# Continuous Random Variables

Until now, we have always considered random variables with a finite or countable number of possible values. However, one often encounters the need of using random variables that can take any value in some interval: the lifetime of a radioactive particle, the size of a drop of rain, the distance from the Earth to the nearest galaxy in a given direction, etc.

In these situations, the framework changes radically. Mathematically, this is the equivalent of going from sums and series (discrete world) to integrals (continuous world). Conceptually, it is the equivalent of going from atoms to fluids: one has to abandon the notion of probability in a given point (it vanishes) and introduce a "probability density" just as one introduces a charge or mass density in physics.

## 1 Basic definitions

## Definition 1 (Continuous random variables, probability density)

A random variable $X$ is continuous if there exists a function $f : \mathbb{R} \longrightarrow \mathbb{R}$ such that:

$\star f \geq 0$

$\star \forall B \subset \mathbb{R}, P(X \in B) = \int_B f(x)dx$.

The function $f$ is called the probability density function associated to the random variable $X$. The set of all continuous random variables over $\Omega$ is denoted $\mathscr{L}(\Omega)$.

Remark: A probability density function is not a probability! In particular, nothing prevents $f$ from being larger than 1 . Indeed, the values that really matter are the integrals of $f$ (that is, the area under the graph of $f$ ).

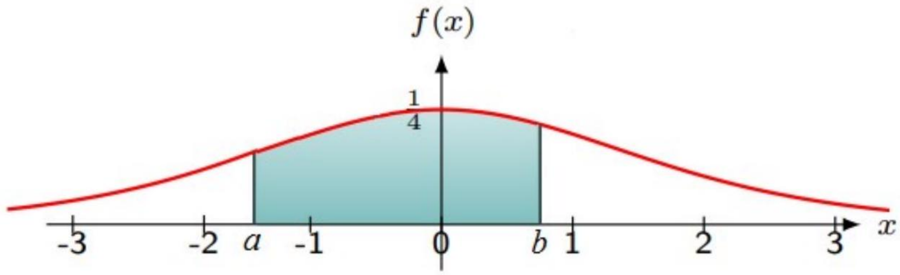



Figure 1.1 - The function $f$ defined by $\forall x \in \mathbb{R}, f(x) = \frac{e^{-x}}{(1+e^{-x})^2}$ is a probability density function.

The probability $P(a \leq X \leq b)$ is the area under the graph of $f$ in the interval $[a, b]$.

## Proposition 1

Let $X$ be a continuous random variable and denote by $f$ its probability density function. Then:

$\star P(a \leq X \leq b) = \int_a^b f(x)dx$.
$\star \int_{-\infty}^{+\infty} f(x)dx = P(X \in \mathbb{R}) = 1$.

## Examples:

$\star$ The lifetime of a computer, in hours of usage, can be modelled as a continuous random variable $T$ whose probability density function is

$$f(t) = \begin{cases} \frac{1}{1000}e^{-t/1000} \text{ if } t \geq 0 \\ 0 \text{ sinon.} \end{cases}$$

One can check that indeed $\int_{-\infty}^{+\infty} f(t)dt = 1$. If one wishes to find the probability that a computer will still be functioning after 2000 hours of usage, one must compute

$$P(T \geq 2000) = \frac{1}{1000}\int_{2000}^{+\infty} e^{-t/1000}dt = e^{-2} \approx 0.14$$

$\star$ In the case of an alpha radioactive decay, an atom $X$ emits a helium particle (called $\alpha$ particle) and transforms into a lighter atom $Y$ :

$${}^{\mathrm{A}}_{\mathrm{Z}}\mathrm{X} \longrightarrow {}^{\mathrm{A-4}}_{\mathrm{Z-2}}\mathrm{Y} + {}^{4}_{2}\mathrm{He}$$

In the reference frame of the particle $X$, one cannot know in advance the direction in which the alpha particle will be emitted but all directions are equally likely. In two dimensions, the direction of emission is then described as a continuous random variable whose probability density function is

$$f(\theta) = \begin{cases} \frac{1}{2\pi} \text{ if } \theta \in [0, 2\pi[ \\ 0 \text{ otherwise.} \end{cases}$$

## - Proposition 2

Let $a \in \mathbb{R}$ and $X$ a continuous random variable. Then, we have $P(X = a) = 0$.

