# Ten Challenges and Explainable Analogs of growth functions and distributions for statistical literacy and fluency

## Challenge 1: Linear function

**Discover**. Consider each cube a value of 1. From left to right: start with 1 (1 cube), then 2, then 3, then 4.

**Present.** A linear function increases by a constant amount (the value of its slope) in each time interval. Real-world example problems solved by linear functions include age, speed, time and distance, pressure and force.

**Enjoy**. Congratulations! You just made a linear function that increases by 1. It can be written $f(n)=n+1$ with n being the number of cubes. You can represent each cube as a square on paper and continue adding 1 square at each iteration. This linear function $f(n)$ increases by 1 (a constant slope) every time $n$ increases by 1.

## Challenge 2: Exponential function

**Discover**. Consider each cube a value of 2. From left to right: start with 2 (1 cube), then 4 (2 cubes), then 6, then 8, then 16. You might wonder now that you are missing some more cubes. Since the value 8 requires 4 cubes, the value 16 will require 8 cubes. To represent this difference, add 4 more cubes to the 4 cubes at value 8. Then, move them to the right. The negative space demonstrates how fast this goes. If you have more cubes you may fill up the negative space or even continue further. Every next value doubles the current one. If you do not have enough cubes, you may draw on paper the space those cubes would occupy or use the pencil as a representation of 6 cubes.

**Present.** An exponential function increases by a constant percentage (or ratio) in each time interval. When you recall the beginning of the COVID-19 pandemic, governments were pushing to flatten the curve. What was meant was to contain or limit the uncontrolled exponential growth of infections to prevent hospitals from being overrun.

**Enjoy**. Well done! You just made an exponential function that increases by *2^n* with n varying from 1 to 4. This can be written *2^1, 2^2, 2^3, 2^4* which corresponds to 2, 4, 8, and 16, respectively. It can be written *f(n) = 2^n*. This exponential function *f(n)* increases by 50 percent (a constant percentage) every time n increases by 1.

## Challenge 3: Uniform distribution

**Discover**. This challenge requires all the cubes in a bag. You are also welcome to ask a friend to join you for this challenge. Draw one cube and declare it your favorite. Set it aside for a second, then put it back with the others. Place them all inside the bag, then without looking draw your favorite cube. Depending on the kind of cubes you use, it may be simply unidentifiable or impossible to tell which is which. It's completely normal, this is a difficult task. After drawing one cube, put it back in the bag. In the best case scenario, you will draw your own cube on the first try. In the worst case scenario, your favorite will never be drawn.

**Present.** A uniform distribution or a rectangular distribution is a distribution that is concerned with events that are equally likely to occur. It has constant probability. Assuming there's no preference for any particular cube, you'd imagine that the probability of each of the twelve cubes 1, 2, 3, …, 12 is the same. Since all the probabilities must add up to 1, a logical conclusion is to assign a probability of 1/12 to each of the 12 options.

