5.4 Probability distributions

We will now study in some detail some of the most important continuous probability distributions that one encounters in many situations.

5.4.1 Uniform distribution

Definition 5.9 (Uniform distribution)
Let X be a continuous random variable and $a < b$ two real numbers. We say that X is *uniform over* $[a,b]$, denoted $X \sim \mathscr{U}([a,b])$, if its probability density function is constant over $[a,b]$ and vanishes otherwise:

$$f(x) = \begin{cases} \dfrac{1}{b-a} \text{ if } x \in [a,b] \\ 0 \text{ otherwise.} \end{cases}$$

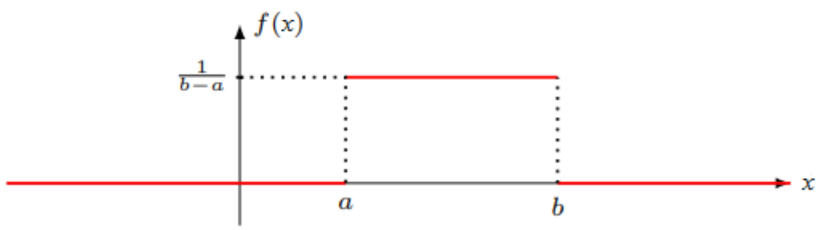


Figure 5.5 – uniform probability density function $\mathscr{U}([a,b])$

♦<u>Examples:</u>

⋆ Read again the example of the direction in which an alpha particle is emitted in radioactive decay (page 102).

⋆ Suppose metro line 8 passes through Place d'Italie station exactly every six minutes between 6am and 8am on working days. An EFREI student travelling to the République campus arrives at Place d'Italie station at a given time, uniformly distributed between 7am and 7.30am. Calculate the probability that the student will wait less than two minutes to take the metro to Place d'Italie. Let T be the student's moment of arrival to Place d'Italie. We have $T \sim \mathscr{U}(0,30)$ (where the value 0 refers to the time 7h). Now, consider the event E "the student waits less than two minutes to get on the next metro carriage". We have

$$E = \{4 \leq X \leq 6\} \cup \{10 \leq X \leq 12\} \cup \{16 \leq X \leq 18\} \cup \{22 \leq X \leq 24\} \cup \{28 \leq X \leq 30\}$$

and therefore, since the different events are incompatible,

$$\begin{aligned}
P(E) &= P(4 \le X \le 6) + P(10 \le X \le 12) + P(16 \le X \le 18) + P(22 \le X \le 24) + P(28 \le X \le 30) \\
&= \frac{1}{30}\left(\int_4^6 dt + \int_{10}^{12} dt + \int_{16}^{18} dt + \int_{22}^{24} dt + \int_{28}^{30} dt\right) \\
&= \frac{1}{3}.
\end{aligned}$$

Proposition 5.10 : Expected value and variance of the uniform distribution

Si $X \sim \mathscr{U}([a,b])$, alors $E(X) = \dfrac{a+b}{2}$ and $V(X) = \dfrac{(b-a)^2}{12}$.

Proof. If $X \sim ([a,b])$, we have

$$E(X) = \int_{-\infty}^{+\infty} x f(x) dx = \frac{1}{b-a}\int_a^b x dx = \frac{b^2-a^2}{2(b-a)} = \frac{a+b}{2}.$$

Similarly, using theorem 5.8, we have

$$E(X^2) = \frac{1}{b-a}\int_a^b x^2 dx = \frac{b^3-a^3}{3(b-a)} = \frac{a^2+ab+b^2}{3}.$$

and thus

$$V(X) = E(X^2) - E(X)^2 = \frac{a^2+ab+b^2}{3} - \frac{(a+b)^2}{4} = \frac{(b-a)^2}{12}.$$

□

5.4.2 Exponential distribution

We now meet an important distribution which appears every time we are dealing with memory-less phenomena such as the lifetime of a radioactive particle (or, less intuitively, the lifetime of a computer).

