# Probability

Adam Klepáč

March 8, 2024

# Probability Distribution

# What Is Probability Distribution?

**Probability distribution** is a function that describes the probability of different *possible values* of a *random variable*.

# What Is Probability Distribution?

**Probability distribution** is a function that describes the probability of different *possible values* of a *random variable*.
*graphs*.

# What Is Probability Distribution?

**Probability distribution** is a function that describes the probability of different *possible values* of a *random variable*.
*graphs*.
For example, imagine the random variable $X$ that describes the outcome of a coin toss, that is $X \in \{\text{heads}, \text{tails}\}$.

# What Is Probability Distribution?

**Probability distribution** is a function that describes the probability of different *possible values* of a *random variable*.
*graphs*.
For example, imagine the random variable $X$ that describes the outcome of a coin toss, that is $X \in \{\text{heads}, \text{tails}\}$.
If the coin is fair then $P(X = \text{heads}) = P(X = \text{tails}) = 1/2$.

# What Is Probability Distribution?

**Probability distribution** is a function that describes the probability of different *possible values* of a *random variable*.
*graphs*.
For example, imagine the random variable $X$ that describes the outcome of a coin toss, that is $X \in \{\text{heads}, \text{tails}\}$.
If the coin is fair then $P(X = \text{heads}) = P(X = \text{tails}) = 1/2$.
The **probability distribution** of this random variable is a function $f : \{\text{heads}, \text{tails}\} \to [0, 1]$ which assigns to the element 'heads' the probability $P(X = \text{heads})$ and to 'tails' the probability $P(X = \text{tails})$.

# What Is Probability Distribution?

**Probability distribution** is a function that describes the probability of different *possible values* of a *random variable*.
*graphs*.
For example, imagine the random variable $X$ that describes the outcome of a coin toss, that is $X \in \{\text{heads}, \text{tails}\}$.
If the coin is fair then $P(X = \text{heads}) = P(X = \text{tails}) = 1/2$.
The **probability distribution** of this random variable is a function $f : \{\text{heads}, \text{tails}\} \to [0, 1]$ which assigns to the element 'heads' the probability $P(X = \text{heads})$ and to 'tails' the probability $P(X = \text{tails})$.
In other words, $f(\text{heads}) = f(\text{tails}) = 1/2$.

# Probability Distribution – Examples

- The **probability distribution** of a random variable representing the value of a dice roll is a function

$$f : \{1, 2, 3, 4, 5, 6\} \to [0, 1]$$

such that $f(k) = 1/6$ for all numbers $k \in \{1, 2, 3, 4, 5, 6\}$.

# Probability Distribution – Examples

- The **probability distribution** of a random variable representing the value of a dice roll is a function

$$f : \{1, 2, 3, 4, 5, 6\} \to [0, 1]$$

  such that $f(k) = 1/6$ for all numbers $k \in \{1, 2, 3, 4, 5, 6\}$.
- The **probability distribution** of a random variable representing the rank of a randomly chosen playing card is a function

# Probability Distribution – Examples

- The **probability distribution** of a random variable representing the value of a dice roll is a function
$$f : \{1, 2, 3, 4, 5, 6\} \to [0, 1]$$
such that $f(k) = 1/6$ for all numbers $k \in \{1, 2, 3, 4, 5, 6\}$.
- The **probability distribution** of a random variable representing the rank of a randomly chosen playing card is a function
$$f : \{2, 3, 4, 5, 6, 7, 8, 9, 10, J, Q, K, A\} \to [0, 1]$$
such that $f(r) = 4/52$ where $r$ is a rank of a playing card.

# Visualizing Probability Distributions – Tables

*Discrete* **probability distributions** (meaning distributions of a *discrete* random variable) can be easily represented using **tables**.

# Visualizing Probability Distributions – Tables

*Discrete* **probability distributions** (meaning distributions of a *discrete* random variable) can be easily represented using **tables**.