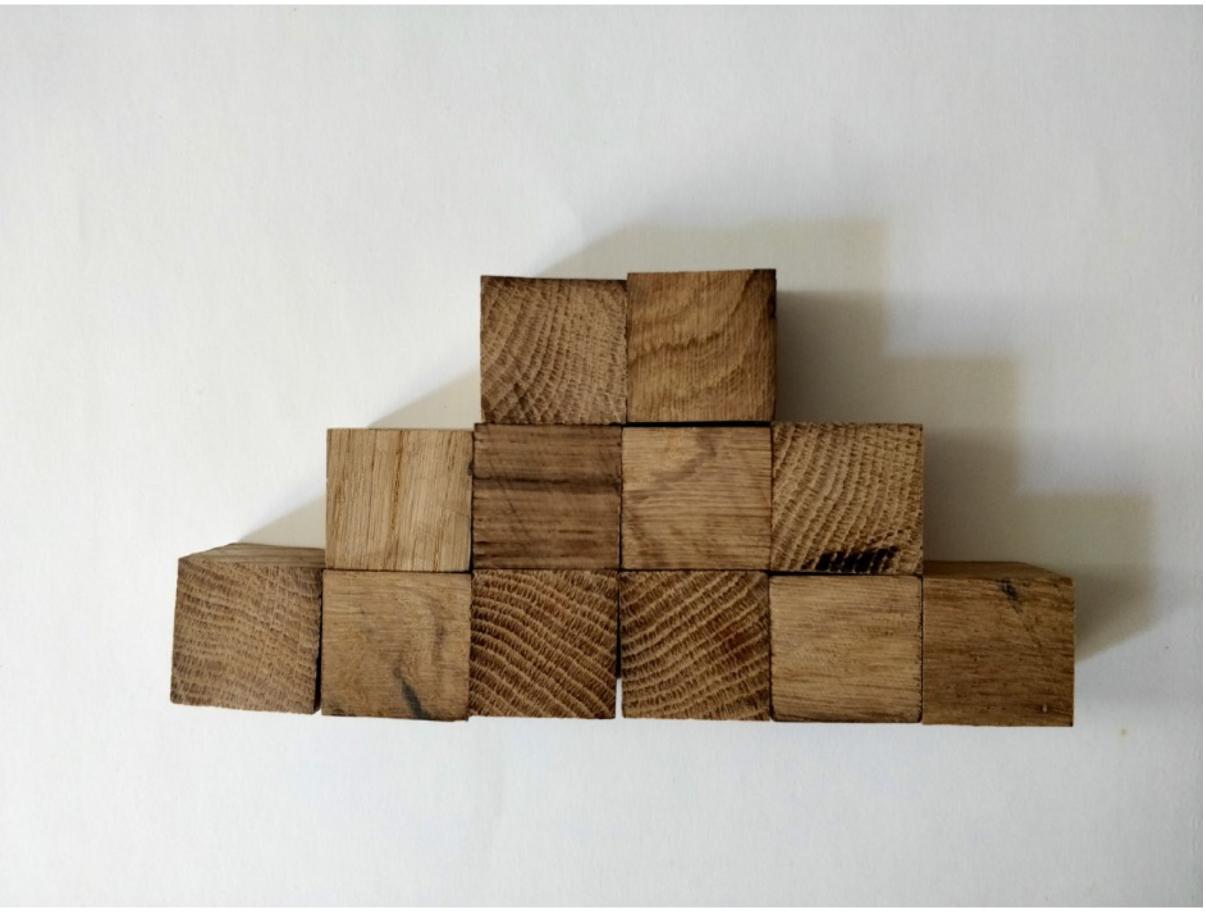
A deck of cards also has a uniform distribution. This is because a person has an equal chance of drawing a spade, heart, club, or diamond. Another example of a uniform distribution is a coin toss. The probability of getting a tail or a head is the same.

**Enjoy.** Kudos! You just experienced the uniform distribution through a randomized experiment in which a drawn item is always returned to the universe of possibilities before the next draw. The probability that you draw your favorite cube follows a uniform distribution because every cube is equally likely to be drawn next. We can write the probability *P* of drawing any cube *x* from 1 to 12 as $P(X = x) = 1/12$ *for* $x = 1, 2, 3, \ldots, 12$.

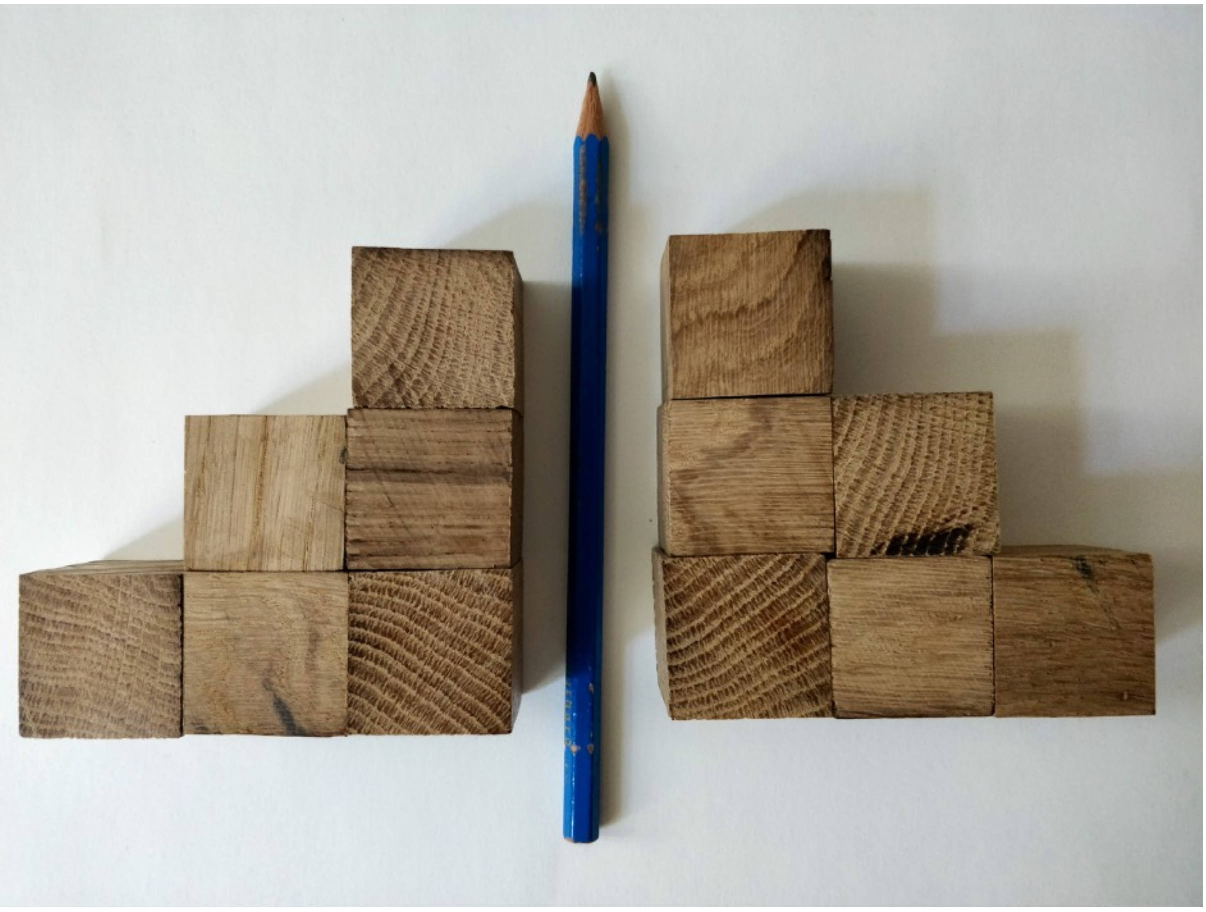
In statistical terms, we say that the probability of the drawing event remains the same throughout the experiment. To understand a uniform distribution, we have to capture the probability of *X* being close to a single number. This is achieved by relying on a probability distribution function. When the probability distribution around a point *x* is large, it means the random variable *X* is likely to be close to *x*. If the probability distribution around a point x equals zero, it means that *X* won't be in that interval. The probability of each event is always between 0 and 1.