Definition 5.10 (Exponential distribution)
Let X be a continuous random variable and λ a strictly positive real number. Then, X is an *exponential random variable of parameter* λ, denoted $X \sim \mathscr{E}(\lambda)$, if its probability density function is given by:

$$f(x) = \begin{cases} \lambda e^{-\lambda x} \text{ if } x \ge 0 \\ 0 \text{ otherwise.} \end{cases}$$

Examples:

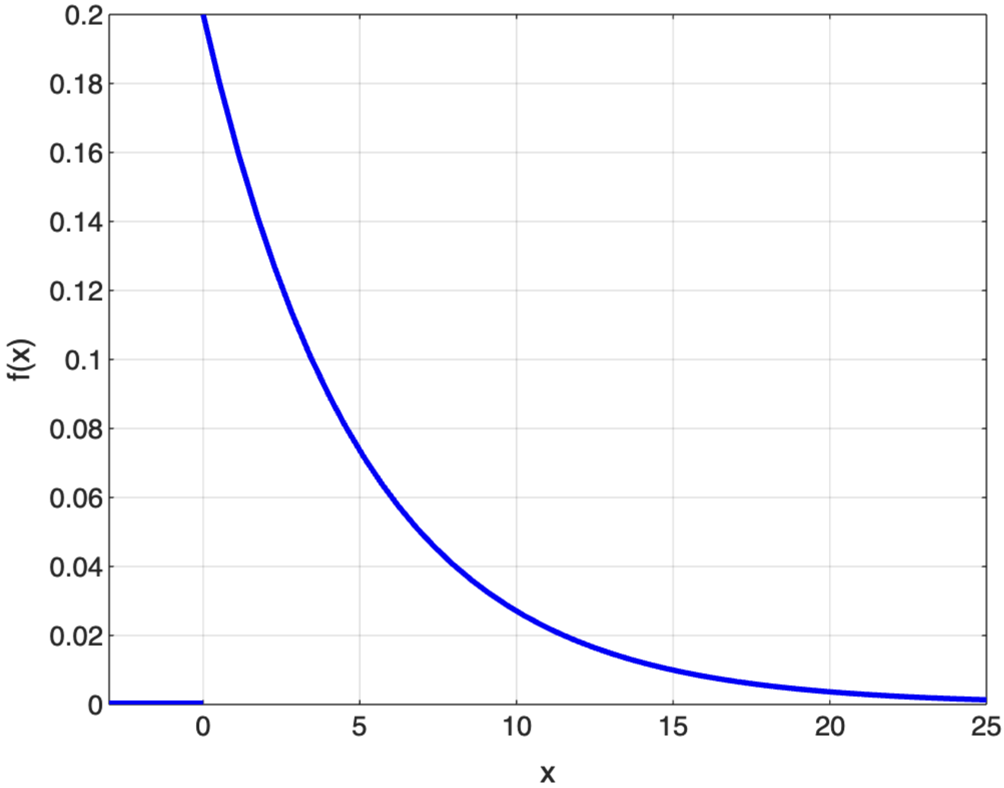


Figure 5.6 – Plot of the probability density function for an exponential random variable of parameter $\lambda = 1/5$.

- ⋆ The already mentioned lifetime of a radioactive atom or a computer(see page 102).
- ⋆ The waiting time in a line, the time of a phone call, the distance travelled by a meteorite in the Earth's atmosphere before it disintegrates completely.

Proposition 5.11 : Expected value and variance for an exponential random variable

If $X \sim \mathscr{E}(\lambda)$, then $E(X) = \dfrac{1}{\lambda}$ and $V(X) = \dfrac{1}{\lambda^2}$.

Proof. Using integration by parts, we have

$$E(X) = \int_{-\infty}^{+\infty} x f(x) dx = \lambda \int_0^{+\infty} x e^{-\lambda x} dx = \frac{1}{\lambda}.$$

Similarly, integrating by parts twice, we find

$$E(X^2) = \int_{-\infty}^{+\infty} x^2 f(x) dx = \lambda \int_0^{+\infty} x^2 e^{-\lambda x} dx = \frac{2}{\lambda^2}$$

and therefore

$$V(X) = E(X^2) - E(X)^2 = \frac{1}{\lambda^2}.$$

□

The next proposition gives an important property characterising exponential distributions:

Proposition 5.12 : Exponential distributions are memory-less

We say that a continuous random variable T is *memory-less* if it satisfies

$$\forall s \geq 0, P(T \geq t+s \middle| T \geq t) = P(T \geq s).$$

Then, we have the following equivalence:

$$T \text{ is memory-less} \iff T \text{ is an exponential random variable.}$$

Proof. See exercise 5.13. □

Remark: We need to understand this property simply: consider, for example, the probability that an electronic component will work for at least s days. This probability is written as $P(T \geq s)$. The next day, if the component is still working, i.e. "knowing that it has lasted at least one day" $(T \geq 1)$, this first probability will be exactly the same: $P(T \geq 1+s \middle| T \geq 1) = P(T \geq s)$. By repeating the reasoning, we realise that, regardless of the number of days that have elapsed, every day the probability that the component will last s days longer will be the same, right up to the moment when it no longer functions.

