Statistics: Continuous Random Variables

S. Maset

Dipartimento di Matematica e Geoscienze, Università di Trieste

PEM 2017-2018

Outline

Introduction

We have defined a discrete random variable as a random variable $X : \Omega \to \mathbb{R}$ whose range $X(\Omega)$ is discrete.

Now we introduce the notion of a **continuous random variable**, which is a random variable whose range is an interval of $\mathbb{R}$ (the exact definition is given above).

This means that a continuous random variable can take any value within some interval.

Examples:

- In the experiment of the falling meteor, the distance between the impact point and our town is a random variable with range the interval $[0, \pi R_E]$, where R_E is the Earth's radius.
- In the experiment given by the life of a person, the lifespan and the height of the person (in adulthood) are random variables with range a finite interval of non-negative real numbers.

- Before to present the formal exact definition of a continuous random variable, we need to introduce the notion of a distribution function.

Definition

Let $X : \Omega \to \mathbb{R}$ be a random variable. The function $F_X : \mathbb{R} \to [0, 1]$ given by

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

is called the **distribution function** or the **cumulative distribution function** of X.

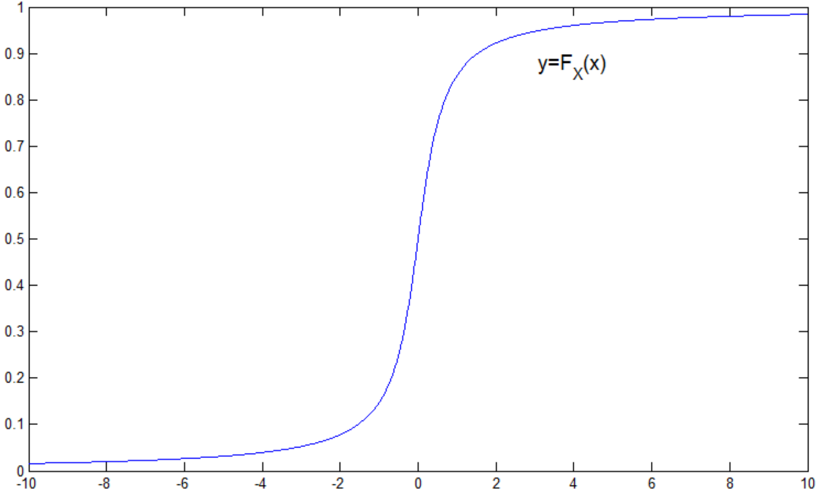
The distribution function F_X has the following three properties:

i) F_X is increasing;

ii) $\lim_{x \to -\infty} F_X(x) = 0$;

iii) $\lim_{x \to +\infty} F_X(x) = 1$.

Graph of a distribution function



In fact:

i) for $x, y \in \mathbb{R}$ such that $x \leq y$, we have

$$X \leq x \Rightarrow X \leq y, \text{ i.e. } X \leq x \subseteq X \leq y,$$

and so

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

ii) Since F_X is an increasing function, $\lim_{x\to-\infty} F_X(x)$ exists. Consider a decreasing sequence $\{x_n\}$ in $\mathbb{R}$ such that $x_n \to -\infty$, $n \to \infty$. Consider the events

$$X \leq x_1 \supseteq X \leq x_2 \supseteq X \leq x_3 \supseteq \cdots$$

By the lower monotone convergence property, we have

$$\begin{aligned}
\lim_{x\to-\infty} F_X(x) &= \lim_{n\to\infty} F_X(x_n) = \lim_{n\to\infty} \mathbb{P}(X \leq x_n) \\
&= \mathbb{P}\left(\bigcap_{n=1}^{\infty} X \leq x_n\right) = \mathbb{P}(\emptyset) = 0.
\end{aligned}$$

iii) Since F_X is an increasing function, $\lim_{x\to+\infty} F_X(x)$ exists. Consider an increasing sequence $\{x_n\}$ in $\mathbb{R}$ such that $x_n \to +\infty$, $n \to \infty$. Consider the events

$$X \leq x_1 \subseteq X \leq x_2 \subseteq X \leq x_3 \subseteq \cdots$$

By the upper monotone convergence property, we have

$$\begin{aligned}\lim_{x\to+\infty} F_X(x) &= \lim_{n\to\infty} F_X(x_n) = \lim_{n\to\infty} \mathbb{P}(X \leq x_n) \\ &= \mathbb{P}\left(\bigcup_{n=1}^{\infty} X \leq x_n\right) = \mathbb{P}(\Omega) = 1.\end{aligned}$$

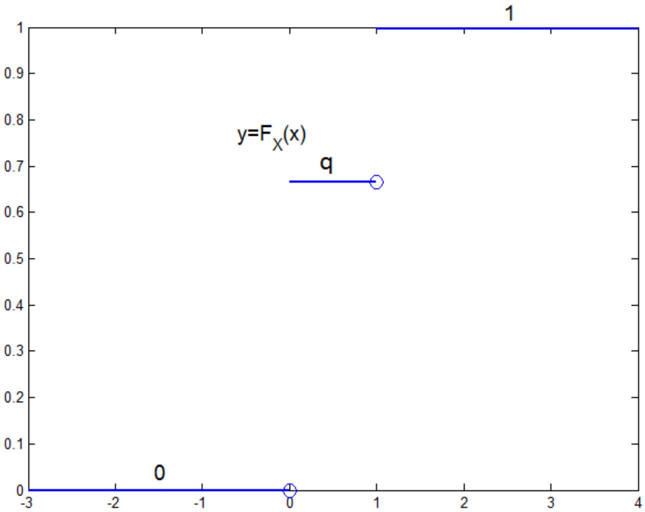
As a very simple example of a distribution function, consider the following discrete random variable X related to a single trial in a Bernoulli process with outcomes α with probability p and β with probability q:

$$X = \begin{cases} 1 \text{ if the outcome is } \alpha \\ 0 \text{ if the outcome is } \beta. \end{cases}$$

We have

$$F_X(x) = \mathbb{P}(X \leq x) = \begin{cases} 0 \text{ if } x < 0 \\ q \text{ if } 0 \leq x < 1 \\ 1 \text{ if } x \geq 1 \end{cases}, \ x \in \mathbb{R}.$$

Graph of F_X



In general, for a discrete random variable X whose values are y_i, $i \in I$, where $I = \{1, 2, \ldots, n\}$ for some positive integer n or $I = \{1, 2, 3, \ldots\}$, and $y_1 < y_2 < y_3 < \cdots$, we have

$$\begin{aligned} F_X(x) &= \mathbb{P}(X \leq x) = \mathbb{P}\Big(\bigcup_{\substack{i \in I \\ y_i \leq x}} X = y_i \Big) = \sum_{\substack{i \in I \\ y_i \leq x}} \mathbb{P}(X = y_i) \\ &= \sum_{\substack{i \in I \\ y_i \leq x}} f_X(y_i) = \begin{cases} 0 \text{ if } x < y_1 \\ f_X(y_1) \text{ if } y_1 \leq x < y_2 \\ f_X(y_1) + f_X(y_2) \text{ if } y_2 \leq x < y_3 \\ f_X(y_1) + f_X(y_2) + f_X(y_3) \text{ if } y_3 \leq x < y_4 \\ \ldots \end{cases}, \ x \in \mathbb{R} \end{aligned}$$

So, F_X is piecewise constant with jumps at the points y_i, $i \in I$.

Moreover, the previous expression of F_X explains why the name "cumulative distribution function" is used.

Exercise. Prove that the distribution function F_X of a random variable X is right-continuous, i.e.

$$\lim_{y \downarrow x} F_X(y) = F_X(x) \text{ for any } x \in \mathbb{R}.$$

To this aim, observe that $\lim_{y \downarrow x} F_X(y)$ exists since F_X is increasing. Then consider a decreasing sequence $\{y_n\}$ in $\mathbb{R}$ such that $y_n \to x$, $n \to \infty$, and the events

$$X \leq x_1 \supseteq X \leq x_2 \supseteq X \leq x_3 \supseteq \cdots$$

Continuous Random Variables and Probability Density Functions

Definition

A random variable X is called **continuous** if F_X has a derivative. The derivative of F_X is denoted by f_X and it is called the **probability density function** (**pdf**) of X.

The pdf of X is also called the **distribution** of X.

Observe that f_X is a non-negative function since F_X is increasing.

- Local meaning of the pdf: for any $x \in \mathbb{R}$, we have

$$\begin{aligned} f_X(x)dx &= F_X(x+dx) - F_X(x) = \mathbb{P}(X \leq x+dx) - \mathbb{P}(X \leq x) \\ &= \mathbb{P}(x < X \leq x+dx) \end{aligned}$$

where the last equality holds since the event $X \leq x$ is included in the event $X \leq x + dx$ and the the event $x < X \leq x + dx$ is their difference.

We can write

$$f_X(x) = \frac{\mathbb{P}(x < X \leq x+dx)}{dx}$$

and this explain the name "probability density function" for f_X: the values of the probability are distributed along the real line and $f_X(x)$ is the probability for length unit at the point x, i.e. the linear density of probability at x.

On the other hand, for a discrete random variable X, the values

$$f_X(x) = \mathbb{P}(X = x), \; x \in X(\Omega),$$

of the "probability mass function" are values of probability concentrated at the points of $X(\Omega)$ and so such points have a mass of probability.

- Global meaning of the pdf: for any $a, b \in \mathbb{R}$ with $a < b$, we have

$$\int_a^b f_X(x)\,dx = F_X(b) - F_X(a) = \mathbb{P}(X \le b) - \mathbb{P}(X \le a) = \mathbb{P}(a < X \le b).$$

By letting $a \to -\infty$, we obtain

$$\int_{-\infty}^b f_X(x)\,dx = \lim_{a \to -\infty} \int_a^b f_X(x)\,dx = \lim_{a \to -\infty} (F_X(b) - F_X(a)) = F_X(b) - \lim_{a \to -\infty} F_X(a) = F_X(b) - 0 = F_X(b).$$

Thus,

$$F_X(x) = \int_{-\infty}^x f_X(y)dy, \; x \in \mathbb{R}.$$

- For a continuous random variable X, F_X is a continuous function since F_X is an integral function.

  Observe that a discrete random variable X cannot be a continuous random variable, since F_X is not a continuous function (it is piecewise constant with jumps).

- Let X be a random variable such that F_X is continuous. Then, we have

$$\mathbb{P}(X = x) = 0 \text{ for any } x \in \mathbb{R}.$$

In fact, consider an increasing sequence $\{x_n\}$ in $\mathbb{R}$ such that $x_n \to x$, $n \to \infty$. Consider the events

$$x_1 < X \leq x \supseteq x_2 < X \leq x \supseteq x_3 < X \leq x \supseteq \cdots$$

By the lower monotone convergence property, we have

$$\lim_{n\to\infty} \mathbb{P}(x_n < X \leq x) = \mathbb{P}(\bigcap_{n=1}^{\infty} x_n < X \leq x) = \mathbb{P}(X = x).$$

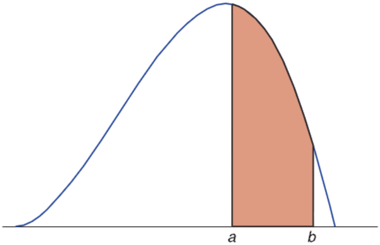
and

$$\begin{aligned}
\lim_{n\to\infty} \mathbb{P}(x_n < X \leq x) &= \lim_{n\to\infty} (\mathbb{P}(X \leq x) - \mathbb{P}(X \leq x_n)) \\
&= \lim_{n\to\infty} (F_X(x) - F_X(x_n)) \\
&= F(x) - \lim_{n\to\infty} F_X(x_n) \\
&= F_X(x) - F_X(x) \text{ since } F_X \text{ is continuous} \\
&= 0.
\end{aligned}$$

- Now, let X be a continuous random variable, so that F_X is continuous. For $a, b \in \mathbb{R}$ with $a < b$, we have

$$\int_a^b f_X(x)\,dx = \mathbb{P}(a < X \leq b) = \mathbb{P}(a \leq X \leq b)$$
$$= \mathbb{P}(a \leq X < b) = \mathbb{P}(a < X < b),$$

since $\mathbb{P}(X = a) = \mathbb{P}(X = b) = 0$ and all these probabilities are the area under the graph di f_X between a and b



$P\{a \leq X \leq b\}$ = area of shaded region

Once we know that probabilities of the events

$$X \in \text{ closed box of } \mathbb{R},$$

i.e the probabilities

$$\mathbb{P}(X \in [a, b]) = \int_a^b f_X(x)\, dx, \;\; a, b \in \mathbb{R} \text{ with } a \leq b,$$

we also know the probabilities of the events

$$X \in \text{ Borel subset of } \mathbb{R}.$$

These probabilities are obtained by the same way with which the Borel subsets of $\mathbb{R}$ are constructed from the closed boxes of $\mathbb{R}$, by applying the properties of a measure of probability.