Proof. Since $X$ is continuous, we have

$$P(X = a) = \int_a^a f(x)dx = 0$$

Remark: Non trivial consequence of this seemingly anodyne result: for a continuous variable, $X(\Omega)$ must be uncountable (otherwise we would have $\sum_{x \in X(\Omega)} P(X = x) = \sum_{x \in X(\Omega)} 0 = 0$, which is absurd) and therefore the sample space $\Omega$ is also uncountable! But we will bury our head in the sand and ignore the subtleties that uncountable spaces bring with them (see section 1.4), since in practice we will always take for the subsets $B \subset \mathbb{R}$ intervals or unions of intervals.

## Definition 2 (Cumulative distribution function)

We call cumulative distribution function of a random variable $X$, denoted $F_X(x)$ (or $F(x)$ if the associated random variable is obvious from context), the function defined, for any real number $x \in \mathbb{R}$, by

$$F(x) = P(X \leq x)$$

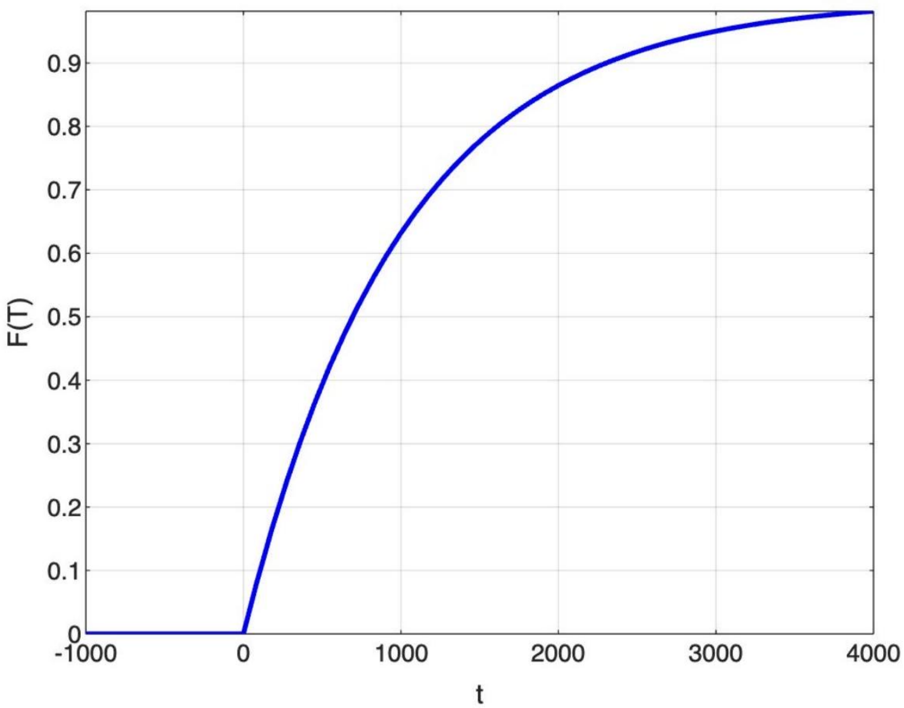
✓ Example: Continuing the example of the lifetime of a computer, modelled as a continuous random variable $T$ whose probability density function is

$$f(t) = \begin{cases} \frac{1}{1000} e^{-t/1000} \text{ if } t \leq 0 \\ 0 \text{ otherwise} \end{cases}$$

we have

$$F(t) = \begin{cases} 0 \text{ if } t \leq 0 \\ 1 - e^{-t/1000} \text{ if } t \geq 0 \end{cases}$$

Here, the cumulative distribution function $F$ represents the probability that the computer no longer functions at instant $t$. Graphically:




Remark: The definition is exactly the same for discrete or continuous random variables. Despite this, as we will see later on in this chapter, the cumulative distribution function is a much more important tool in the framework of continuous random variables than it was in the discrete case.

## Theorem 1 : Link between the cumulative distribution and the probability density functions

Let $X$ be a continuous random variable and denote by $f$ its probability density function and $F$ its cumulative distribution function. Then,

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

$\star f$ is the derivative of $F$.

$\star F$ is the unique anti-derivative of $f$ satisfying $\lim_{x\to+\infty} F(x) = 1$.

Proof. The theorem follows directly from the definition of the cumulative distribution function.