For example, the probability distribution of a dice roll is given simply by

| **Roll** | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| **Probability** | $\frac{1}{6}$ | $\frac{1}{6}$ | $\frac{1}{6}$ | $\frac{1}{6}$ | $\frac{1}{6}$ | $\frac{1}{6}$. |

# Visualizing Probability Distributions – Tables

For a more abstract example, if $X$ can attain any of the four values $a, b, c, d$ with probabilities $P(X = a) = 3/10, P(X = b) = 5/10, P(X = c) = 1/10, P(X = d) = 1/10$, then its probability distribution is

| **Value** | a | b | c | d |
|---|---|---|---|---|
| **Probability** | $\frac{3}{10}$ | $\frac{5}{10}$ | $\frac{1}{10}$ | $\frac{1}{10}$. |

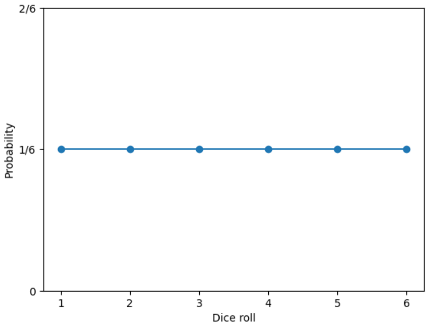
# Visualizing Probability Distributions – Graphs

**Probability distributions** (both *discrete* and *continuous*) can be represented as graphs. These are your typical function graphs which draw inputs on the $x$-axis and outputs on the $y$-axis.

# Visualizing Probability Distributions – Graphs

**Probability distributions** (both *discrete* and *continuous*) can be represented as graphs. These are your typical function graphs which draw inputs on the $x$-axis and outputs on the $y$-axis.

The probability distribution of a dice roll looks like this

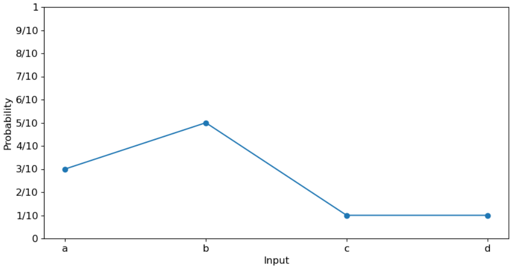


# Visualizing Probability Distributions – Graphs

The probability distribution from this table

| **Value** | a | b | c | d |
|---|---|---|---|---|
| **Probability** | $\frac{3}{10}$ | $\frac{5}{10}$ | $\frac{1}{10}$ | $\frac{1}{10}$. |

looks like this



1

# Discrete Probability Distributions

# Discrete Probability Distribution

Let $X$ be a random variable taking values from a set $A$.

# Discrete Probability Distribution

Let $X$ be a random variable taking values from a set $A$.

- The **probability distribution** (or also **probability mass function**) of $X$ is the function $f : A \to [0, 1]$ defined as $f(a) = P(X = a)$ for $a \in A$.

# Discrete Probability Distribution

Let $X$ be a random variable taking values from a set $A$.

- The **probability distribution** (or also **probability mass function**) of $X$ is the function $f : A \to [0, 1]$ defined as $f(a) = P(X = a)$ for $a \in A$.
- The **cumulative distribution function** of $X$ gives the probability that a random variable is **less than a certain value**. It is defined as $F(a) = P(X \leq a)$ for $a \in A$.

# Discrete Probability Distribution

Let $X$ be a random variable taking values from a set $A$.

- The **probability distribution** (or also **probability mass function**) of $X$ is the function $f : A \to [0, 1]$ defined as $f(a) = P(X = a)$ for $a \in A$.
- The **cumulative distribution function** of $X$ gives the probability that a random variable is **less than a certain value**. It is defined as $F(a) = P(X \leq a)$ for $a \in A$.
- The **mean** of $X$ is defined as $E(X) = \sum_{a \in A} a \cdot P(X = a)$. It represents the 'expected' value of $X$.

# Discrete Probability Distribution

Let $X$ be a random variable taking values from a set $A$.