We obtain, for any Borel subset A of $\mathbb{R}$,

$$\mathbb{P}(X \in A) = \int_{x \in A} f_X(x)\, dx.$$

In particular, we have

$$\int\limits_{x\in\mathbb{R}} f_X(x)\,dx = \mathbb{P}(X\in\mathbb{R}) = 1,$$

i.e. the area under the whole graph of f_X is 1. On the other hand, this can be seen by observing that

$$\int\limits_{x\in\mathbb{R}} f_X(x)\,dx = \lim_{b\to+\infty}\int\limits_{-\infty}^{b} f_X(x)\,dx = \lim_{b\to+\infty} F_X(b) = 1.$$

Finally, observe that, since

$$\mathbb{P}(X = x) = 0 \text{ for any } x\in\mathbb{R},$$

for a discrete subset $A = \{a_i : i\in I\}$ of $\mathbb{R}$, we have

$$\mathbb{P}(X\in A) = \mathbb{P}\left(\bigcup_{i\in I} X = a_i\right) = \sum_{i\in I}\underbrace{\mathbb{P}(X = a_i)}_{=0} = 0.$$

- Here is an interesting interpretation of the pdf of a continuous random variable.

  Consider an experiment with sample space Ω and a continuous random variable $X : \Omega \to \mathbb{R}$.

  Suppose to consider as the new outcome for the experiment $X(\omega)$, rather than ω. So, the new sample space is the continuous set $\Omega^{\text{new}} = \mathbb{R}$, rather than Ω.

  The non-negative integrable function $p : \Omega^{\text{new}} = \mathbb{R} \to \mathbb{R}$ giving the probabilities of the closed boxes of the new sample space $\Omega^{\text{new}} = \mathbb{R}$ is the pdf of X:

$$\mathbb{P}([a, b]) = \mathbb{P}(X \in [a, b]) = \int_a^b f_X(x)\,dx$$

  for a closed box $[a, b]$ of $\Omega^{\text{new}} = \mathbb{R}$.

- Now, we see a first example of a continuous random variable.

  A continuous random variable X is said have the **uniform distribution** $\mathrm{U}(a, b)$, where $a, b \in \mathbb{R}$ with $a < b$, if

$$f_X(x) = \begin{cases} 0 \text{ if } x < a \\ \frac{1}{b-a} \text{ if } a \leq x \leq b \\ 0 \text{ if } x > b \end{cases}, \quad x \in \mathbb{R}.$$

  So, if X has the uniform distribution $\mathrm{U}(a, b)$, then

$$F_X(x) = \int_{-\infty}^{x} f_X(y)dy = \begin{cases} \int_{-\infty}^{x} 0 = 0 \text{ if } x < a \\ \int_{a}^{x} \frac{1}{b-a} dx = \frac{x-a}{b-a} \text{ if } a \leq x \leq b \\ \int_{a}^{b} \frac{1}{b-a} dx = 1 \ \text{ if } x > b \end{cases}, \quad x \in \mathbb{R},$$

and, for any Borel subset A of $[a, b]$,

$$\mathbb{P}(X \in A) = \int_{y \in A} f_X(y)\, dy = \int_{y \in A} \frac{1}{b-a} dy = \frac{\int_{y \in A} dy}{b-a} = \frac{\text{length}(A)}{b-a}.$$

Here are two examples of random variables with uniform distribution:

- *The number obtained by a random number generator in a computer is a random variable with uniform distribution $U(0, 1)$.*
- *Consider a railway connecting the city A (at the position 0 in the railway) to the city B (at the position D in the railway). The position in the railway where the next train with a malfunction will stop is a random variable with uniform distribution $U(0, D)$.*

Exercise. What is the probability that the number obtained by a random number generator is rational?

Exercise. Consider the next time that there will be here an earthquake. Is the time during the day (from 0 h to 24 h) at which the earthquake will strike an uniform random variable? Is the length of time interval from now to the moment when the earthquake will strike a uniform random variable?

Exercise. Suppose it is known the final result 1-0 of a football match, but the time at which the goal is scored is not known. Is this time a uniform random variable?

f_X as derivative of F_X

- We have a defined a continuous random variable as a random variable whose distribution function has a derivative, called the pdf, but a better definition is the following one.

  A random variable X is called continuous if there exists a function f_X, called a pdf of X, such that

$$F_X(x) = \int_{-\infty}^{x} f_X(y)\, dy, \ x \in \mathbb{R}.$$

  So f_X is meant as a derivative in a weaker sense than the usual one: we only require that F_X is the integral function of f_X.

By recalling Analysis courses, we can say the distribution function F_X of a continuous random variable X, as any integral function, is continuous at any point of $\mathbb{R}$ and differentiable at any point of $\mathbb{R}$ except for a set a points of measure (length) zero.

The derivative F'_X, which is defined at any point of $\mathbb{R}$ except for a set a points of measure zero, is only a particular pdf of X, i.e. it is only a particular function whose integral function is F_X.

We have that each pdf of X, i.e. each function whose integral function is F_X, coincides with F'_X except for a set of points of measure zero.

So the pdf f_X of X is unique except for a set of points of measure zero: observe that by changing the values of f_X in a set of points of measure zero does not modify the integrals

$$F_X(x) = \int_{-\infty}^{x} f_X(y)dy, \ x \in \mathbb{R}.$$

Example. if X has the uniform distribution $\mathrm{U}(a, b)$, then

$$F_X(x) = \begin{cases} 0 \text{ if } x < a \\ \\ \frac{x-a}{b-a} \text{ if } a \le x \le b \\ \\ 1 \text{ if } x > b \end{cases}, \ x \in \mathbb{R},$$

is continuous at any point and differentiable at any point except a and b: we have

$$F'_X(x) = \begin{cases} 0 \text{ if } x < a \\ \frac{1}{b-a} \text{ if } a < x < b \\ 0 \text{ if } x > b \end{cases}, \ x \in D := \mathbb{R} \setminus \{a, b\},$$

Thus f_X is any function coinciding con F'_X except for a set of points of measure zero.

Rule of transformation for pdfs

- We have the following **rule of transformation for pdfs.**

  Let X be a continuous random variable and let Y be a random variable implicitly defined by

  $$X = \psi(Y)$$

  where $\psi : I \to \mathbb{R}$, I interval of $\mathbb{R}$, is differentiable and strictly increasing. Observe that $Y(\Omega) \subseteq I$. Then Y is a continuous random variable with pdf

  $$f_Y(y) = f_X(\psi(y)) \cdot \psi'(y), \; y \in I.$$

  In fact, since ψ is strictly increasing we have

  $$F_Y(y) = \mathbb{P}(Y \leq y) = \mathbb{P}(\psi(Y) \leq \psi(y)) = \mathbb{P}(X \leq \psi(y)) = F_X(\psi(y)), \; y \in I,$$

  and so, since ψ is differentiable, we have

  $$f_Y(y) = F'_Y(y) = F'_X(\psi(y))\psi'(y) = f_X(\psi(y))\psi'(y), \; y \in I.$$

  Exercise. Suppose that $I \neq \mathbb{R}$. How is it defined the pdf of Y outside I?

Example. Consider a random variable X with $X(\Omega) = I = [0,1]$ and distribution $\mathrm{U}(0,1)$. We determine the distribution of $Y = X^{\alpha}$, where $\alpha > 0$. We have $X = \psi(Y)$, where $\psi : I \to I$ is given by

$$\psi(y) = y^{\frac{1}{\alpha}}, y \in I.$$

Thus

$$f_Y(y) = f_X(\psi(y))\,\psi'(y) = 1 \cdot \frac{1}{\alpha} y^{\frac{1}{\alpha}-1} = \frac{1}{\alpha} y^{\frac{1}{\alpha}-1},\; y \in I.$$

- Exercise. Prove this other rule of transformation for pdfs. Let X be a continuous random variable and let Y be a random variable implicitly defined by $X = \psi(Y)$, where $\psi : I \to \mathbb{R}$, I interval of $\mathbb{R}$, is differentiable and strictly decreasing. Then Y is a continuous random variable with pdf

$$f_Y(y) = -f_X(\psi(y))\psi'(y), \; y \in I.$$

Exercise. Let X be a continuous random variable. Find the pdf of $Y = aX + b$, where $a, b \in \mathbb{R}$ with $a \neq 0$, in terms of the pdf f_X.

Normal Random Variables

- The most important type of random variable is the normal random variable, or gaussian random variable.

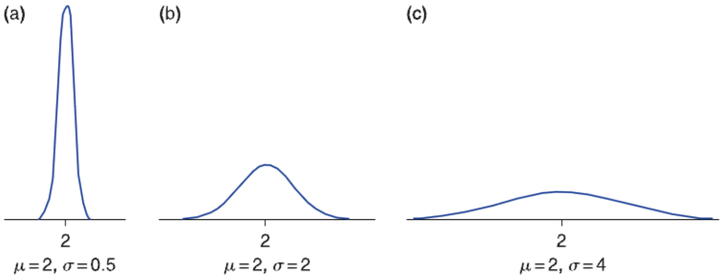
Definition

A random variable X is said to have the **normal distribution** $\mathrm{N}(\mu, \sigma^2)$, where $\mu \in \mathbb{R}$ and $\sigma > 0$, if it is a continuous random variable with pdf

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

X is said a **normal random variable** if it has some normal distribution $\mathrm{N}(\mu, \sigma^2)$.

The pdf f_X of a normal random variable X is a bell-shaped curve.



The curve is symmetric about μ.

In fact, for any $c > 0$, we have

$$\begin{aligned} f_X(\mu + c) &= \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(\mu+c-\mu)^2}{2\sigma^2}} = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{c^2}{2\sigma^2}} \\ &= \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(-c)^2}{2\sigma^2}} = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(\mu-c-\mu)^2}{2\sigma^2}} = f_X(\mu - c)\,. \end{aligned}$$

The variability of the curve is measured by σ: the curve flattens out as σ increases and the peak at μ is smaller as σ increases.

In fact, we have

$$f'_X(x) = \frac{d}{dx}\left(\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}\right) = -\frac{1}{\sigma^3\sqrt{2\pi}}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}}\,(x-\mu)\,,\ x \in \mathbb{R},$$

and

$$|f'_X(x)| = \frac{1}{\sigma^3\sqrt{2\pi}}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}}\,|x-\mu|,\ x \in \mathbb{R},$$

with

$$\frac{d}{d\sigma}\left(\frac{1}{\sigma^3}e^{-\frac{(x-\mu)^2}{2\sigma^2}}\right) = -\frac{3}{\sigma^4}e^{-\frac{(x-\mu)^2}{2\sigma^2}} + \frac{1}{\sigma^6}e^{-\frac{(x-\mu)^2}{2\sigma^2}}(x-\mu)^2$$

$$= \frac{e^{-\frac{(x-\mu)^2}{2\sigma^2}}}{\sigma^6}\left(-3\sigma^2 + (x-\mu)^2\right) < 0,\ \sigma^2 > \frac{(x-\mu)^2}{3},$$

and so at the same $x \in \mathbb{R}$, for $\sigma \geq \frac{|x-\mu|}{\sqrt{3}}$, $|f'_X(x)|$ decreases as σ increases and tends to zero, as $\sigma \to +\infty$.

Moreover, since $f'_X(x)$ is positive for $x < \mu$ and negative for $x > \mu$, f_X at μ has the maximum value

$$f_X(\mu) = \frac{1}{\sigma\sqrt{2\pi}}$$

that decreases as σ increases.