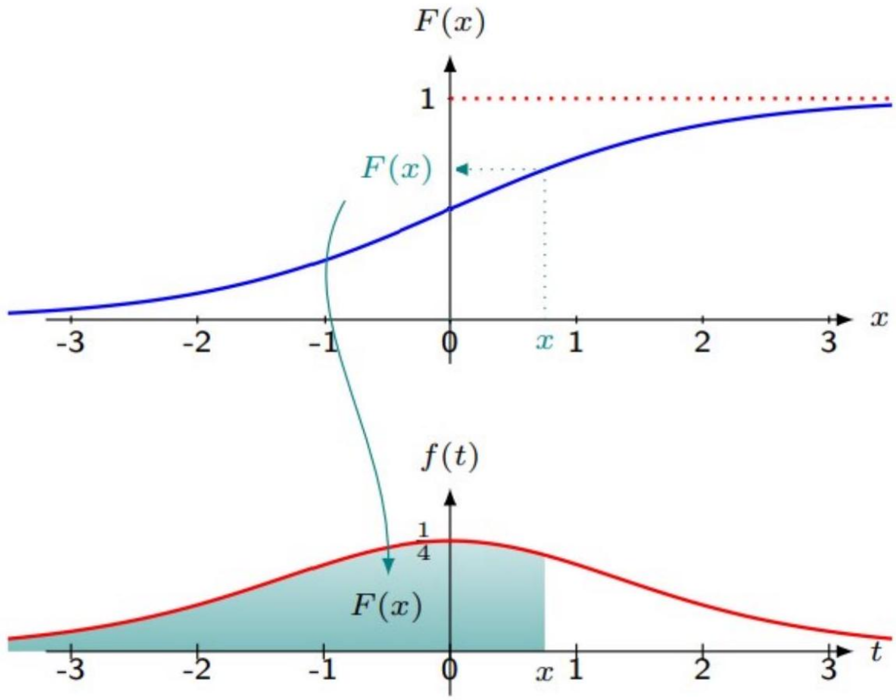



Figure 1.2 - Link between the cumulative distribution and the probability density functions.

The graph below shows the probability density function (in red). The area under the graph corresponds to the value of the cumulative distribution function (plotted above).

## Proposition 3 : Properties of the cumulative distribution function

Let $X$ be a continuous random variable and denote by $F$ its cumulative distribution function. Then:

$\star F$ is continuous.

$\star F$ is increasing.

$\star \lim_{x\to-\infty} F(x) = 0$.

$\star \lim_{x\to+\infty} F(x) = 1$.

$\star \forall (a,b) \in \mathbb{R}^2, P(a \leq X \leq b) = F(b) - F(a)$.

Proof. All these properties follow directly from the definition of the cumulative distribution function.

All what precedes shows that all probability computations about $X$ are completely determined if one knows the cumulative distribution function $F$ or the probability density function $f$.

## Methodological flashcard #1 (Probability distribution of $Y = g(X)$ in the continuous case)

Let $X$ be a continuous random variable of support $X(\Omega)$ and whose probability density function or cumulative distribution function is known. Let $g$ be a real-valued function defined over $X(\Omega)$ and $Y = g(X)$ a new random variable. To find the probability density function of $Y$ :

a) Determine the support $Y(\Omega)$.

b) Express the event $\{Y \leq a\}$ in terms of the random variable $X$.

c) Use this to deduce the cumulative distribution function of $Y$.

d) Differentiate to deduce the probability density function of $Y$.

✓ Example: Let $X$ be a continuous random variable whose probability density function is

$$f_X(x) = \begin{cases} 1 - |x| \text{ if } -1 \leq x \leq 1 \\ 0 \text{ otherwise} \end{cases}$$

Let us find the probability density function of $Y = \sqrt{|X|}$.

a) First, since $X(\Omega) = [-1, 1]$, we have $Y(\Omega) = [0, 1]$.

b) Next, let $a \in \mathbb{R}$. To write the event $\{Y \leq a\}$, we distinguish three cases:

⋆ If $a < 0$, the event $\{Y \leq a\}$ is impossible.

⋆ If $a > 1$, the event $\{Y \leq a\}$ is certain.

⋆ If $a \in [0, 1]$, we have $Y \leq a \Longleftrightarrow \sqrt{|X|} \leq a \Longleftrightarrow |X| \leq a^2 \Longleftrightarrow -a^2 \leq X \leq a^2$.

c) Therefore, for the cumulative distribution function of $Y$, we find:

⋆ If $a < 0, F_Y(a) = P(Y \leq a) = 0$.

⋆ If $a > 0, F_Y(a) = P(Y \leq a) = 1$.