## Challenge 4: Normal or Gaussian distribution

**Discover**. Consider each cube a value of 2. From left to right: start with 2 (1 cube), then 4, then 6. You might wonder now, this looks much like a linear function. Continue adding cubes but in reverse by restarting from the end. Add again 6, then 4, then 2. The second part looks very much like an inverted



version of the first part. It is as if a mirror is placed in the middle. You may use your pencil as if it was a mirror.



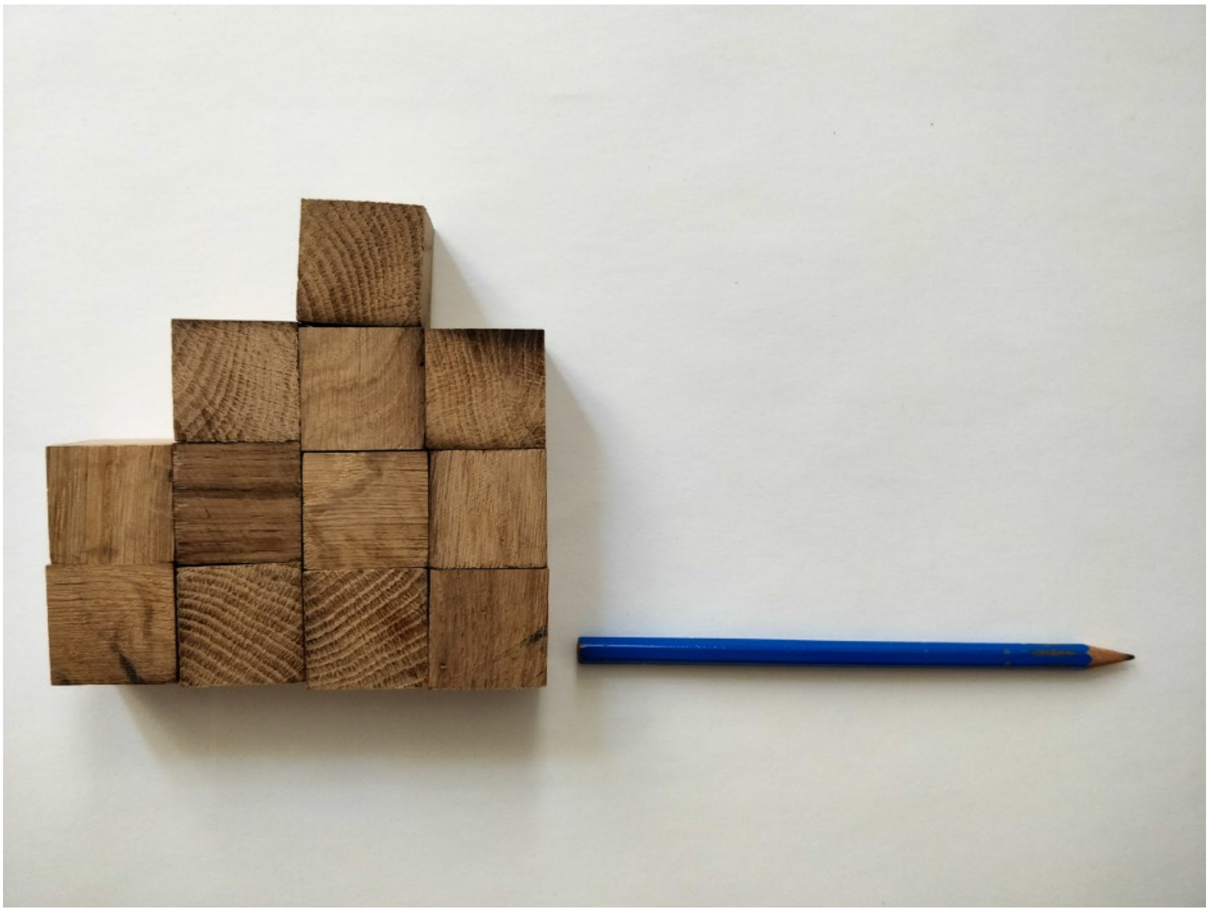
**Present**. A normal distribution is a probability distribution where the values are concentrated in the middle of the interval and the rest shrink symmetrically toward the two extremes. While statisticians and mathematicians uniformly use the term ‘normal distribution’, physicists sometimes refer to it as a Gaussian distribution and, because of its curved shape, social scientists call it the ‘bell curve’.