Example: or a radioactive nucleus, the fact that it has not yet decayed (and therefore that a certain amount of time has passed without it having decayed) has no influence on the probability that the nucleus will decay in the future.

5.4.3 Normal distrbution

From all the distributions, the normal distribution (also known as Gaussian distribution) is by far the most important. This probability distribution is omnipresent in the modelling of physical phenomena: it is used to describe, for example, the size of the stars in a galaxy, the blood pressure of the human body, the size of a human being and the fundamental state of the quantum harmonic oscillator.

Definition 5.11 (Normal or Gaussian distribution)
Let X be a continuous random variable and $(\mu \times \sigma) \in \mathbb{R} \times \mathbb{R}^{+*}$. X is a *normal random variable of parameters* μ *and* σ, denoted $X \sim \mathcal{N}(\mu, \sigma^2)$, if its probability density function is given by

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

In particular, if $X \sim \mathcal{N}(0,1)$, X is said to be *centered reduced normal distribution*.

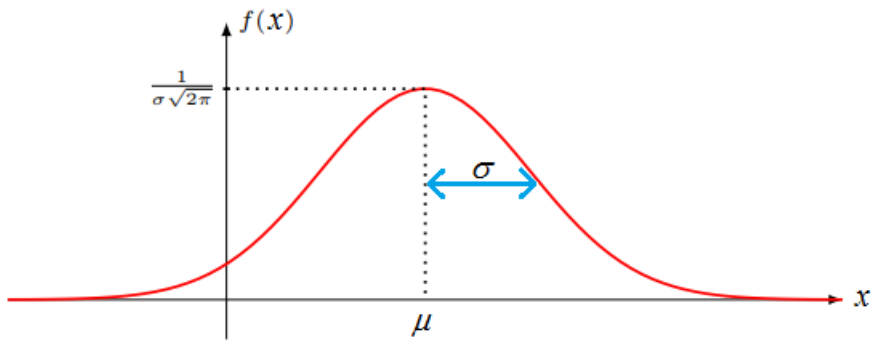
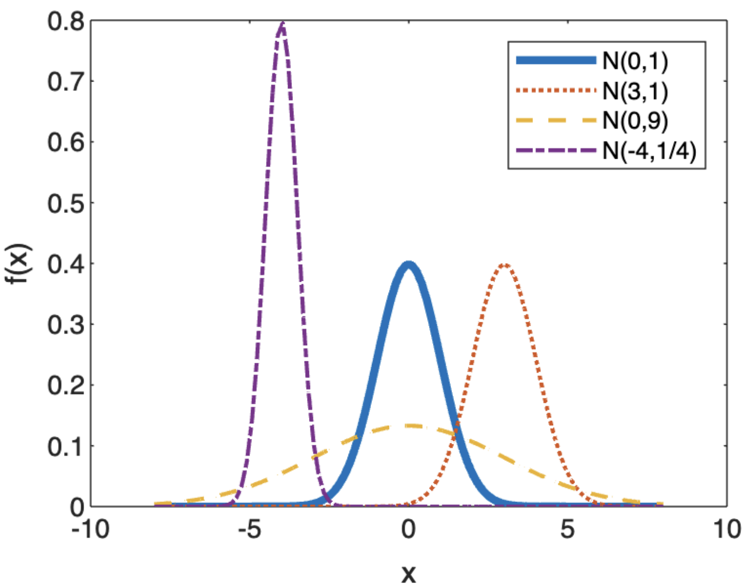


Figure 5.7 – Shape of the normal distribution. The parameter μ fixed the symmetry axis of the distribution and the parameter σ represents the distance from this position to the inflexion point.



Matlab script for this figure:

```
x=linspace(-8,8);
y1=normpdf(x,0,1); y2=normpdf(x,3,1);
y3=normpdf(x,0,3); y4=normpdf(x,-4,1/2);
plot(x,y1,'LineWidth',3)
hold on
plot(x,y2,":",'LineWidth',1.5)
plot(x,y3,"-",'LineWidth',1.5)
plot(x,y4,"-.",'LineWidth',1.5)
hold off
legend(["N(0,1)" "N(3,1)" "N(0,9)"
"N(-4,1/4)"])
xlabel('x')
ylabel('f(x)')
```

Figure 5.8 – Gaussian distribution for different values of the parameters.

Remark: Changing the value of μ moves the plot towards the right or the left. Decreasing σ, the peak becomes more pronounced, whereas decreasing σ makes the curve more spread out.