- The **probability distribution** (or also **probability mass function**) of $X$ is the function $f : A \to [0,1]$ defined as $f(a) = P(X = a)$ for $a \in A$.
- The **cumulative distribution function** of $X$ gives the probability that a random variable is **less than a certain value**. It is defined as $F(a) = P(X \leq a)$ for $a \in A$.
- The **mean** of $X$ is defined as $E(X) = \sum_{a \in A} a \cdot P(X = a)$. It represents the 'expected' value of $X$.
- The **variance** (describing the *dispersion* of the distribution around the mean) of $X$ is defined as

$$\mathrm{Var}(X) = \sum_{a \in A} (a - E(X))^2 \cdot P(X = a).$$

# Discrete Probability Distribution – Example

Let's see what these concepts mean in a simple statistical experiment.

# Discrete Probability Distribution – Example

Let's see what these concepts mean in a simple statistical experiment.

Suppose we measure the height of a randomly picked 20-year-old males. We might get something akin to the following table

| **Height** | 175 | 176 | 177 | 178 | 179 | 180 | 181 | 182 | 183 |
|---|---|---|---|---|---|---|---|---|---|
| **Count** | 13 | 20 | 11 | 17 | 11 | 8 | 10 | 7 | 3 |

# Discrete Probability Distribution – Example

Let's see what these concepts mean in a simple statistical experiment.

Suppose we measure the height of a randomly picked 20-year-old males. We might get something akin to the following table

| **Height** | 175 | 176 | 177 | 178 | 179 | 180 | 181 | 182 | 183 |
|---|---|---|---|---|---|---|---|---|---|
| **Count** | 13 | 20 | 11 | 17 | 11 | 8 | 10 | 7 | 3 |

We can easily calculate the mean and standard deviation of this data.

# Discrete Probability Distribution – Example

| **Height** | 175 | 176 | 177 | 178 | 179 | 180 | 181 | 182 | 183 |
|---|---|---|---|---|---|---|---|---|---|
| **Count** | 13 | 20 | 11 | 17 | 11 | 8 | 10 | 7 | 3 |

Using the formula for the arithmetic mean, we get

$$\bar{x} = \frac{175 \cdot 13 + 176 \cdot 20 + \ldots + 183 \cdot 3}{13 + 20 + \ldots + 3} = 178.1$$

# Discrete Probability Distribution – Example

| Height | 175 | 176 | 177 | 178 | 179 | 180 | 181 | 182 | 183 |
|---|---|---|---|---|---|---|---|---|---|
| Count | 13 | 20 | 11 | 17 | 11 | 8 | 10 | 7 | 3 |

The standard deviation is then

$$\sigma = \sqrt{\frac{13 \cdot (175 - 178.1)^2 + 20 \cdot (176 - 178.1)^2 + \ldots + 3 \cdot (183 - 178.1)^2}{13 + 20 + \ldots + 3}} = 8.203.$$

# Discrete Probability Distribution – Example

| **Height** | 175 | 176 | 177 | 178 | 179 | 180 | 181 | 182 | 183 |
|---|---|---|---|---|---|---|---|---|---|
| **Count** | 13 | 20 | 11 | 17 | 11 | 8 | 10 | 7 | 3 |

Let's now define a random variable $X$ which can be any of those heights in the table above.

# Discrete Probability Distribution – Example

| Height | 175 | 176 | 177 | 178 | 179 | 180 | 181 | 182 | 183 |
|---|---|---|---|---|---|---|---|---|---|
| Count | 13 | 20 | 11 | 17 | 11 | 8 | 10 | 7 | 3 |

Let's now define a random variable $X$ which can be any of those heights in the table above. We define the probabilities that $X$ is a particular height based on the counts above. That gives the following table

| Height | 175 | 176 | 177 | 178 | 179 | 180 | 181 | 182 | 183 |
|---|---|---|---|---|---|---|---|---|---|
| Probability | $\frac{13}{100}$ | $\frac{20}{100}$ | $\frac{11}{100}$ | $\frac{17}{100}$ | $\frac{11}{100}$ | $\frac{8}{100}$ | $\frac{10}{100}$ | $\frac{7}{100}$ | $\frac{3}{100}$ |

