Distribution



Bernoulli Distribution



If $X \sim Bernoulli(p)$,

$$\mathbb{E}(X) = p, \qquad \mathrm{Var}(X) = p(1-p)$$

Geometric Distribution

Suppose we have a coin with $\mathbb{P}(\text{Heads}) = p$ and we toss it until we observe the first Heads, after which we stop. We define X as the total number of coin tosses in this experiment. Then X is said to have *geometric distribution* with parameter p.

Geometric Distribution

Suppose we have a coin with $\mathbb{P}(\text{Heads}) = p$ and we toss it until we observe the first Heads, after which we stop. We define X as the total number of coin tosses in this experiment. Then X is said to have *geometric distribution* with parameter p.
For any natural number $k \geq 1$:

$$\mathbb{P}(X = k) = (1 - p)^{k-1} p$$

Geometric Distribution

Suppose we have a coin with $\mathbb{P}(\text{Heads}) = p$ and we toss it until we observe the first Heads, after which we stop. We define X as the total number of coin tosses in this experiment. Then X is said to have *geometric distribution* with parameter p.
For any natural number $k \geq 1$:

$$\mathbb{P}(X = k) = (1 - p)^{k-1} p$$

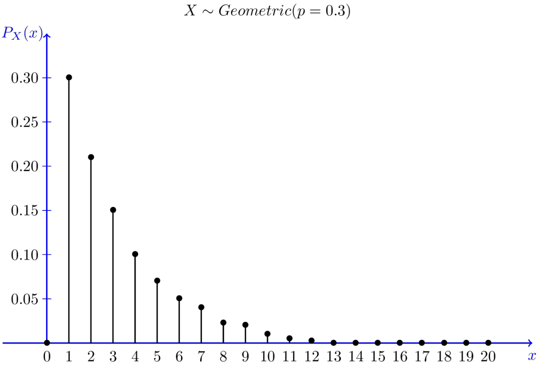
Definition

A random variable X is said to be a **geometric** random variable with parameter p, denoted by $X \sim Geo(p)$, if its PMF is given by:

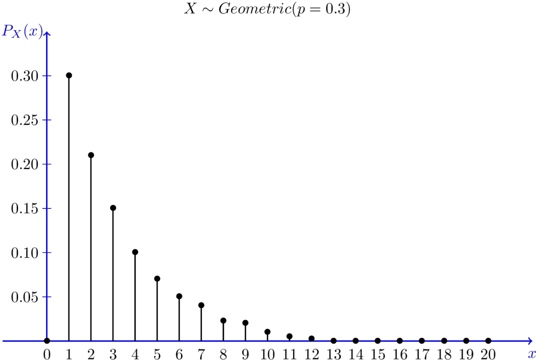
$$\mathbb{P}(X = k) = \begin{cases} (1-p)^{k-1} p & \text{for } k = 1, 2, 3, \ldots \\ 0 & \text{otherwise} \end{cases}$$

where $0 < p < 1$.

Geometric Distribution



Geometric Distribution



If $X \sim Geo(p)$,

$$\mathbb{E}(X) = \frac{1}{p}, \qquad \mathrm{Var}(X) = \frac{1-p}{p^2}$$

Binomial Distribution

Suppose we have a coin with $\mathbb{P}(\text{Heads}) = p$ and we toss it n times. We define X to be the total number of Heads observed. Then X is said to have *binomial distribution* with parameter n and p.

Binomial Distribution

Suppose we have a coin with $\mathbb{P}(\text{Heads}) = p$ and we toss it n times. We define X to be the total number of Heads observed. Then X is said to have *binomial distribution* with parameter n and p.
For any whole number $k \geq 0$:

$$\mathbb{P}(X = k) = C_n^k p^k (1 - p)^{n-k}$$

Binomial Distribution

Suppose we have a coin with $\mathbb{P}(\text{Heads}) = p$ and we toss it n times. We define X to be the total number of Heads observed. Then X is said to have *binomial distribution* with parameter n and p.
For any whole number $k \geq 0$:

$$\mathbb{P}(X = k) = C_n^k p^k (1-p)^{n-k}$$

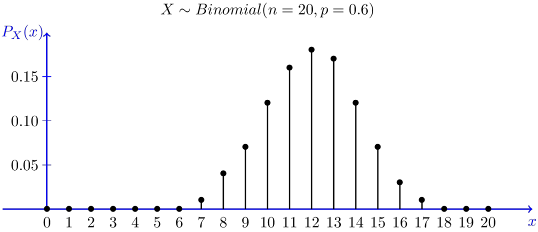
Definition

A random variable X is said to be a **binomial** random variable with parameters n and p, denoted by $X \sim B(n, p)$, if its PMF is given by:

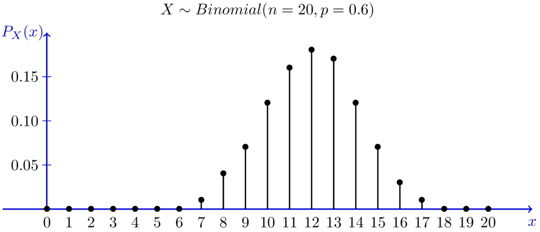
$$\mathbb{P}(X = k) = \begin{cases} C_n^k p^k (1-p)^{n-k} & \text{for } k = 0, 1, 2, \ldots \\ 0 & \text{otherwise} \end{cases}$$

where $0 < p < 1$.

Binomial Distribution



Binomial Distribution



If $X \sim B(n, p)$,

$$\mathbb{E}(X) = np, \qquad \mathrm{Var}(X) = np(1-p)$$

Poisson Distribution

The Poisson distribution is one of the most widely used probability distributions. It is usually used in scenarios where we are counting the occurrences of certain events in an interval of time or space. In practice, it is often an approximation of a real-life random variable.

Poisson Distribution

The Poisson distribution is one of the most widely used probability distributions. It is usually used in scenarios where we are counting the occurrences of certain events in an interval of time or space. In practice, it is often an approximation of a real-life random variable.
Suppose we are counting the number of customers who visit a certain store from 13:00 to 14:00. Based on data from previous days, we know that on average, $\lambda = 15$ customers visit the store.

Poisson Distribution

The Poisson distribution is one of the most widely used probability distributions. It is usually used in scenarios where we are counting the occurrences of certain events in an interval of time or space. In practice, it is often an approximation of a real-life random variable.
Suppose we are counting the number of customers who visit a certain store from 13:00 to 14:00. Based on data from previous days, we know that on average, $\lambda = 15$ customers visit the store. In practice, the number of customers tends to follow the following pattern:

Poisson Distribution

The Poisson distribution is one of the most widely used probability distributions. It is usually used in scenarios where we are counting the occurrences of certain events in an interval of time or space. In practice, it is often an approximation of a real-life random variable.

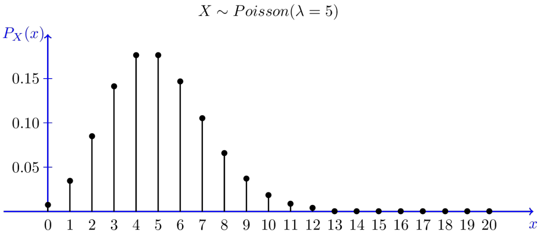
Suppose we are counting the number of customers who visit a certain store from 13:00 to 14:00. Based on data from previous days, we know that on average, $\lambda = 15$ customers visit the store. In practice, the number of customers tends to follow the following pattern:

Definition

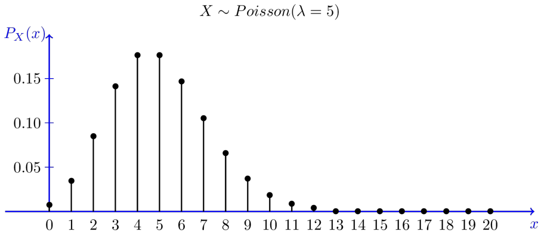
A random variable X is said to be a **Poisson** random variable with parameter λ, denoted by $X \sim Poisson(\lambda)$, if its PMF is given by:

$$\mathbb{P}(X = k) = \begin{cases} \dfrac{\lambda^k}{k!} e^{-\lambda} & \text{for } k = 0, 1, 2, \dots \\ 0 & \text{otherwise} \end{cases}$$