In practice, all the probability computations about a normal distribution are carried out by changing to a centered reduced random variable and using the values of the cumulative distribution function of the centered reduced normal distribution.

Definition 5.12 (Cumulative distribution function of the centered reduced normal random variable)
We denote by Φ the cumulative distribution function of the centered reduced normal random variable. In other words, we have

$$\forall x \in \mathbb{R}, \Phi(x) = P(X \leq x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-x^2/2} dx.$$

Proposition 5.13 : Properties of the cumulative distribution function Φ

Let $X \sim \mathcal{N}(0,1)$, Φ be the cumulative distribution function of X and $a \leq b \in \mathbb{R}$. We have

- ⋆ $P(a \leq X \leq b) = \Phi(b) - \Phi(a)$,
- ⋆ $P(X \geq a) = P(X \leq -a) = 1 - \Phi(a)$,
- ⋆ $\Phi(-a) = 1 - \Phi(a)$.

Proof. The best is to have in mind the following drawings:

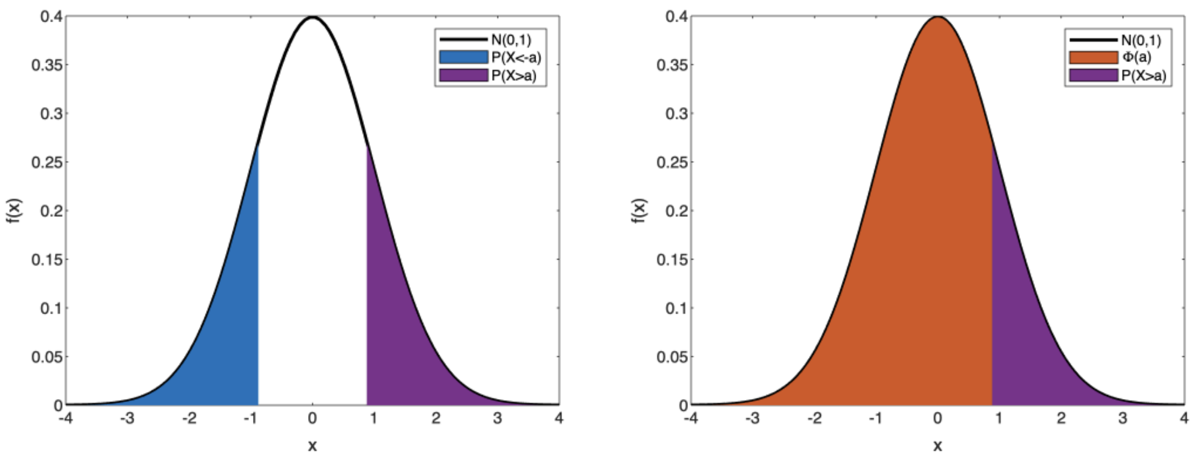


Figure 5.9 – The equalities $P(X \geq a) = P(X \leq -a)$ (left)
and $P(X \geq a) + \Phi(a) = 1$ (right).
(Plots were generated for $a = 0.88$.)

□

Methodological flashcard #7 (Computing the probabilities of a normal random variable)
Let $X \sim \mathcal{N}(\mu, \sigma^2)$ and I_1, I_2 be two intervals of $\mathbb{R}$. To $P(X \in I_1 \big| X \in I_2)$, one performs the following steps:

a) Introduce $Z = \frac{X-\mu}{\sigma}$. In this way, Z is a centered reduced normal variable: $Z \sim \mathcal{N}(0,1)$.

b) Express the events $\{X \in I_1\}$ and $\{X \in I_2\}$ using the new variable Z.

c) Use the properties of proposition 5.13 to express the probability $P(X \in I_1 \big| X \in I_2) = P(Z \in I_3 \big| Z \in I_4)$ in terms of the function Φ.

d) Use the table of values of Φ (see next page) to find an estimated value of $P(X \in I_1 \big| X \in I_2)$.