# Discrete Probability Distribution – Example

| Height | 175 | 176 | 177 | 178 | 179 | 180 | 181 | 182 | 183 |
|---|---|---|---|---|---|---|---|---|---|
| Count | 13 | 20 | 11 | 17 | 11 | 8 | 10 | 7 | 3 |

Let's now define a random variable $X$ which can be any of those heights in the table above. We define the probabilities that $X$ is a particular height based on the counts above. That gives the following table

| Height | 175 | 176 | 177 | 178 | 179 | 180 | 181 | 182 | 183 |
|---|---|---|---|---|---|---|---|---|---|
| Probability | $\frac{13}{100}$ | $\frac{20}{100}$ | $\frac{11}{100}$ | $\frac{17}{100}$ | $\frac{11}{100}$ | $\frac{8}{100}$ | $\frac{10}{100}$ | $\frac{7}{100}$ | $\frac{3}{100}$ |

In other words, this gives a **distribution function** $f$ of $X$ where the set $A = \{175, 176, 177, 178, 179, 180, 181, 182, 183\}$ and the outputs of $f$ on each of these numbers are given by the table above.

# Discrete Probability Distribution – Example

| **Height** | 175 | 176 | 177 | 178 | 179 | 180 | 181 | 182 | 183 |
|---|---|---|---|---|---|---|---|---|---|
| **Probability** | $\frac{13}{100}$ | $\frac{20}{100}$ | $\frac{11}{100}$ | $\frac{17}{100}$ | $\frac{11}{100}$ | $\frac{8}{100}$ | $\frac{10}{100}$ | $\frac{7}{100}$ | $\frac{3}{100}$ |

- The **cumulative distribution function** $F$ describes the probability that a randomly chosen person from the group has height *less than* a particular number.

# Discrete Probability Distribution – Example

| Height | 175 | 176 | 177 | 178 | 179 | 180 | 181 | 182 | 183 |
|---|---|---|---|---|---|---|---|---|---|
| **Probability** | $\frac{13}{100}$ | $\frac{20}{100}$ | $\frac{11}{100}$ | $\frac{17}{100}$ | $\frac{11}{100}$ | $\frac{8}{100}$ | $\frac{10}{100}$ | $\frac{7}{100}$ | $\frac{3}{100}$ |

- The **cumulative distribution function** $F$ describes the probability that a randomly chosen person from the group has height *less than* a particular number. For example,

$$F(178) = P(X \leq 178) = P(X = 175) + P(X = 176) + P(X = 177) + P(X = 178)$$
$$= \frac{13}{100} + \frac{20}{100} + \frac{11}{100} + \frac{17}{100} = \frac{61}{100}.$$

# Discrete Probability Distribution – Example

| **Height** | 175 | 176 | 177 | 178 | 179 | 180 | 181 | 182 | 183 |
|---|---|---|---|---|---|---|---|---|---|
| **Probability** | $\frac{13}{100}$ | $\frac{20}{100}$ | $\frac{11}{100}$ | $\frac{17}{100}$ | $\frac{11}{100}$ | $\frac{8}{100}$ | $\frac{10}{100}$ | $\frac{7}{100}$ | $\frac{3}{100}$ |

- The **mean** of $X$ is the same as the arithmetic mean of the data. Indeed,

$$\begin{aligned} E(X) &= \sum_{a \in A} a \cdot P(X = a) \\ &= 175 \cdot P(X = 175) + 176 \cdot P(X = 176) + \ldots + 183 \cdot P(X = 183) \\ &= 175 \cdot \frac{13}{100} + 176 \cdot \frac{20}{100} + \ldots + 183 \cdot \frac{3}{100} = 178.1. \end{aligned}$$