It represents real-valued random variables whose distributions are not known to us: the average height of NBA players, the average female shoe size, the volume of milk production from cows, the shifting distribution of summer temperatures, *etc*.

**Enjoy**. Well done! You just discovered the normal distribution. A normal distribution in a variate $X$ with mean $\mu$ and variance $\sigma^2$ is probably the most important statistical distribution to know because many phenomena correspond to this distribution (e.g., height, shoe size). The normal distribution is symmetric around the mean, showing that data close to the mean are more frequent than data far from the mean. A typical normal distribution looks much like the one you just built. It has a mean $\mu = 0$ and a standard deviation $\sigma = \sqrt{\sigma^2} = 1$.

## Challenge 5: Poisson distribution

**Discover**. Gather all 12 cubes and put them in their bag. You may ask a friend to join you for this challenge. Consider that this bag is part of a board game. As part of its rules, in a fifteen minutes or a quarter of an hour interval, a player may draw a random number of cubes with one hand, note the number then return the cubes to the bag. This means that players may draw four times in one hour. Depending on the number of cubes drawn, the game stops or continues with further actions. As an example, you may count the number of cubes drawn in a min. To represent its probability distribution, we consider 6 points and use our twelve cubes. We can consider each cube a value of 1. From left to right: start with 2, then 3, then 4, then 3, then none, then none. Since the events span only four time intervals, the last two positions are empty where we can count 0 cubes. You may place the pencil horizontally to show both positions.



**Present.** This corresponds to a count distribution or the Poisson probability distribution which gives how many times this event is likely to occur over a specified period. In other words, it gives the probability that a number of events will occur in a fixed time interval if these events occur at a known average rate and independently of the time since the last event. A Poisson process is characterized by three points. Events are independent of each other. The occurrence of one event does not affect the probability of another event occurring. The average rate (events per time period) is constant. Two events cannot occur at the same time.

For example, a book publisher may be interested in the number of misspelled words in a particular book. It may be that, on average, there are five misspelled words in 100 pages. The interval is 100 pages.