- We also observe that the presence of the factor $\frac{1}{\sigma\sqrt{2\pi}}$ in front of the exponential $e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ is due to the need of the normalization

$$\int\limits_{x\in\mathbb{R}} f_X(x)\,dx = 1.$$

In fact, we have

$$\int\limits_{x\in\mathbb{R}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \sigma\sqrt{2\pi}.$$

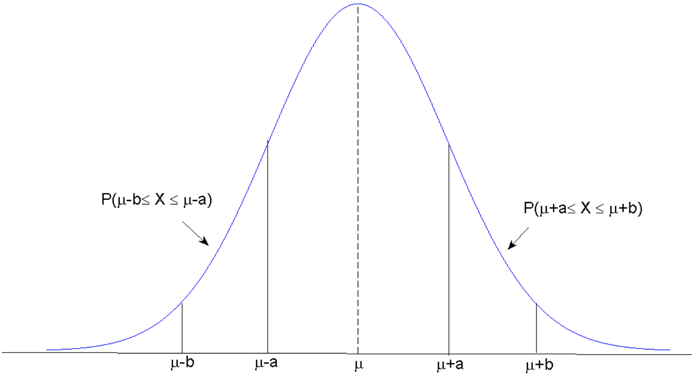
Exercise. Prove this starting from the fact that

$$\int\limits_{y\in\mathbb{R}} e^{-y^2} dy = \sqrt{\pi}.$$

- Exercise. Find the inflection points of the pdf f_X.

- Finally, we observe that the symmetry of the pdf f_X around μ implies: for any $a, b \in [0, +\infty]$ with $a < b$, we have

$$\mathbb{P}(\mu + a \leq X \leq \mu + b) = \mathbb{P}(\mu - b \leq X \leq \mu - a).$$

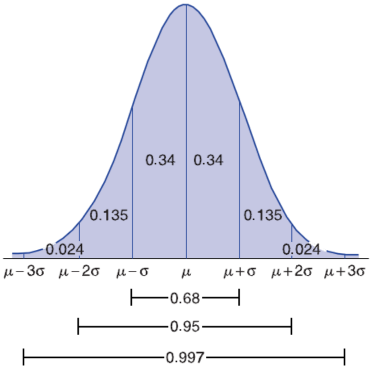


In fact

$$\mathbb{P}(\mu + a \leq X \leq \mu + b) = \int_{\mu+a}^{\mu+b} f_X(x)\, dx = \int_a^b f_X(\mu + y)\, dy, \; y = x - \mu,$$

$$= \int_a^b f_X(\mu - y)\, dy = -\int_{\mu-a}^{\mu-b} f_X(x)\, dx, \; x = \mu - y,$$

$$= \int_{\mu-b}^{\mu-a} f_X(x)\, dx = \mathbb{P}(\mu - b \leq X \leq \mu - a).$$

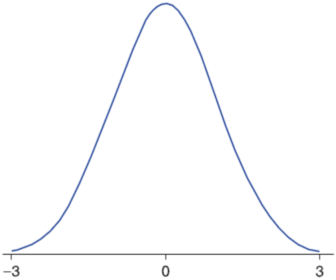


In Figure, we see the probabilities of some intervals $[\mu + a, \mu + b]$ with a and b multiples of σ. From the figure we can deduce that

$$\mathbb{P}\left(-\sigma \leq X - \mu \leq \sigma\right) = 68\%$$
$$\mathbb{P}\left(-2\sigma \leq X - \mu \leq 2\sigma\right) = 95\%$$
$$\mathbb{P}\left(-3\sigma \leq X - \mu \leq 3\sigma\right) = 99.7\%.$$

The standard normal distribution

- The distribution $\mathrm{N}(0, 1)$ is said the **standard normal distribution.** A random variable with the standard normal distribution is said a **standard normal random variable**.



If Z has the standard normal distribution, then

$$f_Z(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}, \; z \in \mathbb{R}.$$

- Let X be a random variable with normal distribution $\mathrm{N}(\mu, \sigma^2)$. Then

$$Z = \frac{X - \mu}{\sigma},$$

called the **standardized form** of X, is a standard normal random variable.

In fact,

$$X = \mu + \sigma Z = \psi(Z)$$

where

$$\psi(z) = \mu + \sigma z, \; z \in \mathbb{R},$$

and so

$$\begin{aligned} f_Z(z) &= f_X(\psi(z))\psi'(z) = f_X(\mu + \sigma z) \cdot \sigma \\ &= \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(\mu+\sigma z-\mu)^2}{2\sigma^2}} \cdot \sigma = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}, \; z \in \mathbb{R}. \end{aligned}$$

Importance of normal random variables

- The great importance of the normal random variables is due to the following fact:
  - in a process that produces a final result, which is programmed in some measure but it is also influenced by many random factors, numerical quantities related to this final result are normally distributed.

  Examples:

  - *Consider the formation process of an individual: when the process has produced the adult individual, its height, blood pressure, length of the feet, cholesterol level, etc., are normally distributed;*
  - *Consider the race of an athlete, for example a marathon or a cycling race: the time to complete the race is normally distributed.*
  - *Consider the weather in a given place and in a given month: quantities as the average of the temperatures and the quantity of rainfall during the month are normally distributed.*

Exercise. Which of the following random variables are normally distributed?

- The life time of a notebook battery.
- The time to travel by car from city A to city B.
- The distance from here to the impact point of the next meteor falling on the Earth.
- The birth weight of a newborn.
- The length of time interval from now to the moment when an earthquake will strike here in this place.
- The length of a piece produced by a machine in an industrial process.

- Once we know that many interesting quantities have a normal distribution, the big problem is to estimate the parameters μ and σ of a normal distribution. This is one of the main task of the Inferential Statistics.

  Another reason for which the normal distribution is important is the **Central Limit Theorem** that will be presented later.

- *Consider an industrial process producing pieces.*

  Let X be a numerical quantity related to a produced piece. For example, if the piece is a disk, X can be the diameter of the piece.

  We can assume that, in the experiment of the production of a piece, X is a random variable with normal distribution $\mathrm{N}(\mu, \sigma^2)$, where μ is some reference value for the quantity X.

  We consider the produced piece as defective if $|X - \mu| > \mathrm{TOL}$, where TOL is a given tolerance.

  Suppose that the standard deviation σ is such that $k\sigma \leq \mathrm{TOL}$.

  As a consequence, the probability that the produced piece will be defective is

$$p = \mathbb{P}(|X - \mu| > \mathrm{TOL}) \leq \mathbb{P}(|X - \mu| > k\sigma) = 1 - \mathbb{P}(|X - \mu| \leq k\sigma).$$

So, for $k = 3$ (Three-Sigma rule), we have

$$p \leq 1 - 99.7\% = 0.3\%$$

and, for $k = 6$ we have

$$p \leq 2 \cdot 10^{-9}.$$

Indeed, this value of p is unattainable.

What is known as Six-Sigma methodology is not the case $k = 6$ but a methodology with a probability of having a defective piece not larger than $3.4 \cdot 10^{-6}$. This is explained later on.

Independence of Continuous Random Variables

- Consider a finite or infinite sequence X_i, $i \in I$, of continuous random variables for the same experiment of sample space Ω.

Definition

The random variables of the sequence X_i, $i \in I$, are called **independent** if, for any sequence $[a_i, b_i]$, $i \in I$, of closed boxes of $\mathbb{R}$, the events

$$X_i \in [a_i, b_i], \ i \in I,$$

are independent.

- As a consequence of the definition of independence we have the following property.

Theorem

Given a sequence X_i, $i \in I$, of independent continuous random variables, for any positive integer k such that $2 \leq k \leq |I|$, for any $i_1, i_2, \ldots, i_k \in I$ distinct and for any Borel subset U of $\mathbb{R}^k$, we have

$$\mathbb{P}((X_{i_1}, \ldots, X_{i_k}) \in U) = \int\limits_{(x_{i_1}, \ldots, x_{i_k}) \in U} f_{X_{i_1}}(x_{i_1}) \cdots f_{X_{i_k}}(x_{i_k})\, dx_{i_1} \ldots dx_{i_k}.$$

Proof.

For any closed box $[a_{i_1}, b_{i_1}] \times \cdots \times [a_{i_k}, b_{i_k}]$ of $\mathbb{R}^k$, we have

$$\mathbb{P}\left((X_{i_1}, \ldots, X_{i_k}) \in [a_{i_1}, b_{i_1}] \times \cdots \times [a_{i_k}, b_{i_k}]\right)$$

$$= \mathbb{P}\left(X_{i_1} \in [a_{i_1}, b_{i_1}] \cap \cdots \cap X_{i_k} \in [a_{i_k}, b_{i_k}]\right)$$

$$= \mathbb{P}\left(X_{i_1} \in [a_{i_1}, b_{i_1}]\right) \cdots \mathbb{P}\left(X_{i_k} \in [a_{i_k}, b_{i_k}]\right)$$

$$= \left(\int_{x_{i_1} \in [a_{i_1}, b_{i_1}]} f_{X_{i_1}}(x_{i_1}) dx_{i_1}\right) \cdots \left(\int_{x_{i_k} \in [a_{i_k}, b_{i_k}]} f_{X_{i_k}}(x_{i_k}) dx_{i_k}\right)$$

$$= \int_{(x_{i_1}, \ldots, x_{i_k}) \in [a_{i_1}, b_{i_1}] \times \cdots \times [a_{i_k}, b_{i_k}]} f_{X_{i_1}}(x_{i_1}) \cdots f_{X_{i_k}}(x_{i_k})\, dx_{i_1} \ldots dx_{i_k}.$$

Proof.

Now, by starting from the formula

$$\mathbb{P}\left((X_{i_1},\ldots,X_{i_k}) \in [a_{i_1},b_{i_1}] \times \cdots \times [a_{i_k},b_{i_k}]\right)$$
$$= \int\limits_{\left(x_{i_1},\ldots,x_{i_k}\right)\in[a_{i_1},b_{i_1}]\times\cdots\times[a_{i_k},b_{i_k}]} f_{X_{i_1}}(x_{i_1})\cdots f_{X_{i_k}}(x_{i_k})\,dx_{i_1}\ldots dx_{i_k}.$$

for closed boxes, the formula

$$\mathbb{P}\left((X_{i_1},\ldots,X_{i_k}) \in U\right)$$
$$= \int\limits_{\left(x_{i_1},\ldots,x_{i_k}\right)\in U} f_{X_{i_1}}(x_{i_1})\cdots f_{X_{i_k}}(x_{i_k})\,dx_{i_1}\ldots dx_{i_k}.$$

for a general Borel subset U is then obtained by the same way with which Borel subsets are constructed from the closed boxes of $\mathbb{R}^k$, simply by applying the properties of a measure of probability. □

Then we have the following other property.

Theorem

Given a sequence X_i, $i \in I$, of independent continuous random variables, for any positive integer k such that $2 \leq k \leq |I|$, for any $i_1, i_2, \ldots, i_k \in I$ distinct and for any Borel subsets $U_{i_1}, \ldots, U_{i_k}$ of $\mathbb{R}$, we have

$$\mathbb{P}\left(X_{i_1} \in U_{i_1} \cap \cdots \cap X_{i_k} \in U_{i_k}\right) = \mathbb{P}\left(X_{i_1} \in U_{i_1}\right) \cdots \mathbb{P}\left(X_{i_k} \in U_{i_k}\right).$$

Exercise. Prove this by taking $U = U_{i_1} \times \cdots \times U_{i_k}$ in the previous formula.

This property can be rewritten as follows.

Theorem

Given a sequence X_i, $i \in I$, of independent continuous random variables, for any sequence U_i, $i \in I$, of Borel subsets of $\mathbb{R}$, the events of the sequence

$$X_i \in U_i, \ i \in I,$$

are independent.

Operations preserving independence

- Similarly to the discrete case, the following operations on a sequence X_i, $i \in I$, of continuous independent random variables preserve the independence relationship.

  First operation. Given a sequence X_i, $i \in I$, of independent continuous random variables, the random variables

  $$Y_i = f_i(X_i),\ i \in I,$$

  are independent, where $f_i : \mathbb{R} \to \mathbb{R}$.

  The proof is the same as in the discrete case.

- Second operation. Given the finite or infinite sequence $X_1, \ldots, X_n, X_{n+1}, X_{n+2}, \ldots$ of independent continuous random variables, the random variables

$$Y = f(X_1, \ldots, X_n), X_{n+1}, X_{n+2}, \ldots$$

are independent, where $f : \mathbb{R}^n \to \mathbb{R}$.

The proof is an adaption to integrals of the proof involving sums given in the discrete case.