# Discrete Probability Distribution – Example

| **Height** | 175 | 176 | 177 | 178 | 179 | 180 | 181 | 182 | 183 |
|---|---|---|---|---|---|---|---|---|---|
| **Probability** | $\frac{13}{100}$ | $\frac{20}{100}$ | $\frac{11}{100}$ | $\frac{17}{100}$ | $\frac{11}{100}$ | $\frac{8}{100}$ | $\frac{10}{100}$ | $\frac{7}{100}$ | $\frac{3}{100}$ |

- The **variance** of $X$ is the same as the standard deviation *squared* (that is, $\mathrm{Var}(X) = \sigma^2$). Indeed,

$$
\begin{aligned}
\mathrm{Var}(X) &= \sum_{a \in A} (a - E(X))^2 \cdot P(X = a) \\
&= (175 - 178.1)^2 \cdot P(X = 175) + \ldots + (183 - 178.1)^2 \cdot P(X = 183) \\
&= 67.29 = 8.203^2.
\end{aligned}
$$

# Some Important Discrete Distributions

# The Bernoulli Distribution

The **Bernoulli** distribution is a discrete distribution of a random variable which can only attain **two distinct values**.

# The Bernoulli Distribution

The **Bernoulli** distribution is a discrete distribution of a random variable which can only attain **two distinct values**.

If we denote these values as 0 and 1, then the Bernoulli distribution is the function

$$f(x) = \begin{cases} p, & \text{if } x = 1, \\ 1 - p, & \text{if } x = 0, \end{cases}$$

where $p \in [0, 1]$ is a **fixed** probability.

# The Bernoulli Distribution – Example

A **coin toss** is a perfect example of a Bernoulli distribution with $p = 1/2$.

# The Bernoulli Distribution – Example

A **coin toss** is a perfect example of a Bernoulli distribution with $p = 1/2$. Indeed, if $f$ is the probability distribution of the result of a coin toss, then

$$f(x) = \begin{cases} \frac{1}{2}, & \text{if } x = \text{heads}, \\ \frac{1}{2}, & \text{if } x = \text{tails}. \end{cases}$$

# The Bernoulli Distribution – Properties

We compute the distribution, cumulative distribution, mean and variance of the Bernoulli distribution. We assume that $X \in \{0, 1\}$ and $p \in [0, 1]$.

# The Bernoulli Distribution – Properties

We compute the distribution, cumulative distribution, mean and variance of the Bernoulli distribution. We assume that $X \in \{0, 1\}$ and $p \in [0, 1]$.

- By definition, $f(1) = p$ and $f(0) = 1 - p$.

# The Bernoulli Distribution – Properties

We compute the distribution, cumulative distribution, mean and variance of the Bernoulli distribution. We assume that $X \in \{0, 1\}$ and $p \in [0, 1]$.

- By definition, $f(1) = p$ and $f(0) = 1 - p$.
- Since we have only two values, $F(0) = P(X \leq 0) = P(X = 0) = f(0) = 1 - p$ and $F(1) = P(X \leq 1) = f(0) + f(1) = 1$.

# The Bernoulli Distribution – Properties

We compute the distribution, cumulative distribution, mean and variance of the Bernoulli distribution. We assume that $X \in \{0, 1\}$ and $p \in [0, 1]$.

- By definition, $f(1) = p$ and $f(0) = 1 - p$.
- Since we have only two values, $F(0) = P(X \leq 0) = P(X = 0) = f(0) = 1 - p$ and $F(1) = P(X \leq 1) = f(0) + f(1) = 1$.
- We calculate,

$$E(X) = \sum_{a \in \{0,1\}} a \cdot f(a) = 0 \cdot f(0) + 1 \cdot f(1) = 0 \cdot (1 - p) + 1 \cdot p = p.$$

# The Bernoulli Distribution – Properties

We compute the distribution, cumulative distribution, mean and variance of the Bernoulli distribution. We assume that $X \in \{0, 1\}$ and $p \in [0, 1]$.