♦Example: Let $X \sim \mathcal{N}(6,4)$. Compute $P(|X-4| < 3 \big| X \geq 2)$.

a) Start by introducing $Z = \frac{X-6}{2}$. Z is a centered reduced normal random variable.

b) Moreover, we have

$$|X-4| < 3 \iff -3 < X-4 < 3 \iff -\frac{5}{2} < Z < \frac{1}{2}$$

and also

$$X \geq 2 \iff Z \geq -2.$$

c) Therefore, we have

$$\begin{aligned} P(|X-4| < 3 \big| X \geq 2) &= P(-\frac{5}{2} < Z < \frac{1}{2} \big| Z \geq -2) \\ &= \frac{P(\{-\frac{5}{2} < Z < \frac{1}{2}\} \cap \{Z \geq -2\}}{P(Z \geq -2)} \\ &= \frac{P(\{-2 \leq Z < \frac{1}{2}\}}{P(Z \geq -2)} \\ &= \frac{\Phi(\frac{1}{2}) - \Phi(-2)}{1 - \Phi(-2)} \\ &= \frac{\Phi(\frac{1}{2}) + \Phi(2) - 1}{\Phi(2)} \end{aligned}$$

d) Using the table of values of the function Φ, we find:

$$P(|X-4| < 3 \big| X \geq 2) \approx \frac{0.69146 + 0.97725 - 1}{0.97725} \approx 0.68428.$$

z	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
+0	.50000	.50399	.50798	.51197	.51595	.51994	.52392	.52790	.53188	.53586
+0.1	.53983	.54380	.54776	.55172	.55567	.55966	.56360	.56749	.57142	.57535
+0.2	.57926	.58317	.58706	.59095	.59483	.59871	.60257	.60642	.61026	.61409
+0.3	.61791	.62172	.62552	.62930	.63307	.63683	.64058	.64431	.64803	.65173
+0.4	.65542	.65910	.66276	.66640	.67003	.67364	.67724	.68082	.68439	.68793
+0.5	.69146	.69497	.69847	.70194	.70540	.70884	.71226	.71566	.71904	.72240
+0.6	.72575	.72907	.73237	.73565	.73891	.74215	.74537	.74857	.75175	.75490
+0.7	.75804	.76115	.76424	.76730	.77035	.77337	.77637	.77935	.78230	.78524
+0.8	.78814	.79103	.79389	.79673	.79955	.80234	.80511	.80785	.81057	.81327
+0.9	.81594	.81859	.82121	.82381	.82639	.82894	.83147	.83398	.83646	.83891
+1	.84134	.84375	.84614	.84849	.85083	.85314	.85543	.85769	.85993	.86214
+1.1	.86433	.86650	.86864	.87076	.87286	.87493	.87698	.87900	.88100	.88298
+1.2	.88493	.88686	.88877	.89065	.89251	.89435	.89617	.89796	.89973	.90147
+1.3	.90320	.90490	.90658	.90824	.90988	.91149	.91308	.91466	.91621	.91774
+1.4	.91924	.92073	.92220	.92364	.92507	.92647	.92785	.92922	.93056	.93189
+1.5	.93319	.93448	.93574	.93699	.93822	.93943	.94062	.94179	.94295	.94408
+1.6	.94520	.94630	.94738	.94845	.94950	.95053	.95154	.95254	.95352	.95449
+1.7	.95543	.95637	.95728	.95818	.95907	.95994	.96080	.96164	.96246	.96327
+1.8	.96407	.96485	.96562	.96638	.96712	.96784	.96856	.96926	.96995	.97062
+1.9	.97128	.97193	.97257	.97320	.97381	.97441	.97500	.97558	.97615	.97670
+2	.97725	.97778	.97831	.97882	.97932	.97982	.98030	.98077	.98124	.98169
+2.1	.98214	.98257	.98300	.98341	.98382	.98422	.98461	.98500	.98537	.98574
+2.2	.98610	.98645	.98679	.98713	.98745	.98778	.98809	.98840	.98870	.98899
+2.3	.98928	.98956	.98983	.99010	.99036	.99061	.99086	.99111	.99134	.99158
+2.4	.99180	.99202	.99224	.99245	.99266	.99286	.99305	.99324	.99343	.99361
+2.5	.99379	.99396	.99413	.99430	.99446	.99461	.99477	.99492	.99506	.99520
+2.6	.99534	.99547	.99560	.99573	.99585	.99598	.99609	.99621	.99632	.99643
+2.7	.99653	.99664	.99674	.99683	.99693	.99702	.99711	.99720	.99728	.99736
+2.8	.99744	.99752	.99760	.99767	.99774	.99781	.99788	.99795	.99801	.99807
+2.9	.99813	.99819	.99825	.99831	.99836	.99841	.99846	.99851	.99856	.99861
+3	.99865	.99869	.99874	.99878	.99882	.99886	.99889	.99893	.99896	.99900

Figure 5.10 – Table of the values of the function Φ. It gives the values, with five significant digits, of $\Phi(z)$ for z between 0 and 3 and defined to the nearest hundredth. For instance, the value $\Phi(1.73)$ is found at the fourth column of the row $+1.7$: one reads $\Phi(1.73) = 0.95818$.