Poisson Distribution



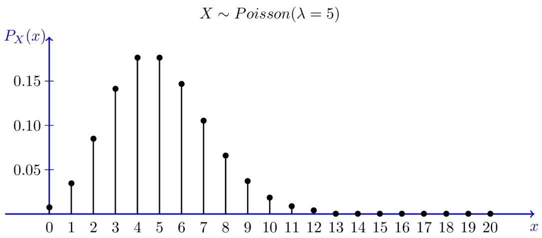
Poisson Distribution



If $X \sim Poisson(\lambda)$,

$$\mathbb{E}(X) = \lambda, \qquad \mathrm{Var}(X) = \lambda$$

Poisson Distribution



If $X \sim Poisson(\lambda)$,

$$\mathbb{E}(X) = \lambda, \qquad \mathrm{Var}(X) = \lambda$$

Remark

If $X \sim B(n, p)$, then

$$\mathbb{P}(X = k) \approx \frac{\lambda^k}{k!} e^{-\lambda}, \qquad \text{where } \lambda = np$$

Uniform Distribution

So far, we have considered only discrete RVs. Let's observe some common distributions for continuous RVs.

Uniform Distribution

So far, we have considered only discrete RVs. Let's observe some common distributions for continuous RVs.

If we pick a random number X from a given interval, without any number being "more probable" than another, then X is said to be *uniformly* distributed.

Uniform Distribution

So far, we have considered only discrete RVs. Let's observe some common distributions for continuous RVs.

If we pick a random number X from a given interval, without any number being "more probable" than another, then X is said to be *uniformly* distributed.

Definition

A random variable X is said to be a **uniform** random variable over the interval $[a, b]$, denoted by $X \sim U(a, b)$, if its PDF is given by:

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{if } x \in (a, b) \\ 0 & \text{if } x \notin (a, b) \end{cases}$$

Uniform Distribution

So far, we have considered only discrete RVs. Let's observe some common distributions for continuous RVs.

If we pick a random number X from a given interval, without any number being "more probable" than another, then X is said to be *uniformly* distributed.

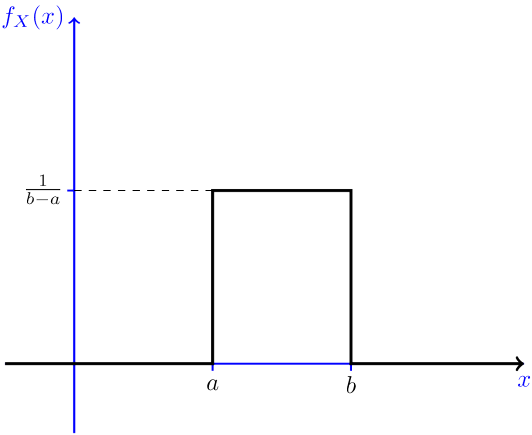
Definition

A random variable X is said to be a **uniform** random variable over the interval $[a, b]$, denoted by $X \sim U(a, b)$, if its PDF is given by:

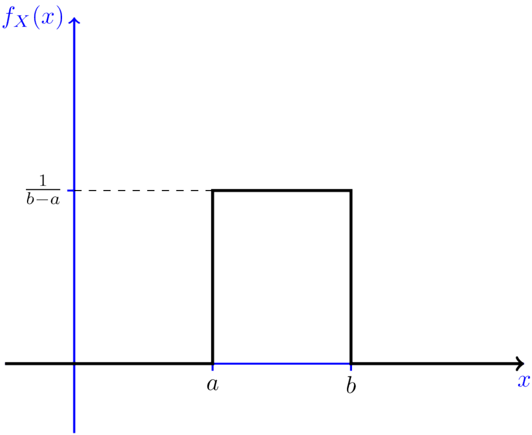
$$f(x) = \begin{cases} \frac{1}{b-a} & \text{if } x \in (a, b) \\ 0 & \text{if } x \notin (a, b) \end{cases}$$

$$F(x) = \begin{cases} 0 & \text{if } x \leq a \\ \frac{x-a}{b-a} & \text{if } a < x < b \\ 1 & \text{if } x \geq b \end{cases}$$

Uniform Distribution



Uniform Distribution



If $X \sim U(a, b)$,

$$\mathbb{E}(X) = \frac{a+b}{2}, \qquad \mathrm{Var}(X) = \frac{(b-a)^2}{12}$$

Exponential Distribution

The *exponential* distribution is the continuous analog of the geometric distribution. It is one of the widely used continuous distributions, and is often used to model the time elapsed between events.