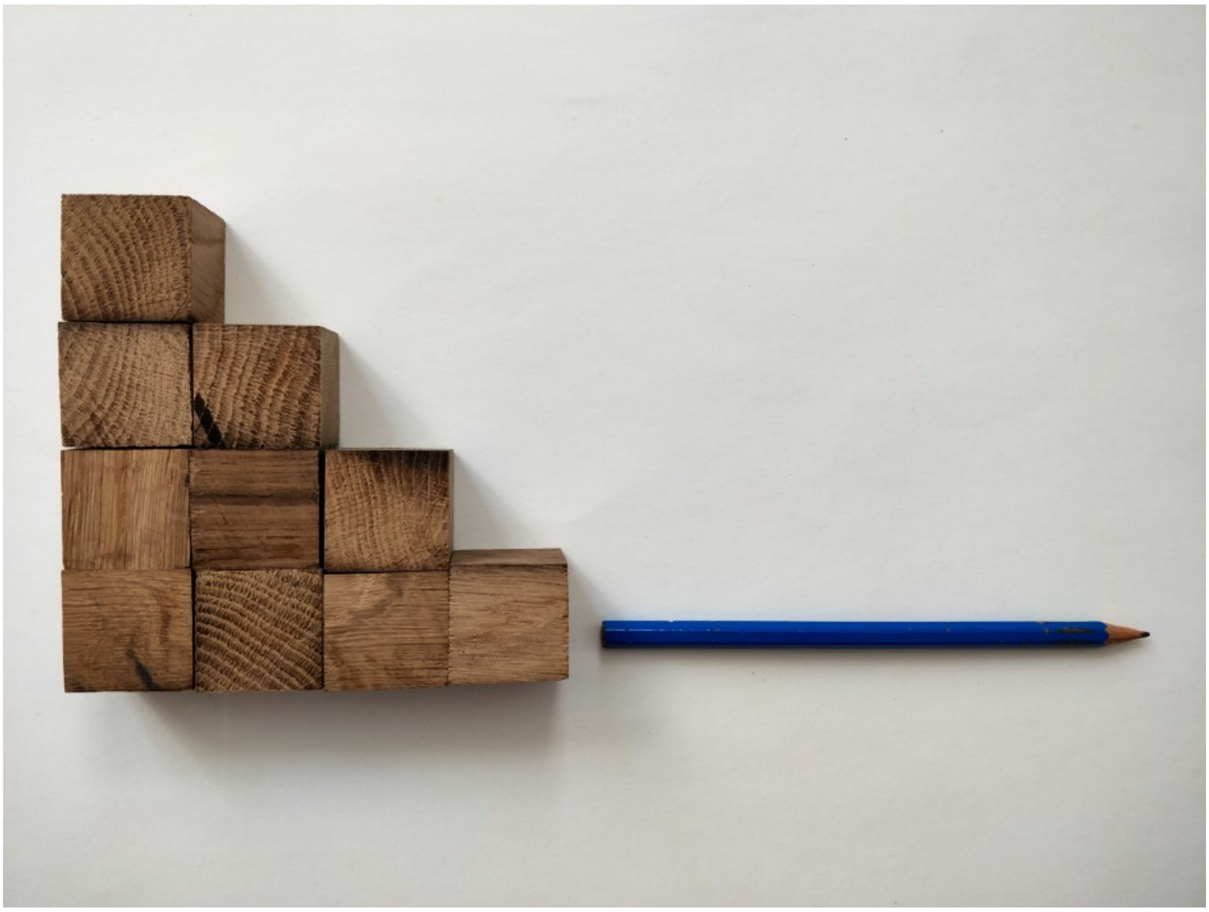
**Enjoy**. Well done! You discovered the Poisson distribution and you are halfway through the challenges. The random variable *X* equals the number of occurrences in the interval of interest. For the board game example, let *X* be the number of cubes a player picks in 15 minutes. If players pick 12 cubes in one hour. It means 12 cubes have been picked in four time intervals. So on average, 12*1/4 = 3 cubes every 15 minutes. The mean for the interval of interest is $\mu = 3$. We note $X \sim P(3)$ for time intervals $x = 1, 2, 3, 4.$ The mean and variance are equal. The standard deviation can be written $\sigma = \sqrt{3} = 1.732$. Extra: A Poisson process is a continuous-time stochastic (random) process which counts the arrival of randomly occurring events.

## Challenge 6: Exponential distribution

**Discover**. Consider each cube a value of 2. From left to right: start with 8 (4 cubes), then 6, then 4, then 2, then 0, 0, and 0. You might wonder how I cannot have cubes or place 0 cubes. Since the value 8 requires 4 cubes, the value 0 is continuous for the last three steps. To represent the zero, use the pencil and place it horizontally. Unlike the exponential function which we discovered earlier which is a function on real numbers, an exponential distribution is a function on measurable sets of real numbers.

**Present.** The exponential distribution gives the probability of each set from occurring. As a shortcut, it can be defined by a probability distribution function which is itself an exponential function, or by a cumulative distribution function which is 1 - an exponential function. That is to say, the inverse of the exponential function we previously discovered.

Examples range from reliability engineering systems function without failure, to traffic management, to physics, to hydrology (study of distribution and movement of water). For example, in physics it is often used to measure radioactive decay, in engineering it is used to measure the time associated with receiving a defective part on an assembly line. In traffic, if it is known that on average about ten cars cross a given highway every minute, the probability that seven cars will cross the same highway the next minute can be easily estimated using the exponential distribution. It can also be used to calculate the time between the passage of two consecutive cars, thus helping traffic managers to reduce traffic problems and avoid collisions.