- By definition, $f(1) = p$ and $f(0) = 1 - p$.
- Since we have only two values, $F(0) = P(X \leq 0) = P(X = 0) = f(0) = 1 - p$ and $F(1) = P(X \leq 1) = f(0) + f(1) = 1$.
- We calculate,

$$E(X) = \sum_{a \in \{0,1\}} a \cdot f(a) = 0 \cdot f(0) + 1 \cdot f(1) = 0 \cdot (1 - p) + 1 \cdot p = p.$$

- And also

$$\mathrm{Var}(X) = \sum_{a \in \{0,1\}} (a - E(X))^2 \cdot f(a) = (0 - p)^2 \cdot (1 - p) + (1 - p)^2 \cdot p = p(1 - p).$$

1

## Digression

↓

## Variations & Combinations

# How To Count The Number Of Repetitions?

Imagine 11 people standing in supermarket queue. How many different ways can they order themselves in that queue?

# How To Count The Number Of Repetitions?

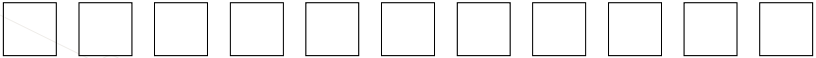
Imagine 11 people standing in supermarket queue. How many different ways can they order themselves in that queue?
The answer is relatively simple.

# How To Count The Number Of Repetitions?

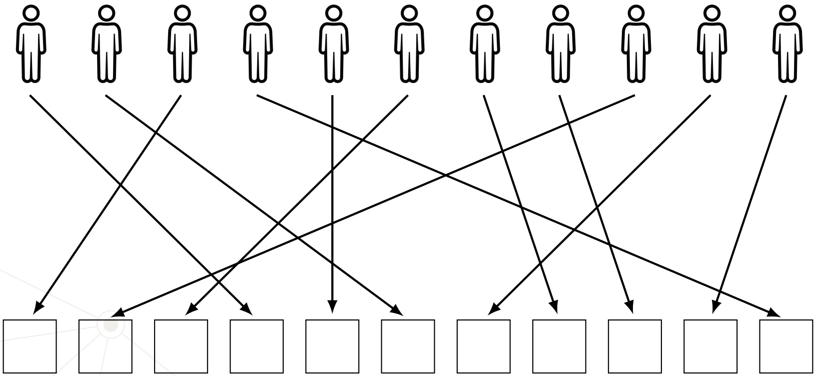
Imagine 11 people standing in supermarket queue. How many different ways can they order themselves in that queue?

The answer is relatively simple.

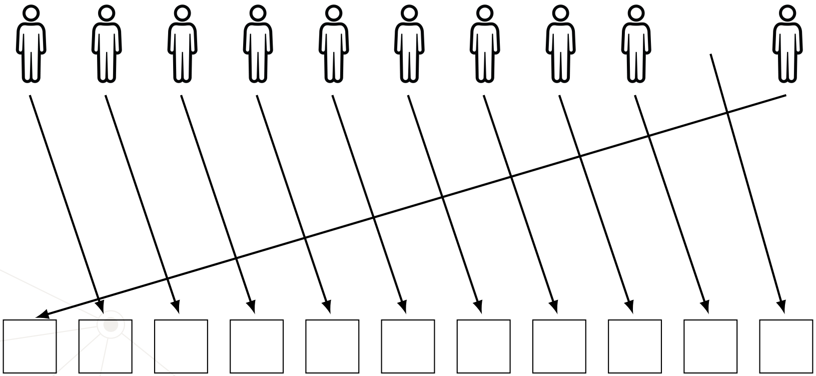
Imagine 11 empty boxes and count how many ways can you distribute 11 people into them.



# How To Count The Number Of Repetitions?



# How To Count The Number Of Repetitions?

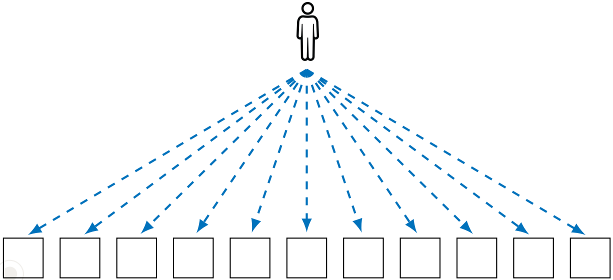


# How To Count The Number Of Repetitions?

What if there were only 1 human?