This table was generated by the website www.ztable.net.

Proposition 5.14 : Expected value and variance of a normal random variable

If $X \sim \mathcal{N}(\mu, \sigma^2)$, then $E(X) = \mu$ and $V(X) = \sigma^2$.

Proof. Let $X \sim \mathcal{N}(\mu, \sigma^2)$. Since $Z = \dfrac{X-\mu}{\sigma} \sim \mathcal{N}(0,1)$ and we have $E(Z) = \dfrac{E(X)-\mu}{\sigma}$ and $V(Z) = \dfrac{V(X)}{\sigma^2}$, it suffices to carry out the computation for $Z \sim \mathcal{N}(0,1)$ to prove the general case.

We have

$$E(Z) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} z e^{-z^2/2} dz = 0$$

since the integrand is an odd function.

Similarly, we have

$$\begin{aligned} V(Z) = E(Z^2) &= \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} z^2 e^{-z^2/2} dz \\ &= -\frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} z \times \left(-z e^{-z^2/2}\right) dz \\ &= \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-z^2/2} dz \\ &= P(Z \in \mathbb{R}) = 1 \end{aligned}$$

where, from the second to the third equation, we perform an integration by parts by noticing that $-ze^{-z^2/2}$ is the derivative of $e^{-z^2/2}$. □

The following theorem, a particular case of the central limit theorem (see the historical note below), largely explains the predominance of the normal distribution in the modelling of physical phenomena:

Theorem 5.15 : Link between binomial and normal random variables

In the case where n tends to infinity, binomial distributions tends to normal distributions. More precisely:

Let $(X_n)_{n\in\mathbb{N}}$ a sequence of discrete random variables such that $X_n \sim \mathcal{B}(n,p)$.

Let $Z_n = \dfrac{X_n - np}{\sqrt{np(1-p)}}$ the centered reduced random variable associated to X_n. Then, we have