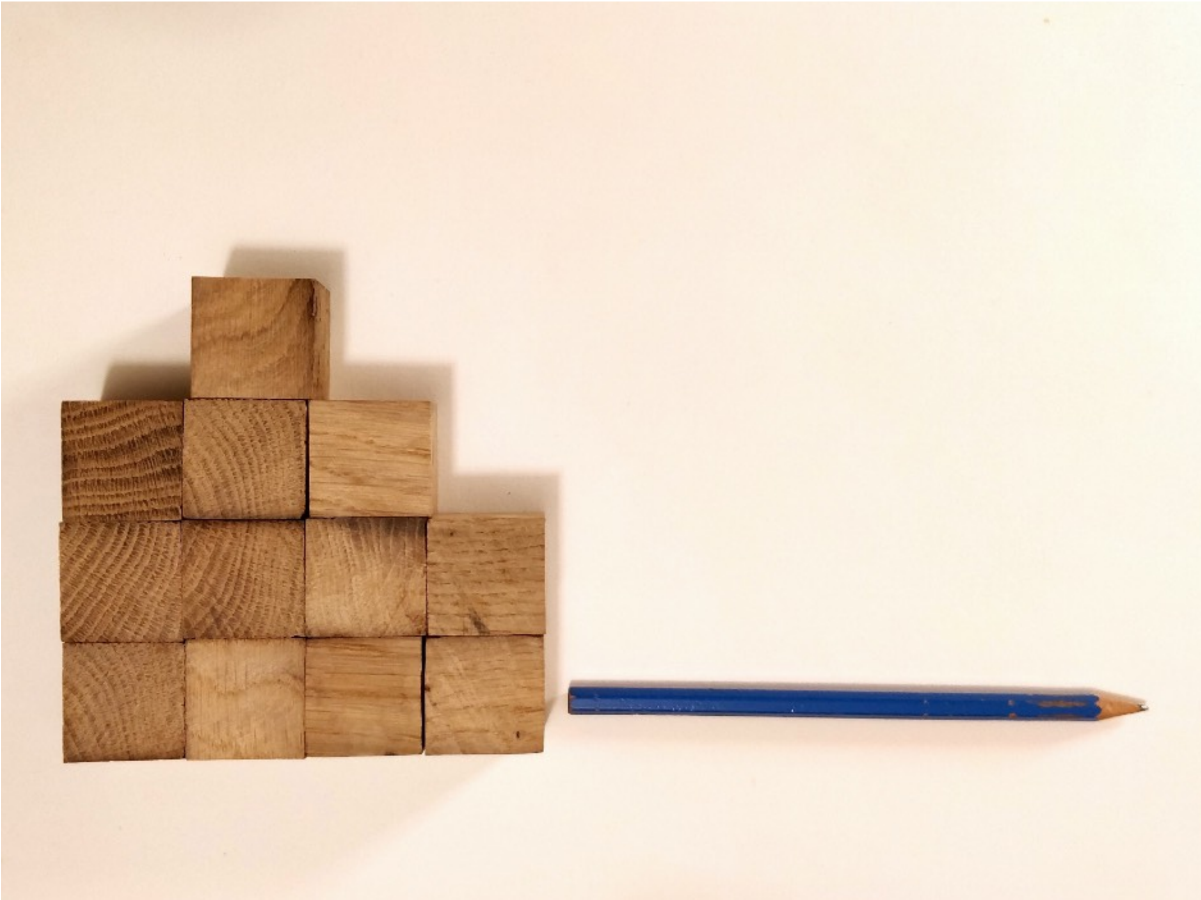


**Enjoy**. Congratulations! You just discovered the exponential distribution. It is commonly used to measure the expected time for an event to occur. Given a Poisson distribution with rate of change $\lambda$, the distribution of waiting times between successive changes (with $k = 0$) is $D(x) = 1- P(X > x) = 1 - e^{(-\lambda x)}$. The probability distribution function is $P(x) = \lambda e^{(-\lambda x)}$.

## Challenge 7: Gamma distribution

**Discover**. Gather the 12 cubes and construct from left to right: start with 3 cubes, then 4, then 3, then 2, then 0, 0, and 0. The representation of zeros may be done as earlier using the pencil.

If we reconsider the drawing cubes as part of a board game from Challenge 5. Poisson distribution. With 12 cubes in a bag and a time-dependent rule, we draw a random number of cubes. Let's repeat this experiment for this challenge but let's suppose that we do not know how many cubes there are in the bag. Consider asking a friend to help out here by changing the number of cubes present in the bag. We can suppose that one expects to get two cubes once every 15 min. But then you wonder, how likely is it that you will have to wait between 30 min to 1 hour before you catch 4 cubes? Using the supposition of 2 cubes every 15 min or 1/4 hour means that you can expect to get on average $2/0.25 = 8$ cubes every hour. The Gamma distribution you constructed models such events.