- The functions $f_i : \mathbb{R} \to \mathbb{R}$ and the function $f : \mathbb{R}^n \to \mathbb{R}$ in the previous operations preserving independence needs to be Borel functions: a function $h : \mathbb{R}^k \to \mathbb{R}$ is said a **Borel function** if, for any Borel subset U of $\mathbb{R}$, the counter-image

$$f^{-1}(U) = \{x \in \mathbb{R}^k : f(x) \in U\}$$

is a Borel subset of $\mathbb{R}^k$.

Any piecewise continuous function is a Borel function. Indeed, any function encountered in theory and applications is a Borel function.

- Finally, we have the following two facts, whose proofs are exactly the same as in the discrete case.

  Third operation. Independent continuous random variables $X_1, X_2, X_3, \ldots$ remain independent if they are presented in any other order

  Fourth operation. A subsequence $X_i,\ i \in J \subseteq I$, of a sequence $X_i,\ i \in I$, of independent continuous random variables is a sequence of independent continuous random variables.

Mean of a Continuous Random Variable

- Here is the definition of mean for a continuous random variable. With respect to the case of a discrete random variable, we simply replace the sum with an integral.

Definition

Let X be a continuous random variable. The **mean** of X is the quantity

$$\mathbb{E}(X) := \int_{x \in \mathbb{R}} x f_X(x)\, dx = \int_{x \in \mathbb{R}} x \cdot \mathbb{P}(x < X \leq x + dx).$$

If X has distribution $\mathrm{U}(a, b)$, then

$$
\begin{aligned}
\mathbb{E}(X) &= \int_{x\in\mathbb{R}} x f_X(x)\,dx = \int_a^b x\frac{1}{b-a}dx \\
&= \frac{1}{b-a}\int_a^b x dx \\
&= \frac{1}{b-a}\cdot\frac{1}{2}\left(b^2-a^2\right) \\
&= \frac{a+b}{2} \text{ middle point of } [a, b].
\end{aligned}
$$

Similarly to the case of a discrete random variable, the mean of a continuous random variable X can be interpreted as center of mass of a distributed mass along a rod with linear density $f_X(x)$ at the point of abscissa x on the rod.

So, it is not a surprise that the mean of a random variable with uniform distribution $\mathrm{U}(a, b)$ is just in the middle of the interval $[a, b]$.

If X has distribution $\mathrm{N}\left(\mu, \sigma^2\right)$, then

$$\mathbb{E}(X) = \int\limits_{x \in \mathbb{R}} x f_X(x)\, dx = \int\limits_{x \in \mathbb{R}} (\mu + x - \mu) f_X(x)\, dx$$

$$= \mu \underbrace{\int\limits_{x \in \mathbb{R}} f_X(x)\, dx}_{=1} + \underbrace{\int\limits_{x \in \mathbb{R}} (x - \mu) f_X(x)\, dx}_{=0 \text{ since } f_X \text{ is symmetric around } \mu} = \mu.$$

Due to the symmetry of f_X around μ, the last integral

$$\int\limits_{x \in \mathbb{R}} (x - \mu) f_X(x)\, dx = \int\limits_{y \in \mathbb{R}} y f_X(\mu + y)\, dy, \quad y = x - \mu,$$

is zero, since it is the integral sum of positive infinitesimal terms

$$y f_X(\mu + y) dy, y > 0,$$

as well as opposite negative infinitesimal terms

$$y f_X(\mu + y) dy = -(-y) f_X(\mu + (-y)), y < 0.$$

- If Ω is continuous, the mean can be also expressed by

$$\mathbb{E}(X) = \int_{x \in \mathbb{R}} x f_X(x)\, dx = \int_{\omega \in \Omega} X(\omega)\, p(\omega)\, d\omega,$$

where $p : \Omega \to \mathbb{R}$ is the non-negative integrable function giving the probabilities of the closed boxes of Ω.

This form of the mean is a continuous analog of the form seen for discrete random variables in case of a discrete sample space. Its proof is an adaption to integrals of the proof involving sums given in the discrete case.

By using this form of the mean, we can prove, in case of Ω continuous, the linearity and the monotonicity of the mean.

Linearity of the mean: for $X, Y : \Omega \to \mathbb{R}$ continuous random variables,

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

and

$$\mathbb{E}(cX) = c\mathbb{E}(X) \text{ for any } c \in \mathbb{R}.$$

As a consequence, we have

$$\mathbb{E}(X_1 + X_2 + \cdots + X_n) = \mathbb{E}(X_1) + \cdots + \mathbb{E}(X_n)$$

for an arbitrary number of continuous random variable $X_1, X_2, \ldots, X_n : \Omega \to \mathbb{R}$.

Monotonicity of the mean: for $X, Y : \Omega \to \mathbb{R}$ continuous random variables,

$$X \leq Y \Rightarrow \mathbb{E}(X) \leq \mathbb{E}(Y).$$

We have the **rule of multiplication of the means in case of independence**: for $X_1, X_2, \ldots, X_n : \Omega \to \mathbb{R}$ independent continuous random variables,

$$\mathbb{E}\left(X_1 X_2 \cdots X_n\right) = \mathbb{E}\left(X_1\right) \mathbb{E}\left(X_2\right) \cdots \mathbb{E}\left(X_n\right).$$

The proof of this property for $n = 2$ is an adaption to integrals of the proof involving sums of the analogous result in the discrete case.

Another useful formula is

$$\mathbb{E}\left(h\left(X\right)\right) = \int_{x \in \mathbb{R}} h\left(x\right) f_X\left(x\right) dx$$

for a continuous random variable X and $h : \mathbb{R} \to \mathbb{R}$ Borel function.

The proof of this formula is once again an adaption to integrals of the proof of the discrete analog.

Exercise. Let X be a continuous random variable such that $X(\Omega) \subseteq I$, I interval of $\mathbb{R}$. By using the rule of transformation for pdfs, prove the formula

$$\mathbb{E}\left(h\left(X\right)\right) = \int\limits_{x \in I} h\left(x\right) f_X\left(x\right) dx$$

in case of a function $h : I \to \mathbb{R}$ differentiable and strictly increasing.

Exercise. Prove that

$$\mathbb{E}(X + c) = \mathbb{E}(X) + c \text{ for any } c \in \mathbb{R},$$

for a continuous random variable X.

Variance

- Variance and standard deviation for continuous random variables are defined exactly as in case of discrete random variables.

  The **variance** of X is

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

  and the **standard deviation** of X is

  $$\mathrm{SD}\,(X) := \sqrt{\mathrm{Var}\,(X)}.$$

  As in the discrete case, we have

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

  and we have that $\mathrm{Var}\,(X)$ can intepreted as momentum of inertia of a rod with distributed mass of linear density f_X around the axis passing through the center of mass.

Moreover, we have the **Chebyshev's inequality**, that is proved exactly as in the discrete case : for a continuous random variable X with mean μ, we have

$$\mathbb{P}\left(|X-\mu| \geq c\right) \leq \frac{\mathrm{Var}\left(X\right)}{c^2},\ c>0,$$

or the equivalent form

$$\mathbb{P}\left(|X-\mu| < k\mathrm{SD}\left(X\right)\right) \geq 1-\frac{1}{k^2},\ k>0.$$

Finally, with proofs exactly as in the discrete case, we have the properties of the variance : for $X, Y : \Omega \to \mathbb{R}$ continuous random variables,

$$\mathrm{Var}\,(cX) = c^2 \mathrm{Var}\,(X) \quad \text{for any } c \in \mathbb{R}$$

and

$$\mathrm{Var}\,(X + c) = \mathrm{Var}\,(X) \quad \text{for any } c \in \mathbb{R}.$$

Moreover, we have the **rule of addition of the variances in case of independence**: for $X_1, X_2, \ldots, X_n : \Omega \to \mathbb{R}$ independent continuous random variables,

$$\mathrm{Var}\,(X_1 + X_2 + \cdots + X_n) = \mathrm{Var}\,(X_1) + \mathrm{Var}\,(X_2) + \cdots + \mathrm{Var}\,(X_n)\,.$$

- Now, we determine the variance of a normal random variable.

  If Z has distribution $\mathrm{N}(0,1)$, then

$$
\begin{aligned}
\mathbb{E}\left(Z^2\right) &= \int\limits_{z\in\mathbb{R}} z^2 f_Z(z)\,dz = \int\limits_{z\in\mathbb{R}} z^2 \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}\,dz \\
&= -\int\limits_{z\in\mathbb{R}} \frac{1}{\sqrt{2\pi}} zD\left(e^{-\frac{z^2}{2}}\right)dz \\
&= -\left(\underbrace{\left[\frac{1}{\sqrt{2\pi}} ze^{-\frac{z^2}{2}}\right]_{-\infty}^{+\infty}}_{=0-0=0} - \underbrace{\int\limits_{z\in\mathbb{R}} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}\,dz}_{=1}\right) \\
&= 1
\end{aligned}
$$

  and so

$$
\mathrm{Var}(Z) = \mathbb{E}\left(Z^2\right) - \mathbb{E}(Z)^2 = 1 - 0^2 = 1.
$$

If X has distribution $\mathrm{N}(\mu, \sigma^2)$, then

$$X = \mu + \sigma Z,$$

where Z has distribution $\mathrm{N}(0, 1)$, and then

$$\begin{aligned}
\mathrm{Var}(X) &= \mathrm{Var}(\mu + \sigma Z) = \mathrm{Var}(\sigma Z) = \sigma^2 \mathrm{Var}(Z) = \sigma^2 \\
\mathrm{SD}(X) &= \sqrt{\mathrm{Var}(X)} = \sqrt{\sigma^2} = \sigma.
\end{aligned}$$

So, the parameters μ and σ of the normal distribution $N(\mu, \sigma^2)$ are, respectively, mean and standard deviation of a random variable with that distribution.

Note that, whereas Chebyshev's inequality says that

$$\mathbb{P}(|X-\mu| < 2\sigma) \geq \frac{3}{4} = 75\% \text{ and } \mathbb{P}(|X-\mu| < 3\sigma) \geq \frac{8}{9} = 88.9\%,$$

we actually have

$$\mathbb{P}(|X-\mu| < 2\sigma) = 95\% \text{ and } \mathbb{P}(|X-\mu| < 3\sigma) = 99.7\%.$$

- Exercise. Find the variance and the standard deviation of a continuous random variable U with distribution $\mathrm{U}(0, 1)$. Then, compute the probabilities

$$\mathbb{P}\left(|U - \mu| < 2\mathrm{SD}(U)\right) \text{ and } \mathbb{P}\left(|U - \mu| < 3\mathrm{SD}(U)\right),$$

where $\mu = \mathbb{E}(U)$, and compare them with the lower bounds given by the Chebyshev's inequality.

Exercise. Let X be a continuous random variable with distribution $\mathrm{U}(a, b)$. Show that

$$U = \frac{X - a}{b - a}$$

has distribution $\mathrm{U}(0, 1)$. Find the variance and the standard deviation of X.

Finding probabilities for Standard Normal Random Variables

- Let Z be a standard normal random variable. We denote by Φ its distribution function:

$$\Phi(x) := F_Z(x) = \mathbb{P}(Z \leq x) = \mathbb{P}(Z < x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz, \; x \in \mathbb{R}.$$

In the next table, we see the values $\Phi(x)$ for $x \geq 0$.

Table 6.1 Standard Normal Probabilities

x	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998

How to use this table?

For example, suppose that we want to compute $\Phi(x)$ for $x = 1.22$.

We split x as

$$x = 1.22 = \underbrace{1.2}_{\text{row}} + \underbrace{0.02}_{\text{column}}$$

and we use the entry of the table with row 1.2 and column 0.02.

x	**0.00**	**0.01**	**0.02**	**0.03**	**0.04**	...	**0.09**
0.0	0.5000	0.5040					
⋮							
1.1	0.8413						
1.2	0.8849	0.8869	0.8888				
1.3	0.9032						

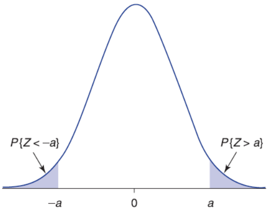
We obtain $\Phi(1.22) = 0.8888$.