⋆ If $a \in [0, 1]$,

$$\begin{aligned} F_Y(a) = P(Y \leq a) &= P\left(-a^2 \leq X \leq a^2\right) \\ &= \int_{-a^2}^{a^2} (1 - |x|) dx \\ &= 2 \int_0^{a^2} (1 - x) dx \\ &= 2a^2 - a^4. \end{aligned}$$

(One can check that $F(0) = 0$ and $F(1) = 1$.) d) Lastly, by differentiating $F_Y$, one obtains the probability density function of $Y$ :

$$f_Y(y) = \begin{cases} 4y\left(1 - y^2\right) \text{ if } y \in [0, 1] \\ 0 \text{ otherwise.} \end{cases}$$

(One can check that indeed $\int_{-\infty}^{+\infty} f_Y(y) dy = 1$.)

## 2 Expected Value, Variance, Standard deviation

# Definition 3 (Expected value of a continuous random variable)

Let $X$ be a continuous random variable. If it exists, the expected value of $X$ is the quantity

$$E(X)=\int_{-\infty}^{+\infty} x f(x) d x$$

The set of all continuous random variables over $\Omega$ whose expected value exist is denoted $\mathscr{L}_1(\Omega)$.

✓ Examples:

⋆ Coming back to the example of the lifetime of a computer (pages 102 and 103), we have

$$E(T)=\frac{1}{1000} \int_0^{+\infty} t e^{-t / 1000} d t=\int_0^{+\infty} e^{-t / 1000} d t=1000$$

⋆ Here is one well known example of a continuous random variable whose expected value diverges: we say that $X$ follows a standard Cauchy distribution [1] if the probability density function is defined by

$$f(x)=\frac{1}{\pi\left(1+x^2\right)}$$

One can check that this indeed defines a probability density function since

$$\int_{-\infty}^{+\infty} f(x) d x=\frac{1}{\pi}[\arctan (x)]_{-\infty}^{+\infty}=1$$

Moreover, the expected value of $X$ is given by

$$E(X)=\int_{-\infty}^{+\infty} x f(x) d x=\int_{-\infty}^{+\infty} \frac{x}{\pi\left(1+x^2\right)} d x$$

and the integral diverges since $x f(x) \underset{+\infty}{\sim} \frac{1}{\pi x}$.

![](https://cdn.mathpix.com/cropped/2024_02_14_e2257b03d51125f41c35g-2.jpg?height = 948width = 1205top_left_y = 840top_left_x = 425)

Figure 5.4 - Plot of the probability density function for the standard Cauchy distribution.

[01] Despite what the name suggests, this distribution was first studied by Siméon Denis Poisson in 1824.

## - Proposition 4

The expected value is a linear map over the real vector space $\mathscr{L}_1(\Omega)$.

In other words, let $X$ and $Y$ be two continuous random variables over $\Omega$ whose expected values exist, and let $\alpha, \beta \in \mathbb{R}$. Then, we have

$$E(\alpha X + \beta Y) = \alpha E(X) + \beta E(Y).$$

## Theorem 2

Let $X \in \mathscr{L}_1(\Omega)$ be a continuous random variable and denote by $f_X$ the probability density function of $X$. Let $g$ be a real-valued function defined over $X(\Omega)$. Then, provided the expected value of $Y = g(X)$ exists, we have

$$E(g(X)) = \int_{X(\Omega)} g(x) f_X(x) dx$$

Proof. Admitted.

✓ Example: We randomly choose a number in the interval $[0,2]$ and then compute its square. To find the expected value of this random experiment, we denote by $X$ the number chosen at random. The probability density function of $X$ is hence

$$f_X(x) = \begin{cases} \frac{1}{2} \text{ if } x \in [0,2] \\ 0 \text{ otherwise.} \end{cases}$$

To compute $E\left(X^2\right)$, we simply use the previous theorem:

$$E\left(X^2\right) = \int_{-\infty}^{+\infty} x^2 f_X(x) dx = \frac{1}{2} \int_0^2 x^2 dx = \frac{4}{3}$$

Remark: In the previous example, one could have computed $E\left(X^2\right)$ using the methodological flashcard 106 to first find the probability density function of $Y = X^2$ and then compute $E(Y)$ :

⋆ Since $X(\Omega) = [0,2]$, we have $Y(\Omega) = [0,4]$.

⋆ Since $X$ and $Y$ are both positive, we have, for any $a \in [0,4], \{Y \leq a\} = \{X \leq \sqrt{a}\}$.