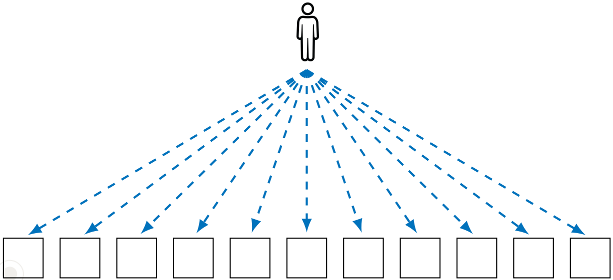
# How To Count The Number Of Repetitions?

What if there were only 1 human?



# How To Count The Number Of Repetitions?

What if there were only 1 human?



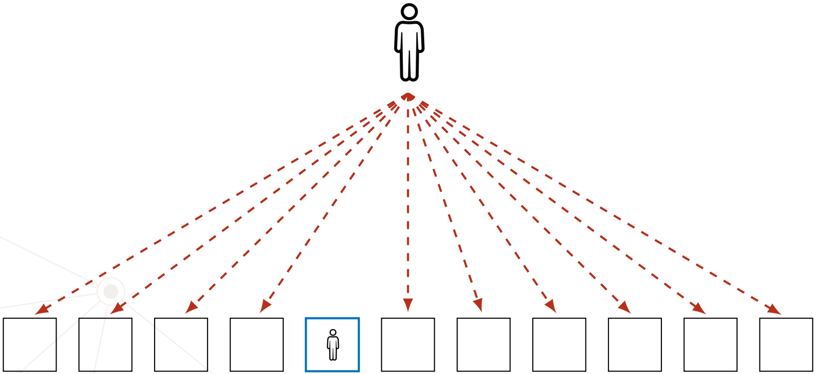
I'd have exactly 11 options where to put him.

# How To Count The Number Of Repetitions?

So, I put him in a random box and in comes the second human.

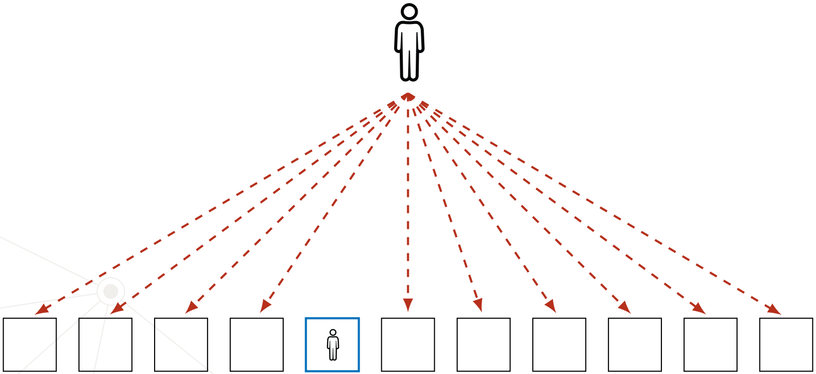
# How To Count The Number Of Repetitions?

So, I put him in a random box and in comes the second human.



# How To Count The Number Of Repetitions?

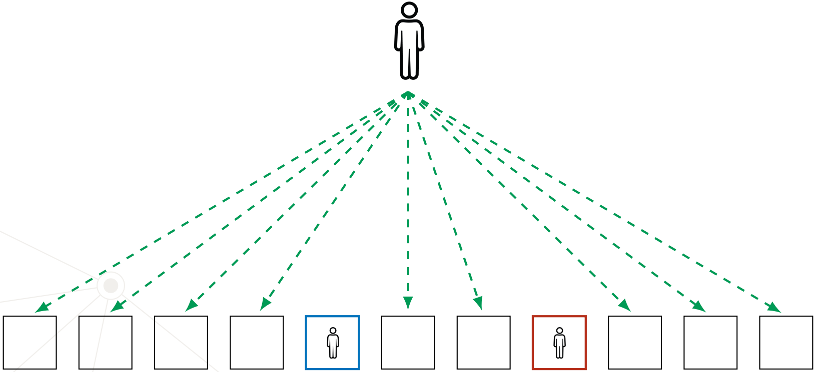
So, I put him in a random box and in comes the second human.