- Now observe that, for $x \geq 0$,

$$\mathbb{P}(Z \geq x) = \mathbb{P}(Z > x) = 1 - \mathbb{P}(Z \leq x) = 1 - \Phi(x).$$

So, for example,

$$\mathbb{P}(Z > 1.52) = 1 - \Phi(1.52) = 1 - 0.9357 = 0.0643.$$



- Moreover,

$$\Phi\,(x) = 1 - \Phi\,(-x)\,,\ x < 0.$$

In fact, due to the symmetry of the pdf f_Z of Z, we have, for $x < 0$,

$$\Phi\,(x) = \mathbb{P}\,(Z \leq x) = \mathbb{P}\,(Z \geq -x) = 1 - \Phi\,(-x)\,.$$

So, for example,

$$\Phi(-0.14) = 1 - \Phi(0.14) = 1 - 0.5557 = 0.4443.$$

- Finally, for $a, b \in \mathbb{R}$ with $a < b$, we can compute the probability

$$\mathbb{P}(a \le Z \le b) = \mathbb{P}(a < Z \le b) = \mathbb{P}(a \le Z < b) = \mathbb{P}(a < Z < b)$$

as

$$\mathbb{P}(a < Z \le b) = \mathbb{P}(Z \le b) - \mathbb{P}(Z \le a) = \Phi(b) - \Phi(a)\,.$$

Examples:

$$\mathbb{P}(0.5 \le Z \le 1.48) = \Phi(1.48) - \Phi(0.5) = 0.9306 - 0.6915 = 0.2391.$$

$$\mathbb{P}(-1.12 \le Z \le 0.73) = \Phi(0.73) - \Phi(-1.12) = \Phi(0.73) - (1 - \Phi(1.12))$$
$$= \Phi(0.73) + \Phi(1.12) - 1 = 0.7673 + 0.8686 - 1 = 0.6359.$$

$$\mathbb{P}(-2.38 \le Z \le -1.94) = \Phi(-1.94) - \Phi(-2.38)$$
$$= 1 - \Phi(1.94) - (1 - \Phi(2.38)) = \Phi(2.38) - \Phi(1.94) = 0.9913 - 0.9738$$
$$= 0.0175.$$

- Exercise. Prove that, for $c > 0$, we have

$$\mathbb{P}\left(-c \leq Z \leq c\right) = 2\Phi\left(c\right) - 1$$

and

$$\mathbb{P}\left(|Z| \geq c\right) = 2\left(1 - \Phi\left(c\right)\right).$$

Finding probabilities for general Normal Random Variables

- Now, let X be a normal random variable with distribution $\mathrm{N}\left(\mu, \sigma^2\right)$.

For $a, b \in \mathbb{R}$ with $a < b$, the probabilities

$$\mathbb{P}\left(X \leq a\right) = \mathbb{P}\left(X < a\right)$$
$$\mathbb{P}\left(X \geq b\right) = \mathbb{P}\left(X > b\right)$$
$$\mathbb{P}\left(a \leq X \leq b\right) = \mathbb{P}\left(a < X \leq b\right) = \mathbb{P}\left(a \leq X < b\right) = \mathbb{P}\left(a < X < b\right)$$

can be determined by reducing them to probabilities relevant to the standardized form

$$Z = \frac{X - \mu}{\sigma}$$

of X, which is a standard normal variable.

We have

$$\mathbb{P}(X \le a) = \mathbb{P}\left(Z = \frac{X-\mu}{\sigma} \le \frac{a-\mu}{\sigma}\right) = \Phi\left(\frac{a-\mu}{\sigma}\right)$$

and

$$\mathbb{P}(X \ge b) = \mathbb{P}\left(Z = \frac{X-\mu}{\sigma} \ge \frac{b-\mu}{\sigma}\right) = 1 - \Phi\left(\frac{b-\mu}{\sigma}\right)$$

and

$$\begin{aligned}\mathbb{P}(a \le X \le b) &= \mathbb{P}\left(\frac{a-\mu}{\sigma} \le Z = \frac{X-\mu}{\sigma} \le \frac{b-\mu}{\sigma}\right) \\ &= \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right).\end{aligned}$$

Example. The IQ score for a sixth-grader (a student after six years of school) is normally distributed with mean $\mu = 100$ *and standard deviation* $\sigma = 14.2$*.*

- *What is the probability that a sixth-grader has a score greater than* 130*?*
- *What is the probability that a sixth-grader has a score between* 90 *and* 110*?*

Indeed, the IQ score is a discrete random variable, since the possible scores are finite. But here, we approximate it by a continuous normal random variable by imagining that the possible scores are all the real numbers in an interval.

The experiment relevant to this situation is the formation process of a given child from her/his birth to the sixth year of school.

Let X be the random variable IQ score for a sixth grader. X has distribution $\mathrm{N}\left(100, 14.2^2\right)$*.*

We have

$$\begin{aligned}\mathbb{P}(X \geq 130) &= \mathbb{P}\left(Z \geq \frac{130-100}{14.2} = 2.11\right) = 1 - \Phi(2.11) \\ &= 1 - 0.9826 = 1.74\%\end{aligned}$$

and

$$\begin{aligned}\mathbb{P}(90 \leq X \leq 110) &= \mathbb{P}\left(\frac{90-100}{14.2} \leq Z \leq \frac{110-100}{14.2}\right) \\ &= \mathbb{P}(-0.70 \leq Z \leq 0.70) \\ &= 2\Phi(0.70) - 1 = 2 \cdot 0.7580 - 1 = 51.6\%.\end{aligned}$$

- Let X be a normal random variable with distribution $\mathrm{N}\left(\mu, \sigma^2\right)$.

  For $k > 0$, we have

$$
\begin{aligned}
\mathbb{P}\left(|X-\mu| \leq k\sigma\right) &= \mathbb{P}\left(\mu - k\sigma \leq X \leq \mu + k\sigma\right) \\
&= \mathbb{P}\left(-k \leq Z = \frac{X-\mu}{\sigma} \leq k\right) \\
&= 2\Phi\left(k\right) - 1.
\end{aligned}
$$

  In particular:

$$
\begin{aligned}
&\mathbb{P}\left(|X-\mu| \leq \sigma\right) = 2\Phi\left(1\right) - 1 = 2 \cdot 0.8413 - 1 = 68.26\% \\
&\mathbb{P}\left(|X-\mu| \leq 2\sigma\right) = 2\Phi\left(2\right) - 1 = 2 \cdot 0.9772 - 1 = 95.44\% \\
&\mathbb{P}\left(|X-\mu| \leq 3\sigma\right) = 2\Phi\left(3\right) - 1 = 2 \cdot 0.9987 - 1 = 99.74\%
\end{aligned}
$$

  as we have already seen.

- In MATLAB, the values of the (cumulative) distribution function of a normal random variable are computed by the function $\mathrm{normcdf}$:

$$\mathrm{normcdf(x)}$$

computes $\Phi(x) = F_Z(x)$, where Z is a standard normal variable, and

$$\mathrm{normcdf}(\mathrm{x}, \mu, \sigma)$$

computes $F_X(x)$, where X has distribution $\mathrm{N}(\mu, \sigma^2)$.

Exercise. The previous table of values $\Phi(x)$ is up to 3.49. By using MATLAB compute $\Phi(x)$ for $x = 4, 5, 6, \ldots, 10$.

- Exercise. The time that a given runner will run the 100m men race at the Olympic Games is normally distributed with mean 9.70 s and standard deviation 0.06 s. What is the probability that he will obtain the world record for this race?

  Exercise. The height of a male individual of the italian population is a random normal variable with mean 176 cm and standard deviation 7 cm. Estimate the number of italian males whose height is more than 2 m.

  Exercise. Assume that a given shot put male athlete throws the shot at a distance normally distributed with mean 19.9 m and standard deviation 45 cm. What is the probability that he will overcome 21 m in at least one of three independent throws.

Exercise. A worker has two possible paths A and B for reaching her/his work place by car. She/he leaves home at $8:00$ and has to be at work at $8:30$. The travel time for the path A (with heavy traffic) is a normal random variable with distribution $\mathrm{N}(\mu_A, \sigma_A^2)$ with $\mu_A = 23\ \mathrm{min}$ and $\sigma_A = 3.5\ \mathrm{min}$. On the other hand, the travel time for the path B (with much less traffic) is a normal random variable with distribution $\mathrm{N}(\mu_B, \sigma_B^2)$ with $\mu_B = 27\ \mathrm{min}$ and $\sigma_B = 1.4\ \mathrm{min}$. What is the best path?

Exercise. It is advisable to change the timing belt of a given car after an usage of $120000\ \mathrm{km}$ (the warranty time). Assume that the lifespan of such a timing belt is a normal random variable with distribution $\mathrm{N}(\mu, \sigma^2)$, where $\sigma = 10\%\mu$ and $\mu - 3\sigma = 120000\ \mathrm{km}$. For such a car, compute the probability that the timing belt breaks before an usage of $120000\ \mathrm{km}$. Moreover, assume that the timing belt of such a car is still working after an usage of $135000\ \mathrm{km}$. What is the probability that the timing belt does not break in the next $1000\ \mathrm{km}$?

Exercise. The lifespan of a given object before it breaks for the usage is a normal random variable with distribution $\mathrm{N}(\mu, \sigma^2)$. Assume that after a time T the object is still working. Given $a > 0$, find the limit, as $T \to +\infty$, of the probability that the object is still working at the time $T + a$,

Properties of Normal Random Variables

- Let X be a normal random variable with mean μ and standard deviation σ.

  For any $c \in \mathbb{R}$, $Y = X + c$ is still a normal random variable: Y has distribution $\mathrm{N}(\mu + c, \sigma^2)$.

  In fact,

$$X = Y - c = \psi(Y)$$

  and so for the rule of transformation for pdfs:

$$\begin{aligned} f_Y(y) &= f_X(\psi(y))\,\psi'(y) = f_X(y-c)\cdot 1 \\ &= \frac{1}{\sigma\sqrt{2\pi}}\cdot e^{-\frac{(y-c-\mu)^2}{2\sigma^2}} = \frac{1}{\sigma\sqrt{2\pi}}\cdot e^{-\frac{(y-(\mu+c))^2}{2\sigma^2}},\ y \in \mathbb{R}. \end{aligned}$$

Moreover, for any $c \in \mathbb{R} \setminus \{0\}$, also $Y = cX$ is still a normal random variable: Y has distribution $\mathrm{N}\left(c\mu, (|c|\,\sigma)^2\right)$.

In fact,

$$X = \frac{1}{c} Y = \psi(Y)$$

and so for the rule of transformation for pdfs:

$$\begin{aligned} f_Y(y) &= \operatorname{sign}(\psi') \, f_X(\psi(y)) \, \psi'(y) = \operatorname{sign}\left(\frac{1}{c}\right) f_X\left(\frac{1}{c} y\right) \cdot \frac{1}{c} \\ &= \frac{1}{|c|\,\sigma\sqrt{2\pi}} \cdot e^{-\frac{\left(\frac{1}{c} y - \mu\right)^2}{2\sigma^2}} = \frac{1}{|c|\,\sigma\sqrt{2\pi}} \cdot e^{-\frac{(y - c\mu)^2}{2(|c|\sigma)^2}}, \; y \in \mathbb{R}. \end{aligned}$$

Exercise. Let X be a normal random variable with mean μ and standard deviation σ. Find the distribution of

$$Y = aX + b,$$

where $a, b \in \mathbb{R}$ with $a \neq 0$.

Sum of independent normal random variables

- Another important property concerns the sum

$$S = X_1 + X_2 + \cdots + X_n$$

of n independent normal random variables $X_1, X_2, \ldots, X_n : \Omega \to \mathbb{R}$.

If X_i, $i \in \{1, 2, \ldots, n\}$, has distribution $\mathrm{N}(\mu_i, \sigma_i^2)$, then

$$\mathbb{E}(S) = \mathbb{E}(X_1) + \mathbb{E}(X_2) + \cdots + \mathbb{E}(X_n) = \mu_1 + \mu_2 + \cdots + \mu_n$$

and

$$\mathrm{Var}(S) = \mathrm{Var}(X_1) + \mathrm{Var}(X_2) + \cdots + \mathrm{Var}(X_n) = \sigma_1^2 + \sigma_2^2 + \cdots + \sigma_n^2.$$

But, what is the distribution of S?

To answer this, we have to introduce the following notion.