**Present**. A gamma distribution is a general type of statistical distribution that arises naturally in processes for which the waiting times between Poisson distributed events are relevant. That is to say, it can be considered as a waiting time between events distributed by Poisson distribution. Gamma distributions have either three, or two free parameters. They are the shape, the scale, and the threshold. Often statisticians set the threshold parameter to zero, then it is a two-parameter gamma distribution. If we hold the shape and scale parameters constant, the threshold shifts the distribution left and right. You can use your hand to shift all the cubes left or right from their current position.

Using this distribution, analysts can specify the number of events, for example by modeling the time until the 2nd or 3rd accident occurs. The Gamma distribution is used to model the size of insurance claims, rainfall, and many other events.

**Enjoy**. Kudos! You just discovered the Gamma distribution. A Gamma distribution with shape parameter $\alpha$ and scale parameter $\beta$ is the same as an exponential distribution when $\alpha = 1$ and $\beta = \lambda$. This distribution has a skewed shape. The skewness reduces as the value of $\alpha$ increases. The Gamma distribution is right skewed. The distribution will be positively skewed (the peak will be on the left side of the distribution, with relatively fewer observations on the right). This also means that the tail of the function (longer end) is towards the right.

Extra 1: What happens when the scale parameter (beta) changes? The scale parameter represents the variability present in the gamma distribution. Higher values cause the distribution to expand to the right and decrease in height. Conversely, lower values contract the distribution to the left and increase its height. Also, if you use the rate form of the parameter, higher rates narrow the distribution while lower rates spread it. The relationship between the scale parameter and the spread of the gamma distribution makes sense when you understand that the scale represents the average time between events. When the time between events is longer, the probabilities of extended times logically increase. In this case, you may imagine using your hand and pushing down the peak while the distribution slowly shifts to the right.

Extra 2: What happens when the shape parameter (alpha) changes? The shape parameter for the gamma distribution specifies the number of events you are modeling. For example, if you want to evaluate probabilities for the elapsed time of three accidents, the shape parameter equals 3. Indeed, increasing the shape leads to increasing the elapsed times. This makes sense because, holding everything else constant, we would expect the duration to increase when we increase the number of events. For example, three accidents ($\alpha = 3$) will take longer to occur than one ($\alpha = 1$), and five ($\alpha = 5$) will take longer than 3. In this case, you may imagine pushing down the distribution with your hand and obtaining a fairly similar outcome to Extra 1 but with a little more elongated or larger tail for larger elapsed times.

## Challenge 8: Bernoulli distribution

**Discover**. Consider each cube a value of 1. Use your pencil as an imaginary line and point it away from you. Gather all 12 cubes and take a few steps back. Prepare to throw each of the twelve cubes in the direction of the pencil. One at a time, each cube will fall towards the left or right of the pencil. Gather all the cubes from each side and put them on top of each other. For example, 5 cubes are on the left and 7 are on the right.

**Present.** This random experiment with only two possible outcomes (left or right) defines a Bernoulli trial. The probability of success is the same every time. The Bernoulli distribution is a discrete distribution for one trial. A discrete distribution is one in which the data can only take on certain values, for example, 1 or 2.

Since the experiment has two possible results, success or failure, many real-world examples exist. For example, a team will win or not win a championship, a student will pass or fail an exam, and a rolled die will show either a 6 or another number.

**Enjoy.** Congratulations! You just ran a Bernoulli trial by throwing the twelve cubes. The gathered cubes from each side are a representation of a Bernoulli distribution.

## Challenge 9: Binomial distribution

**Discover**. Rerun the previous throwing experiment, in Challenge 7, two times or more. Create a table for each trial with twelve cubes. For four trials:

| | t1 | t2 | t3 | t4 |
|---|---|---|---|---|
| Left | 6 | 7 | 5 | 4 |
| Right | 6 | 5 | 7 | 8 |

Suppose this is what we obtain after repeating the trial four times. We could continue further by doing another throw or by picking a random number from an interval such as from 0 to 1. If we want to maximize a trial to be successful, we already know that in the best case we get as many cubes on the left as there are on the right. In reality, this is not so perfect. If your cube throws were mostly going to the right side of the pencil and you did not make many trials, it is likely that your data is different from the assumed success rate in a Bernoulli trial.