I'd have only 10 boxes left to place him into.

# How To Count The Number Of Repetitions?

For the third human, only 9 boxes are left, etc.



# Factorial

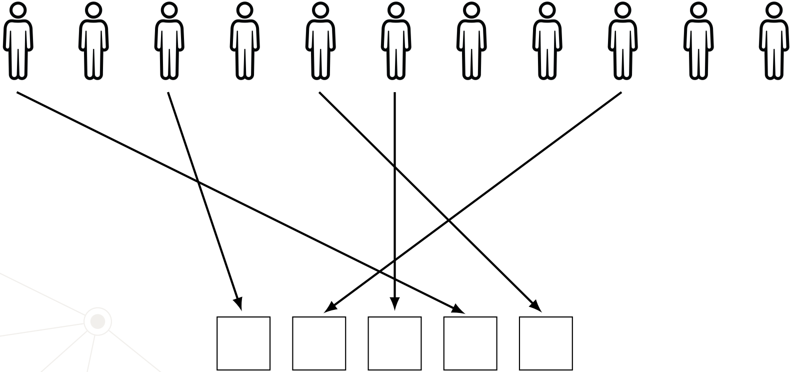
## Factorial

Overall, given $n$ objects, I have

$$n \cdot (n-1) \cdot (n-2) \cdot \ldots \cdot 1$$

ways how to order them. This number is written as $n!$ and read $n$ **factorial**.

# What If I Have Fewer Boxes?



# What If I Have Fewer Boxes?

The same argument applies.

# What If I Have Fewer Boxes?

The same argument applies.

If I have only 5 boxes and 11 humans, I have

$$11 \cdot 10 \cdot 9 \cdot 8 \cdot 7$$

ways to put 5 of those humans into all the boxes.

# What If I Have Fewer Boxes?

The same argument applies.

If I have only 5 boxes and 11 humans, I have

$$11 \cdot 10 \cdot 9 \cdot 8 \cdot 7$$

ways to put 5 of those humans into all the boxes.

## Variations

If I have $n$ elements, the number of ways how to order any $k$ of them is

$$n \cdot (n-1) \cdot (n-2) \cdot \ldots \cdot (n-k+1) = \frac{n!}{(n-k)!}.$$

This number is sometimes called the number of **variations** of $k$ elements out of $n$.

# What If I Don't Care About The Order?

Finally, suppose I'm not interested in the differences between individual humans and I just want to place some 5 of them into boxes.

# What If I Don’t Care About The Order?

Finally, suppose I’m not interested in the differences between individual humans and I just want to place some 5 of them into boxes.
In other words, I don’t care about the order I put them into those boxes. It doesn’t matter to me which human goes to which box as long as all the boxes are full.

# What If I Don't Care About The Order?

Finally, suppose I'm not interested in the differences between individual humans and I just want to place some 5 of them into boxes.

In other words, I don't care about the order I put them into those boxes. It doesn't matter to me which human goes to which box as long as all the boxes are full.

This is a similar problem. The only difference is that I disregard all the ways I can order those $k$ elements inside the boxes.

# Combinations

## Combinations

If I have $n$ elements, the number of ways I can choose any $k$ of them regardless of order, is

$$\frac{n \cdot (n-1) \cdot (n-2) \cdot \ldots \cdot (n-k+1)}{k!} = \frac{n!}{(n-k)!k!}.$$

This number is typically written as $\binom{n}{k}$ and read '$n$ **choose** $k$'.