Definition

Let X be a discrete or continuous random variable. The **moment generating function** of X is the function $M_X : (-\gamma(X), \gamma(X)) \to \mathbb{R}$ given by

$$M_X(\alpha) = \mathbb{E}\left(e^{\alpha X}\right), \; \alpha \in (-\gamma(X), \gamma(X)),$$

Observe that

$$\mathbb{E}(e^{\alpha X}) = \begin{cases} \sum_{x \in X(\Omega)} \underbrace{e^{\alpha x}\mathbb{P}(X = x)}_{\geq 0} & \text{if } X \text{ is discrete} \\ \int_{x \in \mathbb{R}} \underbrace{e^{\alpha x} f_X(x)}_{\geq 0} & \text{if } X \text{ is continuous} \end{cases}$$

and so $0 \leq \mathbb{E}(e^{\alpha X}) < +\infty$ or $\mathbb{E}(e^{\alpha X}) = +\infty$.

$\gamma(X) \in [0, +\infty]$, which depends on X, is the largest number γ in $[0, +\infty]$ such that $\mathbb{E}\left(e^{\alpha X}\right) < +\infty$ for $\alpha \in (-\gamma, \gamma)$. When $\gamma(X) = +\infty$, the domain of M_X is the whole $\mathbb{R}$.

The name "moment generating function" comes from the fact that

$$M_X(\alpha) = \mathbb{E}(e^{\alpha X}) = \mathbb{E}(1 + \alpha X + \frac{1}{2}\alpha^2 X^2 + \frac{1}{3!}\alpha^3 X^3 + \cdots)$$
$$= 1 + \alpha\mathbb{E}(X) + \frac{1}{2}\alpha^2\mathbb{E}(X^2) + \frac{1}{3!}\alpha^3\mathbb{E}(X^3) + \cdots,$$

where α is in a neighborhood of 0, and so the **moments** $\mathbb{E}(X^r)$, $r \in \{1, 2, \ldots\}$, of X are given by $\mathbb{E}(X^r) = M_X^{(r)}(0)$.

- Let Z be a standard normal random variable. We determine its moment generating function.

  For $\alpha \in \mathbb{R}$, we have

$$M_Z(\alpha) = \mathbb{E}\left(e^{\alpha Z}\right) = \int\limits_{z\in\mathbb{R}} e^{\alpha z} f_Z(z)\, dz = \int\limits_{z\in\mathbb{R}} e^{\alpha z} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}\, dz$$

$$= \int\limits_{z\in\mathbb{R}} \frac{1}{\sqrt{2\pi}} e^{\alpha z - \frac{z^2}{2}}\, dz = \int\limits_{z\in\mathbb{R}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left((z-\alpha)^2 - \alpha^2\right)} dz$$

$$= e^{\frac{1}{2}\alpha^2} \int\limits_{z\in\mathbb{R}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(z-\alpha)^2} dz = e^{\frac{1}{2}\alpha^2}$$

  since

$$\int\limits_{z\in\mathbb{R}} \underbrace{\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(z-\alpha)^2}}_{=f_Y(z)} dz = 1$$

  where Y has distribution $\mathrm{N}(\alpha, 1)$. We have $\gamma(Z) = +\infty$.

Now, let X be a normal random variable with distribution $\mathrm{N}(\mu, \sigma^2)$.

We have

$$X = \mu + \sigma Z,$$

where Z is a standard normal distribution.

Thus

$$\begin{aligned} M_X(\alpha) &= \mathbb{E}\left(e^{\alpha X}\right) = \mathbb{E}\left(e^{\alpha(\mu+\sigma Z)}\right) \\ &= \mathbb{E}\left(e^{\mu\alpha+\sigma\alpha Z}\right) = \mathbb{E}\left(e^{\mu\alpha} e^{\sigma\alpha Z}\right) \\ &= e^{\mu\alpha}\mathbb{E}\left(e^{\sigma\alpha Z}\right) = e^{\mu\alpha} M_Z(\sigma\alpha) \\ &= e^{\mu\alpha} e^{\frac{1}{2}\sigma^2\alpha^2} = e^{\mu\alpha+\frac{1}{2}\sigma^2\alpha^2}, \ \alpha \in \mathbb{R}. \end{aligned}$$

We have $\gamma(X) = +\infty$.

Exercise. Determine the first four moments of a standard normal random variable by using its moment generating function. Then determine the first four moments of a general normal random variable.

- Exercise. Let U be a random variable with distribution $\mathrm{U}(0,1)$. Determine the moment generating function of U. Determine the first four moments of U by using its moment generating function. Then determine the moment generating function and the first four moments of a random variable X with distribution $\mathrm{U}(a,b)$.

- The following fact holds: the moment generating function determines the distribution function: let X and Y be random variables, if there exists γ with $0 < \gamma < \min\{\gamma(X), \gamma(Y)\}$ such that

$$M_X(\alpha) = M_Y(\alpha),\ \alpha \in (-\gamma, \gamma),$$

then

$$F_X = F_Y.$$

- Next theorem says how to determine the moment generating function of a sum of independent random variables.

Theorem

Let $X_1, X_2, \ldots, X_n : \Omega \to \mathbb{R}$ be discrete or continuous random variables. If $X_1, X_2, \ldots, X_n$ are independent, then

$$M_{X_1+X_2+\cdots+X_n}(\alpha) = M_{X_1}(\alpha)\, M_{X_2}(\alpha) \cdots M_{X_n}(\alpha)\,,\ \alpha \in (-\gamma, \gamma)\,,$$

where $\gamma = \min_{i\in\{1,\ldots,n\}} \gamma(X_i)$.

Proof.

Let $\alpha \in (-\gamma, \gamma)$. If $X_1, X_2, \ldots, X_n$ are independent, then $e^{\alpha X_1}, e^{\alpha X_2}, \ldots, e^{\alpha X_n}$ are independent and so

$$\begin{aligned} M_{X_1+X_2+\cdots+X_n}(\alpha) &= \mathbb{E}\left(e^{\alpha(X_1+X_2+\cdots+X_n)}\right) = \mathbb{E}\left(e^{\alpha X_1} e^{\alpha X_2} \cdots e^{\alpha X_n}\right) \\ &= \mathbb{E}\left(e^{\alpha X_1}\right) \mathbb{E}\left(e^{\alpha X_2}\right) \cdots \mathbb{E}\left(e^{\alpha X_n}\right) \\ &= M_{X_1}(\alpha)\, M_{X_2}(\alpha) \cdots M_{X_n}(\alpha)\,. \end{aligned}$$

- Next theorem answers to the question about the distribution of a sum of normal random variables.

Theorem

Let $X_1, X_2, \dots, X_n : \Omega \to \mathbb{R}$ be normal random variables, where X_i has distribution $\mathrm{N}\left(\mu_i, \sigma_i^2\right)$, $i \in \{1, \cdots, n\}$. If $X_1, X_2, \dots, X_n$ are independent, then

$$S = X_1 + X_2 + \cdots + X_n$$

is a normal random variable with distribution

$$\mathrm{N}\left(\mu_1 + \mu_2 + \cdots + \mu_n, \sigma_1^2 + \sigma_2^2 + \cdots + \sigma_n^2\right).$$

Proof.

We have, for $i \in \{1, 2, \ldots, n\}$,

$$M_{X_i}(\alpha) = e^{\mu_i \alpha + \frac{1}{2}\sigma_i^2 \alpha^2}, \ \alpha \in \mathbb{R}.$$

Since $X_1, X_2, \ldots, X_n$ are independent, we have

$$\begin{aligned} M_S(\alpha) &= M_{X_1}(\alpha) M_{X_2}(\alpha) \cdots M_{X_n}(\alpha) \\ &= e^{\mu_1 \alpha + \frac{1}{2}\sigma_1^2 \alpha^2} e^{\mu_2 \alpha + \frac{1}{2}\sigma_2^2 \alpha^2} \ldots e^{\mu_n \alpha + \frac{1}{2}\sigma^2 \alpha_n^2} \\ &= e^{(\mu_1 + \mu_2 + \cdots + \mu_n)\alpha + \frac{1}{2}\left(\sigma_1^2 + \sigma_2^2 + \cdots + \sigma_n^2\right)\alpha^2}, \ \alpha \in \mathbb{R}. \end{aligned}$$

This is the moment generating function of a random variable of distribution $\mathrm{N}\left(\mu_1 + \mu_2 + \cdots + \mu_n, \sigma_1^2 + \sigma_2^2 + \cdots + \sigma_n^2\right)$. Since the moment generating function determines the distribution, we can conclude that S has this normal distribution. □

Example. Suppose that the lifespan of a light bulb before burning out is a normal random variable with mean $\mu = 400\ h$ *and standard deviation* $\sigma = 40\ h$.

An individual purchases n such light bulbs and use them one after the other: when a bulb burns out, it is replaced with another bulb. What is the probability that the total lifespan of the bulbs will exceed $k\mu$ *with* $k > 0$*?*

Let X_i, $i \in \{1, 2, \dots, n\}$, *be the lifespan of the* $i-$*th light bulb. It is reasonable to assume that* $X_1, X_2, \dots, X_n$ *are independent. Then, the total lifespan* $S = X_1 + X_2 + \cdots + X_n$ *of the bulbs has distribution*

$$\mathrm{N}\left(n\mu, n\sigma^2\right) = \mathrm{N}\left(n\mu, (\sqrt{n}\sigma)^2\right).$$

Thus

$$\mathbb{P}(S \geq k\mu) = \mathbb{P}\left(Z = \frac{S - n\mu}{\sqrt{n}\sigma} \geq \frac{k\mu - n\mu}{\sqrt{n}\sigma}\right)$$
$$= \mathbb{P}\left(Z \geq -\frac{\mu}{\sigma} \cdot \frac{n-k}{\sqrt{n}}\right) = 1 - \mathbb{P}\left(Z \leq -\frac{\mu}{\sigma} \cdot \frac{n-k}{\sqrt{n}}\right)$$
$$= 1 - \Phi\left(-\frac{\mu}{\sigma} \cdot \frac{n-k}{\sqrt{n}}\right) = \Phi\left(\frac{\mu}{\sigma} \cdot \frac{n-k}{\sqrt{n}}\right).$$

Now, suppose that the individual wants $S \geq k\mu$, where k is given integer, because the she/he plans to go back to the light bulbs shop not before a time $k\mu$. How many bulbs does the individual need to purchase in order to be quite sure of this?

By purchasing $n = k$ light bulbs, we have

$$\mathbb{P}(S \geq k\mu) = \Phi(0) = \frac{1}{2}.$$

By purchasing $n = k + 1$ *light bulbs, we have*

$$\mathbb{P}(S \geq k\mu) = \Phi\left(\frac{\mu}{\sigma} \cdot \frac{1}{\sqrt{k+1}}\right).$$

and if

$$\frac{\mu}{\sigma} \cdot \frac{1}{\sqrt{k+1}} \geq 3.49,$$

where 3.49 *is the largest* x *in the table of the values* $\Phi(x)$*, i.e.*

$$k \leq \left(\frac{\frac{\mu}{\sigma}}{3.49}\right)^2 - 1 = \left(\frac{10}{3.49}\right)^2 - 1 = 7.21,$$

then

$$\mathbb{P}(S \geq k\mu) \geq \Phi(3.49) = 99.98\%.$$

So, for $k \leq 7$*, to be almost sure that the total lifespan of the bulbs will exceed* $k\mu$*, it is sufficient to purchase* $k + 1$ *bulbs.*

Example. In the place A, the yearly rainfall is normally distributed with mean $\mu_A = 998$ mm *and standard deviation* $\sigma_A = 156$ mm. *In the place B, very very far from A, the yearly rainfall is normally distributed with mean* $\mu_B = 1212$ mm *and standard deviation* $\sigma_B = 180$ mm.

What is the probability that A will have more rainfall than B in the next year?

Let X and Y be the yearly rainfall at A and B with distribution $\mathrm{N}\left(\mu_A, \sigma_A^2\right)$ *and* $\mathrm{N}\left(\mu_B, \sigma_B^2\right)$, *respectively.*

We look for the probability

$$\mathbb{P}(X > Y) = \mathbb{P}(X - Y > 0).$$

Since A is very very far from B, it is reasonable to assume that X and Y are independent.

So X and $-Y$ are independent, where $-Y$ has distribution $\mathrm{N}\left(-\mu_B, \sigma_B^2\right)$.