$$\lim_{n\to+\infty} Z_n \sim \mathcal{N}(0,1).$$

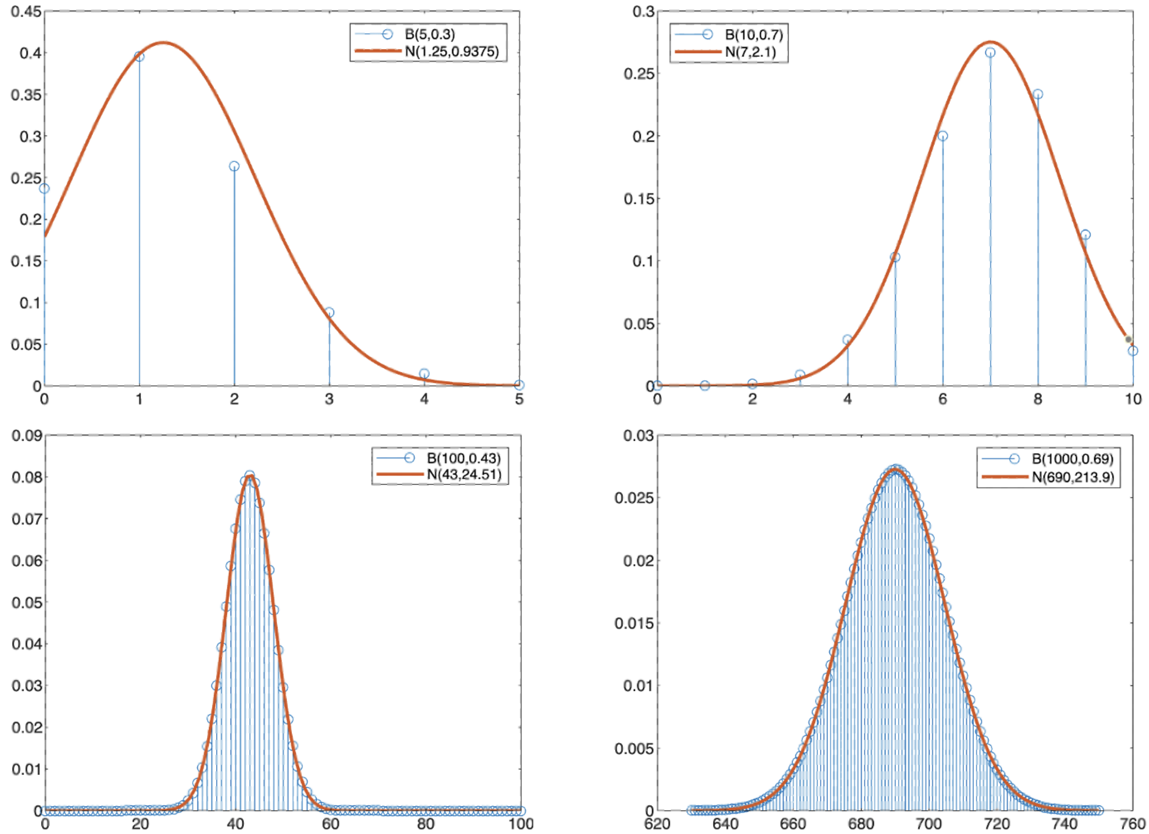


Figure 5.11 – Comparison of the binomial distribution and normal distributions for different values of the parameters n and p. We see that, even for small values of n, it is a good approximation.

Proposition 5.16 : Linear combination of normal distributions

Let X,Y be two normal random variables. Then, any linear combination of X and Y is again a normal random variable.

More precisely, suppose that $X \sim \mathcal{N}(\mu_1, \sigma_1^2)$ and $Y \sim \mathcal{N}(\mu_2, \sigma_2^2)$ and consider $Z = aX + bY$ with $a,b \in \mathbb{R}$:

- ⋆ If X and Y independent, then $Z \sim \mathcal{N}(a\mu_1 + b\mu_2, a^2\sigma_1^2 + b^2\sigma_2^2)$.
- ⋆ If X and Y not independent, then $Z \sim \mathcal{N}(a\mu_1 + b\mu_2, a^2\sigma_1^2 + b^2\sigma_2^2 + 2abCov(X,Y))$.

Proof. In the case of independent variables, see exercise 5.19. The second case is admitted. □

♦Example: Denote by X the size of a man in the French population and Y the size of a woman in the French population. Given that $X \sim \mathcal{N}(176, 64)$ and $Y \sim \mathcal{N}(164, 49)$, find the probability that, if one takes at random one man and one woman, the woman selected is taller than the man.

We therefore wish to $P(Y > X)$ which can also be written as $P(Y - X > 0)$. According to the previous proposition, we have $Y - X \sim \mathcal{N}(-12, 113)$. To compute the sought-for probability, apply the steps of method 7 (page 121):

a) Introduced the centered reduced variable $Z = \dfrac{Y - X + 12}{\sqrt{113}}$.

b) $Y - X > 0 \iff Z > \dfrac{12}{\sqrt{113}}$.

c) Therefore $P(Y > X) = P(Z > \dfrac{12}{\sqrt{113}}) = 1 - \Phi(\dfrac{12}{\sqrt{113}})$

d) Since $\dfrac{12}{\sqrt{113}} \approx 1.13$, we have, according to the table of values Φ (page 122):

$$P(Y > X) \approx 1 - \Phi(1.13) \approx 1 - 0.87076 \approx 0.12924.$$

Historical Note #3 (The normal distribution and the central limit theorem)

The normal distribution was first introduced (rather obscurely) by Abraham de Moivre in 1733 to study the limit of binomial distributions. It was Pierre-Simon Laplace who, in his work *Théorie analytique des probabilités* of 1812, proved an improved version of Bernoulli's golden theorem (see page 87):

$$\forall x \in \mathbb{R}, \lim_{n \to +\infty} P\left(\frac{S_n - np}{\sqrt{npq}} < x\right) = \frac{1}{2\pi} \int_{-\infty}^{x} e^{-y^2/2} dy.$$

This theorem shows that, for big numbers, the *discrete* Bernouilli trial can be described by a *continuous* function that will become a key concept in the theory: the *normal or Gaussian distribution* (thus called because of the use the German mathematician Carl Friedrich Gauss made of it).

Over half a century later (in 1887), the Russian school lead by Pafnuty Chebyshev succeeds in rigorously stating and proving *the central limit theorem*, a generalisation of Laplace's result which had been suggested without proof by Laplace himself. This theorem might be the most important result in probability theory:

Central limit theorem Let $\{X_n\}_{n \in \mathbb{N}}$ be a sequence of identical and independent random variables of expected value μ and standard deviation σ. Then,

$$\lim_{n \to +\infty} \frac{X_1 + \ldots + X_n}{n} \sim \mathcal{N}(\mu, \frac{\sigma^2}{n}).$$

In other words, *the normal distribution is a good approximation of the average of any random experiment whatsoever that repeats many times*! This is the profound reason behind the omnipresence of the normal distribution.