⋆ Therefore

$$F_Y(a) = P(Y \leq a) = P(X \leq \sqrt{a}) = \frac{1}{2} \int_0^{\sqrt{a}} dx = \frac{\sqrt{a}}{2}.$$

⋆ Thus, we find that the probability density function of $Y$ is equal to

$$f_Y(y) = \begin{cases} \frac{1}{4\sqrt{y}} \text{ if } y \in [0,4] \\ 0 \text{ otherwise.} \end{cases}$$

(En passant, we can see that, although the values of $X$ are equally likely, those of $X^2$ are not.)

⋆ Finally, we find for the expected value

$$E(Y) = \int_0^4 \frac{\sqrt{y}}{4} dy = \frac{4}{3}.$$

One can clearly see here how much the theorem simplifies our computations!

## Definition 4 (Variance and standard deviation of a continuous random variable)

Let $X$ be a random variable (continuous or discrete). If it exists, the variance of $X$ is the quantity

$$V(X) = E\left((X - E(X))^2\right) = E\left(X^2\right) - E(X)^2.$$

Moreover, we call standard deviation of $X$ the quantity $\sigma(X) = \sqrt{V(X)}$.

The set of all continuous random variables over $\Omega$ whose variance exists is denoted $\mathscr{L}_2(\Omega)$.

Remark: The definition and interpretation of the variance and standard deviation are exactly the same in the continuous and discrete frameworks.

## Examples:

⋆ Returning to the previous example (a number taken at random in the interval $[0,2]$ ), we clearly have $E(X) = 1$ and thus $V(X) = E\left(X^2\right) - E(X)^2 = \frac{1}{3}$ and $\sigma(X) = \frac{1}{\sqrt{3}} \approx 0.58$.

⋆ Concerning the example of the lifetime of a computer (pages 68, 73 and 73), we have

$$E\left(T^2\right) = \frac{1}{1000}\int_0^{+\infty} t^2 e^{-t/1000} dt = 2\int_0^{+\infty} te^{-t/1000} dt = 2\times 1000\int_0^{+\infty} e^{-t/1000} dt = 2\times(1000)^2$$

(where be have integrated by parts twice). Therefore, we have

$$V(X) = E\left(X^2\right) - E(X)^2 = 2\times(1000)^2 - (1000)^2 = 1000^2$$

and

$$\sigma(X) = 1000$$

## Proposition 5

Let $X \in \mathscr{L}_2(\Omega)$ be a continuous random variable and $a, b \in \mathbb{R}$. We have

$$V(aX + b) = a^2 V(X).$$

Proof. The proof is exactly the same as for 4.5 (page 76).

## Definition 5 (Moment of order $n$ )

Let $X$ be a random variable (continuous or discrete) and $n \in \mathbb{N}$. If it exists, the moment of order $n$ is the quantity

$$m_n(X) = E\left(X^n\right) = \sum_{x \in X(\Omega)} x^n P(X = x)$$

Similarly, if it exists, the centered moment of order $n$ is the quantity

$$\mu_n(X) = E\left((X - E(X))^n\right).$$

Remark: As suggested in the introduction, transforming the formulas from the discrete to the continuous case is achieved by the following transformation:

$$\begin{aligned} \text{Discrete} &\longleftrightarrow \text{Continuous} \\ x_i &\longleftrightarrow x \\ P\left(X = x_i\right) &\longleftrightarrow f(x)dx \\ \sum &\longleftrightarrow \int \end{aligned}$$

| | Discrete | Continuous |
|---|---|---|
| Expected value: $E(X)$ | $\sum x_i P\left(X = x_i\right)$ | $\int_{-\infty}^{+\infty} x f(x) dx$ |
| Moment of order 2: $E\left(X^2\right)$ | $\sum x_i^2 P\left(X = x_i\right)$ | $\int x^2 f(x) dx$ |
| Moment of order $k : E\left(X^k\right)$ | $\sum x_i^k P\left(X = x_i\right)$ | $\int x^k f(x) dx$ |
| Cumulative distribution function: $F(t) = P(X \leq t)$ | $\sum_{x_i \leq t} P\left(X = x_i\right)$ | $\int_{x \leq t} f(x) dx$ |
| $P(a \leq X \leq b)$ | $\sum_{a \leq x_i \leq b} P\left(X = x_i\right)$ | $\int_{a \leq x \leq b} f(x) dx$ |

Of course, this is not a rigorous transformation. It is stated here just to help you in perceiving the similarity. But one has to be cautious since this "translation" does not always work: for example in the continuous framework one has $P(X = k) = 0$ for any value $k$ but this is obviously not true in the discrete framework.