Then, $X - Y$ has distribution

$$\mathrm{N}\left(\mu_A - \mu_B, \sigma_A^2 + \sigma_B^2\right) = \mathrm{N}\left(\mu_A - \mu_B, \left(\sqrt{\sigma_A^2 + \sigma_B^2}\right)^2\right).$$

Therefore

$$\mathbb{P}\left(X - Y > 0\right)$$

$$= \mathbb{P}\left(Z = \frac{X - Y - (\mu_A - \mu_B)}{\sqrt{\sigma_A^2 + \sigma_B^2}} > \frac{0 - (\mu_A - \mu_B)}{\sqrt{\sigma_A^2 + \sigma_B^2}} = \frac{\mu_B - \mu_A}{\sqrt{\sigma_A^2 + \sigma_B^2}}\right)$$

$$= \mathbb{P}\left(Z > 0.90\right) = 1 - \mathbb{P}\left(Z \leq 0.90\right) = 1 - \Phi\left(0.90\right) = 18.4\%.$$

Exercise. A tradesman possesses three supermarkets, denoted by A, B and C. During a saturday, the takings of the supermarkets are normal random variables:

- *the taking of A has mean* $\mu_A = 70K$ Euro *with standard deviation* $\sigma_A = 8K$ Euro*.*
- *the taking of B has mean* $\mu_B = 40K$ Euro *with standard deviation* $\sigma_B = 4.5K$ Euro*.*
- *the taking of C has mean* $\mu_C = 35K$ Euro *with standard deviation* $\sigma_C = 2.8K$ Euro*.*

Assume that the three takings are independent. What is the probability that, during a saturday, the total taking of the three supermarkets is over $160K$ Euro*? What is the probability that in the total taking the contribution of A is at least* 60%*?*

- Exercise. Let X and Y be independent random variables with distribution $\mathrm{U}(0,1)$. Determine the moment generating function of $X+Y$. Has $X+Y$ a uniform distribution $\mathrm{U}(a,b)$, for some $a,b\in\mathbb{R}$ with $a<b$?

The *k*-Sigma rule and the *k*-Sigma methodology

- Consider an industrial process producing pieces.

  Let X be a numerical quantity related to a produced piece. For example, if the piece is a disk, X can be the diameter of the piece.

  We can assume that, in the experiment of the production of a piece, X is a random variable with normal distribution $\mathrm{N}(\mu, \sigma^2)$, where μ is the required value for the quantity X.

  We consider the produced piece as defective if $|X - \mu| > \mathrm{TOL}$, where TOL is a given tolerance.

  We say that the industrial process adopt the ***k*-Sigma rule**, where $k > 0$, if the standard deviation σ is such that $k\sigma \leq \mathrm{TOL}$.

  As a consequence, the probability that the produced piece will be defective is, with Z the standardized form of X,

$$\begin{aligned} p &= \mathbb{P}(|X - \mu| > \mathrm{TOL}) \\ &\leq \mathbb{P}(|X - \mu| > k\sigma) = \mathbb{P}(|Z| > k) = 2\,(1 - \Phi(k)) \,. \end{aligned}$$

So, for $k = 3$ (Three-Sigma rule), we have

$$p \leq 2(1 - \Phi(3)) = 2.7 \cdot 10^{-3}$$

and, for $k = 6$ (Six-Sigma rule), we have

$$p \leq 2(1 - \Phi(6)) = 2.0 \cdot 10^{-9}.$$

- We have assumed that *X* has distribution $\mathrm{N}(\mu, \sigma^2)$ but this holds only in a short period.

  In a long period, we can assume that *X* has distribution $\mathrm{N}(\mu + \Delta, \sigma^2)$, where the constant Δ is some systematic error introduced in the production process.

  The difference between the *k*-Sigma rule and the ***k*-Sigma methodology** is that in the *k*-Sigma methodology it is assumed that *X* has distribution $\mathrm{N}(\mu + \Delta, \sigma^2)$ with

$$\frac{|\Delta|}{\sigma} \leq 1.5$$

  We write

$$X = \mu + \Delta + E,$$

  where *E* has distribution $\mathrm{N}(0, \sigma^2)$.

In the k-Sigma methodology, the probability p that the piece is defective satisfies

$$\begin{aligned} p &= \mathbb{P}\left(|X-\mu| > \mathrm{TOL}\right) \\ &\leq \mathbb{P}\left(|X-\mu| > k\sigma\right) = 1 - \mathbb{P}\left(|X-\mu| \leq k\sigma\right) \\ &= 1 - \mathbb{P}\left(|\Delta + E| \leq k\sigma\right) \end{aligned}$$

with

$$\begin{aligned} &\mathbb{P}\left(|\Delta + E| \leq k\sigma\right) = \mathbb{P}\left(-k\sigma \leq \Delta + E \leq k\sigma\right) \\ &= \mathbb{P}\left(-k\sigma - \Delta \leq E \leq k\sigma - \Delta\right) = \mathbb{P}\left(-k - \frac{\Delta}{\sigma} \leq \frac{E}{\sigma} \leq k - \frac{\Delta}{\sigma}\right) \\ &= \Phi\left(k - \frac{\Delta}{\sigma}\right) - \Phi\left(-k - \frac{\Delta}{\sigma}\right) = \Phi\left(k - \frac{\Delta}{\sigma}\right) - \left(1 - \Phi\left(k + \frac{\Delta}{\sigma}\right)\right) \\ &= \Phi\left(k + \frac{\Delta}{\sigma}\right) + \Phi\left(k - \frac{\Delta}{\sigma}\right) - 1 = \Phi\left(k + \frac{|\Delta|}{\sigma}\right) + \Phi\left(k - \frac{|\Delta|}{\sigma}\right) - 1. \end{aligned}$$

Then

$$\begin{aligned} p &\leq 1 - \mathbb{P}\left(|\Delta + E| \leq k\sigma\right) = 1 - \left(\Phi\left(k + \frac{|\Delta|}{\sigma}\right) + \Phi\left(k - \frac{|\Delta|}{\sigma}\right) - 1\right) \\ &= 2 - \left(\Phi\left(k + \frac{|\Delta|}{\sigma}\right) + \Phi\left(k - \frac{|\Delta|}{\sigma}\right)\right). \end{aligned}$$

Since

$$\begin{aligned}\frac{d}{dx}\left(\Phi\left(k+x\right)+\Phi\left(k-x\right)\right) &= \Phi'\left(k+x\right)-\Phi'\left(k-x\right)\\ &= f_Z\left(k+x\right)-f_Z\left(k-x\right)<0\end{aligned}$$

for $k-x>0$, i.e. $x<k$, the function

$$x\mapsto\Phi\left(k+x\right)+\Phi\left(k-x\right),\ x\leq k,$$

is decreasing.

So, for $k\geq 1.5\geq\frac{|\Delta|}{\sigma}$, we have

$$\begin{aligned}p &\leq 2-\left(\Phi\left(k+\frac{|\Delta|}{\sigma}\right)+\Phi\left(k-\frac{|\Delta|}{\sigma}\right)\right)\\ &\leq 2-\left(\Phi\left(k+1.5\right)+\Phi\left(k-1.5\right)\right).\end{aligned}$$

Indeed,

$$\begin{aligned} p &\leq 2 - (\Phi(k+1.5) + \Phi(k-1.5)) \\ &= 1 - \Phi(k-1.5) + 1 - \Phi(k-1.5) \\ &= (1 - \Phi(k-1.5))\left(1 + \frac{1-\Phi(k+1.5)}{1-\Phi(k-1.5)}\right) \\ &\approx 1 - \Phi(k-1.5) \end{aligned}$$

since

$$\frac{1-\Phi(k+1.5)}{1-\Phi(k-1.5)} \text{ is small.}$$

So, for $k = 3$ (Three-Sigma methodology), we have

$$p \lessapprox 1 - \Phi(1.5) = 6.7\%.$$

and, for $k = 6$ (Six-Sigma methodology), we have

$$p \lessapprox 1 - \Phi(4.5) = 3.4 \cdot 10^{-6}.$$

- In general, we say that a process producing some type di result, which can be not defective or defective, adopts the Six-Sigma methodology if the probability of having a defective result has order of magnitude 10^{-6}.

  For example, a Six-Sigma surgery has a rate of complications after surgery with order of magnitude 10^{-6}.

  The Six-Sigma methodology introduced by Motorola is used also by General Electric, Toyota, Honeywell and Microsoft.

  On the other hand, we say that the process adopts the Three-Sigma rule if the probability of having a defective result has order of magnitude 10^{-3}.

Percentiles

Definition

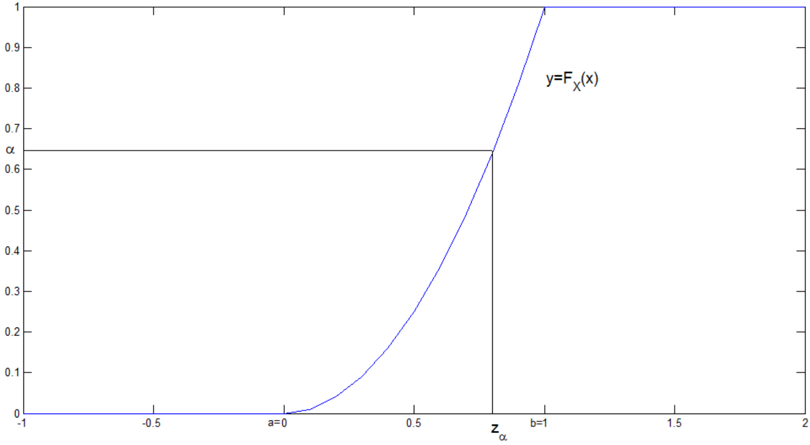
Let X be a continuous random variable whose distribution function F_X is such that:

- $F_X(x) = 0$ for $x \leq a$;
- F_X is strictly increasing in (a, b);
- $F_X(x) = 1$ for $x \geq b$;

for some a, b with $-\infty \leq a < b \leq +\infty$. For any $\alpha \in (0, 1)$, the 100αth **percentile**, or the **quantile** α, of X is the number $z_\alpha \in (a, b)$ such that

$$\mathbb{P}(X \leq z_\alpha) = F_X(z_\alpha) = \alpha.$$

An illustration of z_α:



Observe that such a number z_α exists and it is unique since F_X is continuous and strictly increasing in (a, b) with range $(0, 1)$.

Moreover, z_α is determined by F_X and then it depends only on the distribution of X, not on the particular random variable X with that distribution. So, we can also say that z_α is the 100αth percentile of the distribution of X.

Also observe that the conditions on F_X are satisfied if there exist a, b with $-\infty \leq a < b \leq +\infty$ such that f_X is positive in (a, b) and zero outside.

The percentiles 25th, 50th and 75th are called the **quartiles** of X. The 50th percentile is called the **median** of X.

Exercise. Consider a random variable X with uniform distribution $U(a, b)$. Determine z_α, for $\alpha \in (0, 1)$.

Exercise. Explain why it is not possible to introduce the notion of a percentile for a discrete random variable.

- In the frequentist interpretation, where we repeat the experiment relevant to the random variable X a very large number n of independent times with outcomes $\omega_1^{\text{obs}}, \ldots, \omega_n^{\text{obs}}$, the 100αth percentile of X, $\alpha \in (0,1)$, is the value z_α such that

$$\alpha = \mathbb{P}(X \leq z_\alpha) \approx \frac{\left|\left\{i \in \{1,2,\ldots,n\} : X\left(\omega_i^{\text{obs}}\right) \leq z_\alpha\right\}\right|}{n}.$$

  So, z_α divides the ordered data $\boldsymbol{x}^{\text{ord}}$, where

$$\boldsymbol{x} = \left(X\left(\omega_1^{\text{obs}}\right), \ldots, X\left(\omega_n^{\text{obs}}\right)\right),$$

  in two parts, one has the components not larger than z_α and the other has the components larger than z_α, and the sizes of these two parts are (approximately) proportional to α and $1-\alpha$, respectively.

  We conclude that z_α is (approximately) the 100αth percentile of the data $\boldsymbol{x}$.

- In the following, the 100αth percentile z_α of a random variable X will be denoted by $z_\alpha(X)$, by reserving the symbol z_α without any indication of a random variable for the 100αth percentile of a standard normal random variable.

  Observe that z_α is such that $\Phi(z_\alpha) = \alpha$, i.e.

  $$z_\alpha = \Phi^{-1}(\alpha),$$

  where Φ is the distribution function of a a standard normal random variable.