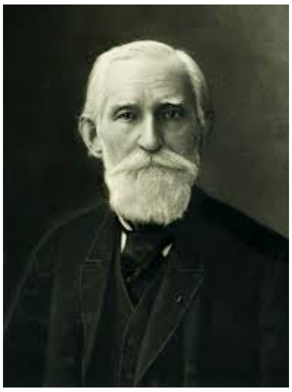
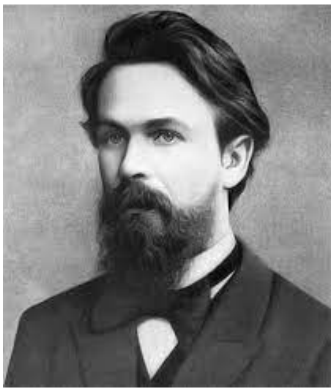
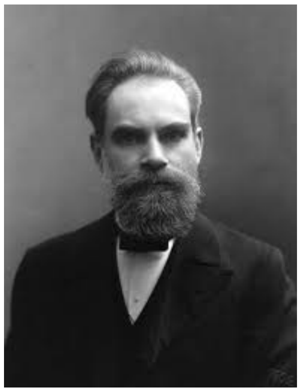




Figure 5.12 – The Russian school of Saint-Petersburg that proved the central limit theorem: Pafnuty Chebyshev and his two students Andrey Markov and Aleksandr Lyapunov.

Interpretation of the central limit theorem: Let us take a population with any distribution, for example with two peaks. Suppose, for example, that this distribution represents the weight of each individual in the population. Let us randomly draw samples containing a fixed number of individuals within the population. The mean weights calculated for each sample will then follow a Gaussian distribution with parameter $\mathcal{N}(\mu, \frac{\sigma^2}{n})$ where σ and μ are the standard deviation and the mean weight in the total population.

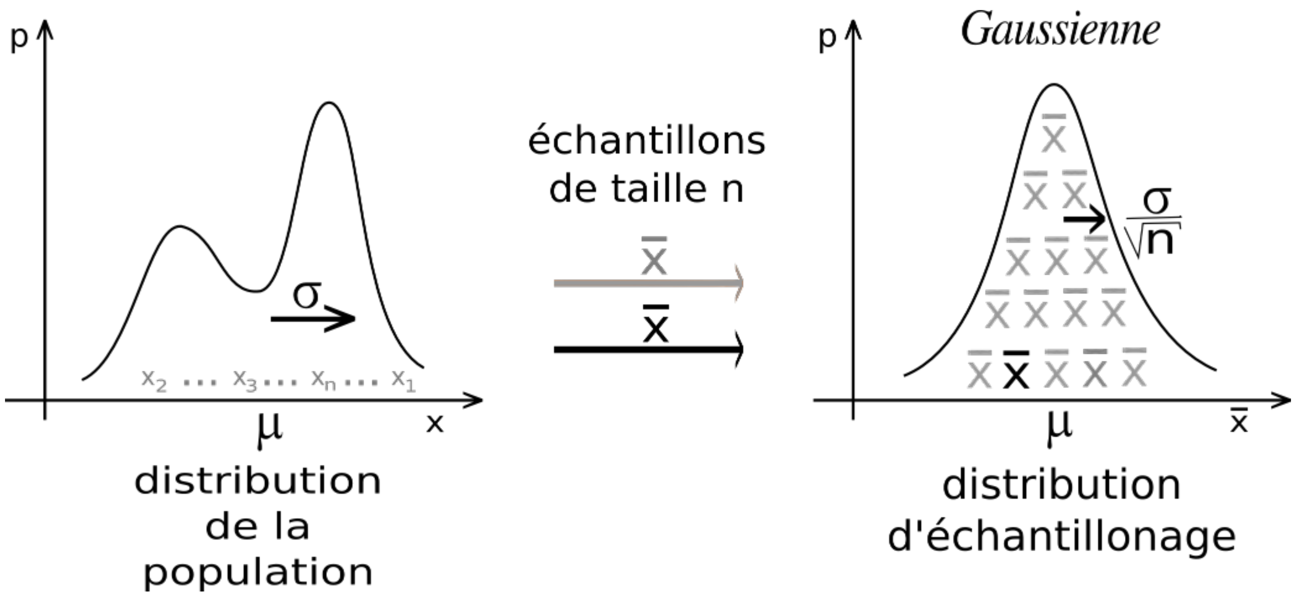


Figure 5.13 – The population's distribution with two peaks is shown in the left. The right plot represents the average $\bar{x}$ computed over samples of a fixed size n.

Figure taken from the book "Calcul d'incertitudes" of Mathieu Rouaud.