Exponential Distribution

The *exponential* distribution is the continuous analog of the geometric distribution. It is one of the widely used continuous distributions, and is often used to model the time elapsed between events.

Definition

A random variable X is said to be a **exponential** random variable with parameter $\lambda > 0$, denoted by $X \sim Exp(\lambda)$, if its PDF is given by:

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & \text{if } x > 0 \\ 0 & \text{otherwise} \end{cases}$$

Exponential Distribution

The *exponential* distribution is the continuous analog of the geometric distribution. It is one of the widely used continuous distributions, and is often used to model the time elapsed between events.

Definition

A random variable X is said to be a **exponential** random variable with parameter $\lambda > 0$, denoted by $X \sim Exp(\lambda)$, if its PDF is given by:

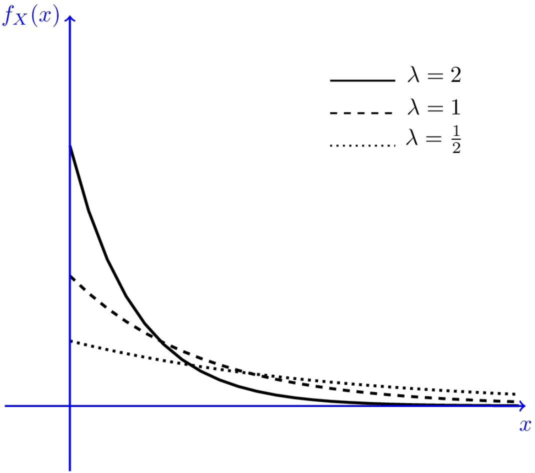
$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & \text{if } x > 0 \\ 0 & \text{otherwise} \end{cases}$$

Remark

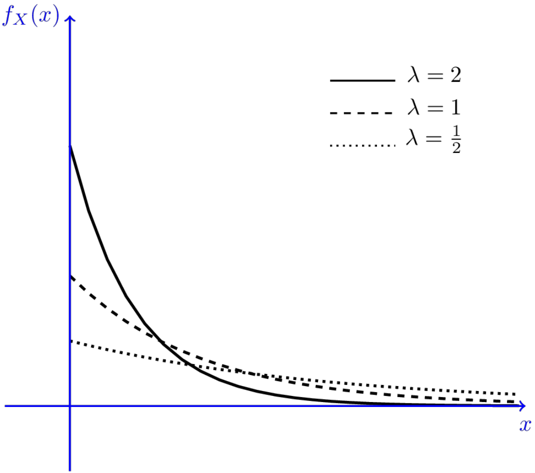
If $X \sim Exp(\lambda)$, then X is a **memoryless** random variable, that is

$$\mathbb{P}(X > x + a \mid X > a) = \mathbb{P}(X > x), \quad \text{for } a, x \geq 0.$$

Exponential Distribution



Exponential Distribution



If $X \sim Exp(\lambda)$,

$$\mathbb{E}(X) = \frac{1}{\lambda}, \qquad \mathrm{Var}(X) = \frac{1}{\lambda^2}$$

Normal Distribution

The *normal* distribution is by far the most important probability distribution.

Normal Distribution

The *normal* distribution is by far the most important probability distribution.
One of the main reasons for that is that if you add a large number of random variables, the distribution of the sum will be approximately normal under certain conditions. In real life, many RVs can be expressed as the sum of a large number of random variables, and their distribution is normal.

Normal Distribution

The *normal* distribution is by far the most important probability distribution.
One of the main reasons for that is that if you add a large number of random variables, the distribution of the sum will be approximately normal under certain conditions. In real life, many RVs can be expressed as the sum of a large number of random variables, and their distribution is normal.

Definition

A random variable X is said to be a **normal** random variable with mean μ and variance σ^2, denoted by $X \sim N(\mu, \sigma^2)$, if its PDF is given by:

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \qquad x \in \mathbb{R}$$

Normal Distribution