If X has distribution $\mathrm{N}\left(\mu, \sigma^2\right)$ (in this case $a = -\infty$ and $b = +\infty$), the $100\alpha-$th percentile of X, where $\alpha \in (0, 1)$, is the number $z_\alpha\left(X\right)$ such that

$$\alpha = \mathbb{P}\left(X \leq z_\alpha\left(X\right)\right) = \mathbb{P}\left(Z = \frac{X - \mu}{\sigma} \leq \frac{z_\alpha\left(X\right) - \mu}{\sigma}\right).$$

We see that

$$\frac{z_\alpha\left(X\right) - \mu}{\sigma}$$

is the 100αth percentile z_α of Z and we conclude that

$$z_\alpha\left(X\right) = \mu + \sigma z_\alpha.$$

So, percentiles of a normal distribution are computed from percentiles of the standard normal distribution.

- Now, we see how to compute the percentiles z_α, $\alpha \in (0, 1)$, of a standard normal distribution.

  For z_α with $\alpha \geq 0.5$, we can use the table of the values $\Phi(x)$, $x \geq 0$. We have to find z_α such that

  $$z_\alpha = \Phi^{-1}(\alpha).$$

  Suppose that α is included in the interval $[\Phi(x_1), \Phi(x_2)]$, where x_1 and x_2 are two consecutive tabulated abscissas in the table.

  For example, $\alpha = 0.6$ is included in the interval

  $$[0.5987, 0.6026] = [\Phi(0.25), \Phi(0.26)].$$

A first rough approximation of z_α is given by

$$z_\alpha = \begin{cases} x_1 \text{ if } \alpha \text{ is closer to } \Phi(x_1) \text{ than } \Phi(x_2) \\ x_2 \text{ otherwise.} \end{cases}$$

In our example of $\alpha = 0.6$, the rough approximation is $z_{0.6} = 0.25$, since 0.6 is closer to $\Phi\,(0.25) = 0.5987$ than $\Phi\,(0.26) = 0.6026$.

A better approximation of z_α can be obtained by the **linear interpolation** of the values $\Phi(x_1)$ and $\Phi(x_2)$.

We replace, in the interval $[x_1, x_2]$, the function $\Phi(x)$ with the straigth line passing through the points $(x_1, \Phi(x_1))$ and $(x_2, \Phi(x_2))$:

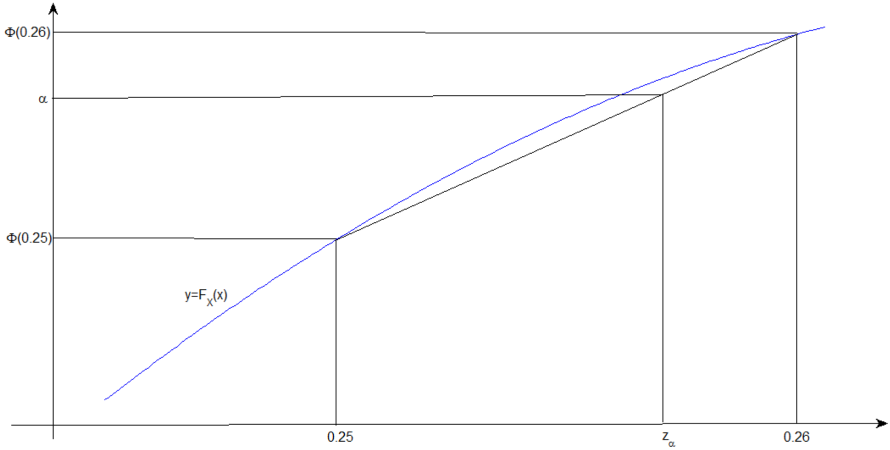
$$y = f(x) = \Phi(x_1) + \frac{\Phi(x_2) - \Phi(x_1)}{x_2 - x_1}(x - x_1)$$

and then we look for z_α such that

$$f(z_\alpha) = \alpha,$$

instead of $\Phi(z_\alpha) = \alpha$.

An illustration of the linear interpolation:



By solving

$$f(z_\alpha) = \Phi(x_1) + \frac{\Phi(x_2) - \Phi(x_1)}{x_2 - x_1}(z_\alpha - x_1) = \alpha$$

we obtain

$$z_\alpha = x_1 + \frac{x_2 - x_1}{\Phi\left(x_2\right) - \Phi\left(x_1\right)} \cdot \left(\alpha - \Phi\left(x_1\right)\right).$$

In our example of $\alpha = 0.6$, the approximation is

$$\begin{aligned} z_{0.6} &= 0.25 + \frac{0.26 - 0.25}{\Phi\left(0.26\right) - \Phi\left(0.25\right)} \cdot \left(0.6 - \Phi\left(0.25\right)\right) \\ &= 0.25 + \frac{0.01}{0.6026 - 0.5987} \cdot (0.6 - 0.5987) = 0.2533. \end{aligned}$$

For z_α with $\alpha < 0.5$, we have

$$\alpha = \Phi\left(z_\alpha\right) = 1 - \Phi\left(-z_\alpha\right)$$

and so

$$z_\alpha = -\Phi^{-1}\left(1 - \alpha\right) = -z_{1-\alpha}$$

with $1 - \alpha > 0.5$ and so z_α can be computed by the table as just described.

For example, for $\alpha = 0.2$, we have

$$z_{0.2} = -z_{0.8}.$$

Since 0.8 is included in

$$[0.7995, 0.8023] = [\Phi\left(0.84\right), \Phi(0.85)],$$

the rough approximation is $z_{0.2} = -0.84$ and the approximation by linear interpolation is

$$z_{0.2} = -\left(0.84 + \frac{0.01}{0.8023 - 0.7995} \cdot (0.8 - 0.7995)\right) = -0.8418$$

- Exercise. Find the three quartiles of a standard normal random variable. Then, find the three quartiles of a general normal random variable with distribution $\mathrm{N}(\mu, \sigma^2)$.

- *Example. The IQ test on sixth-graders produces a score that is normally distributed with* $\mu = 100$ *and* $\sigma = 14.2$*.*

  Assume that a large number of sixth-graders do the test. What is the value v over which there is the top 1 percent of all scores?

  Let X be the random variable score, whose distribution is $N(\mu, \sigma^2)$*. Since (by using the frequentist interpretation)*

$$\begin{aligned} &99\% \\ &= \frac{\textit{number of sixth graders who did the test not in the top } 1\%}{\textit{number of sixth graders who did the test}} \\ &\approx \mathbb{P}(X \leq v), \end{aligned}$$

  we have $v = z_{99\%}(X) = \mu + \sigma z_{99\%}$*.*

  The rough approximation of the 99*th percentile* $z_{99\%}$ *of a standard normal variable is* 2.33*:* 0.99 *is included in*

$$[0.9898, 0.9901] = [\Phi(2.32), \Phi(2.33)]$$

So, the rough approximation of the 99th percentile of X is

$$z_{99\%}(X) = \mu + \sigma z_{99\%} = 100 + 14.2 \cdot 2.33 = 133.1.$$

We conclude that the top 1% has scores larger than 133.

Exercise. Find the 99th percentile of X by using the approximation by linear interpolation.

- Exercise. By assuming that the height of an italian male is normally distributed $\mathrm{N}\left(176\ \mathrm{cm}, (7\ \mathrm{cm})^2\right)$, find the range of heights between the first and third quartiles.

  Exercise. What is the probability that a continuous random variable X lies between its 100αth and 100βth percentiles (with $\alpha < \beta$)?

  Exercise. Consider the example of the bulb lights. How many light bulbs the individual needs to purchase in order to have

  $$\mathbb{P}(S \geq k\mu) \geq C\%$$

  for arbitrary $k > 0$ and $C\% \in (0, 1)$ (in the previous example, we answered in case of $k \leq \left(\frac{\frac{\mu}{\sigma}}{3.49}\right)^2 - 1 = 7.21$ and $C\% = 99.98\%$).

- In MATLAB, the percentiles of a normal random variable are computed by the function norminv:

$$\text{norminv}(\alpha),$$

where $\alpha \in (0, 1)$, computes $\Phi^{-1}(\alpha)$, i.e. the 100αth percentile of a standard normal random variable, and

$$\text{norminv}(\alpha, \mu, \sigma)$$

computes the 100αth percentile of a normal random variable with distribution $\mathrm{N}(\mu, \sigma^2)$.

Exercise. By using MATLAB, compute the 99th percentile of the normal random variable IQ score and compare such value with the values previously computed by the table.

Exercise. Find the deciles of a normal random variable with mean μ and standard deviation σ.

- Observe that when we will deal with the "Testing Statistical Hypotheses", the symbol z_α, $\alpha \in (0, 1)$, will denote

  $\Phi^{-1}(1-\alpha) = 100(1-\alpha)$th percentile of the standard normal distribution,

  i.e. our $z_{1-\alpha}$.

  So, in this context, $z_{0.01} = z_{1\%}$ is our $z_{0.99} = z_{99\%} = 2.33$.

Mixed random variables

- There are random variables which are neither discrete nor continuous, but a mixed between the two types.

  As an example, consider a car policy proposed by an insurance company. Let X be the random variable yearly claim of a policyholder.

  X is neither discrete nor continuous. In fact, X cannot be continuous since $\mathbb{P}(X = 0)$, the probability that the policyholder has not car accidents during the year, is not zero. On the other hand, X cannot be discrete since the yearly claim of a policyholder can be any positive number and we can assume, for any $x \in \mathbb{R}$,

$$\mathbb{P}(0 < X \leq x) = \int_0^x f_X(y)dy$$

  for some pdf f_X.

Exercise. Describe and draw the graph of the distribution function F_X.

Exercise. If the policy has a deductible and a maximum coverage, what is the form of the pdf f_X?

- The previous random variable X is an example of a mixed random variable.

Definition

A random variable $Y : \Omega \to \mathbb{R}$ is called **mixed** if there exist points $y_1, y_2, \ldots, y_n \in \mathbb{R}$ with $y_1 < y_2 < \cdots < y_n$ such that

$$\mathbb{P}(Y = y_i) = p_i, \; i \in \{1, \ldots, n\},$$
$$\mathbb{P}(\infty < Y < y_1) = \int_{-\infty}^{y_1} g(y)dy$$
$$\mathbb{P}(y_i < Y < y_{i+1}) = \int_{y_i}^{y_{i+1}} g(y)dy, \; i \in \{1, \ldots, n-1\}$$
$$\mathbb{P}(y_n < Y < +\infty) = \int_{y_n}^{+\infty} g(y)dy.$$

for some numbers p_i, $i \in \{1, \ldots, n\}$, and for some function $g : \mathbb{R} \to \mathbb{R}$.

Observe that the distribution F_Y of a mixed random variable Y is a piecewise integral function with jumps:

$$F_Y(y) = \sum_{\substack{i=1 \\ y_i \leq y}}^{n} p_i + \int_{-\infty}^{y} g(s)ds, \; y \in \mathbb{R}.$$

Exercise. Draw the graph of F_Y.

- A more general definition of a mixed random variable includes the situation where instead of a finite sequence of points $y_1, y_2, \ldots, y_n$, we have an infinite sequence of them. A random variable Y is called mixed if there exist a discrete subset A of $\mathbb{R}$, a function $p : A \to \mathbb{R}$ and a function $g : \mathbb{R} \to \mathbb{R}$ such that

$$F_Y(y) = \sum_{\substack{s \in A \\ s \leq y}} p(s) + \int_{-\infty}^{y} g(s) ds, \ y \in \mathbb{R}.$$

  Observe that the mixed random variable Y has both a pmf, the function p, and a pdf, the function g.

  Moreover, the discrete random variables are the mixed random variables with $g = 0$ and the continuous random variable are the mixed random variables with $A = \emptyset$.

- For a mixed random variable Y we have

$$\sum_{s \in A} p(s) + \int_{-\infty}^{+\infty} g(s)ds = 1.$$

The mean of the mixed random variable Y is defined as

$$\mathbb{E}(Y) = \sum_{s \in A} s \cdot p(s) + \int_{-\infty}^{+\infty} s \cdot g(s)ds.$$

Variance and standard deviation for a mixed random variable are defined as usual, the notion of independence of a sequence of mixed random variables is exactly the same as for continuous random variables and all the results regarding discrete or continuous random variables can be extended to mixed random variables.