**Present.** The binomial distribution gives the discrete probability distribution $P_p(n|N)$ of obtaining exactly $n$ successes out of $N$ Bernoulli trials (where the result of each Bernoulli trial is true with $p$ the probability of success and false with a probability of failure $q=1-p$). The sum of probabilities is always equal to 1. The shape of a binomial distribution is symmetrical when *p equals 0.5*. Essentially, this means that it looks like the normal distribution we previously made in Challenge 4. Alternatively, when the value is different it's skewed. When $p$ is greater than 0.5, it is positively skewed. When it's smaller, it's the opposite: negatively skewed. You may recreate the normal distribution again and shift the location of the two top central cubes at values 6 to the left. This represents a positively skewed distribution with the tail of the function (longer end) to the right.

There are many examples of binomial distributions in real life. For example, physicians use the binomial distribution to model the probability that a certain number of patients will experience side effects as a result of taking new drugs. Also if a new drug is introduced to cure a disease, it either cures the disease (it is a success) or it does not cure it (it is a failure). Email companies use the binomial distribution to model the probability of a certain number of spam messages landing in an inbox per day. Coin flips and die tosses are other gamified examples that assume a long-run frequency of 1/2 and 1/6, respectively.

**Enjoy**. Well done! You just discovered the binomial distribution. The long-run frequency is the ratio of the number of occurrences of an event in a large number of trials to the number of trials.
Extra: Bernoulli deals with the outcome of the single trial of the event, while Binomial deals with the outcome of multiple trials of the single event. That is to say, Bernoulli is used when the outcome of an event is required only once,

while Binomial is used when the outcome of an event is required multiple times.

## Challenge 10: Chi-squared distribution

**Discover**. Let's start this challenge with a thought experiment. Let us consider that one of the cubes is a die. In a perfect die, the long-run frequency is assumed to be 1/6. If the die is unfair, this means every side is not uniformly likely to happen. Let's assume that it is unfair. To see how this affects the long-frequency run, we repeat the experiment 500 times and get the following table.

| | | | | | | |
|---|---|---|---|---|---|---|
| Face | 1 | 2 | 3 | 4 | 5 | 6 |
| Freq | 49 | 92 | 75 | 63 | 55 | 166 |

If we read the observed frequencies for the six faces in the table, we notice that the die is indeed unfair. If it would be fair, we would expect a frequency of 100 for each of the six faces.

**Present**. The Chi-squared distribution is valuable to either determine whether or not a categorical variable follows a hypothesized distribution, or if there is a significant association between two categorical variables. With the Chi-squared distribution, we could look at every observed frequency *X* as being Poisson events each with the mean $\mu=\lambda=100$ and variance $\sigma^2=\lambda=100$. You may refer to concepts seen in Challenge 5. Poisson distribution. Since this does not match with the provided data or the observed frequencies in the above table, it is strong evidence that the die in the thought experiment is unfair.

Real-world examples that determine a significant association between two variables using the Chi-squared distribution are very common. For example, if gender is associated with the political party preference, or if marital status is associated with the education level.

**Enjoy**. Big Congratulations! You just discovered the Chi-squared distribution and you made it to the very last challenge! The Chi-Squared distribution is used in variety of ways: It describes the distribution of a sum of squared random variables, it is used to test the goodness of fit of a data distribution, or to determine whether data series are independent, and to estimate the confidence surrounding the variance and standard deviation of a random variable from a normal distribution.

A random variable can be a discrete one. In this case, a discrete random variable describes an event that has a specific set of values. For example, the discrete random variable that represents tossing a fair coin can only have the values heads or tails. The discrete random variable that represents choosing a card from a deck of cards can only have 52 possible values, 2 of hearts, 9 of clubs, queen of diamonds, ace of spades, etc.

A random sample of individuals may be obtained and categorized, then a chi-square test is conducted. If the *p-value* is below a certain threshold, then there is significant association. It is the threshold beyond which results can be declared statistically significant and it is set to 5 %. This means that the results are unlikely to be random.

## Extra Challenge: Distribution Comparisons

Gamma, exponential and Poisson distributions all model different characteristics of a Poisson process. In a Poisson process, independent events occur at a constant average rate. All these distributions can use lambda as a parameter, which represents this average rate of occurrence. Here is how to compare these three distributions.

The gamma distribution models the time between events. Time is a continuous variable, and the gamma distribution is, likewise, a continuous probability distribution. In contrast, the Poisson distribution models the number of events in a given time frame. A number is a discrete variable and the Poisson distribution is a discrete probability distribution.

Gamma and exponential distributions are equivalent when the gamma distribution has a shape value of 1. Recall that the shape value equals the number of events and the exponential distribution models the times for an event. Therefore, a gamma distribution with a shape = 1 is the same as an exponential distribution. Yet another example is possible, a gamma distribution with shape = 1 and scale = 3 is equivalent to an exponential distribution with scale = 3.