THE CENTRAL LIMIT THEOREM

OLIVIER CASTÉRA

Abstract. The central limit theorem states that the normal distribution appears whenever the observed random quantity can be presented as the sum of a sufficiently large number of independent or weakly linked elementary components, each of which has little influence on the sum.

Contents

1 Random Variable. Expectation. Moments. Variance.

Definition 1.1. We call a discrete random variable X a function that associates to each outcome of an experiment an unknown value x_i in advance, from a set of possible values $x_1, \dots, x_n$.

Definition 1.2. We call a continuous random variable X a function that associates to each outcome of an experiment an unknown value x in advance, over a finite or infinite set of possible values $[a, b]$.

Date: December 17, 2025.

Definition 1.3. We call the mathematical expectation or mean value of the discrete random variable X the sum of the products of all possible values x_i of this variable by the probabilities p_i of these values. It is denoted $E[X]$ or $M[X]$ or m_x or also $\bar{x}$. We have

$$E[X] \triangleq \sum_{i=1}^{n} x_i p_i$$

In the case of a continuous random variable X, we have

$$E[X] \triangleq \int_{-\infty}^{+\infty} x f(x) dx$$

Definition 1.4. We call the initial moment of order s of a discrete random variable X the function $\alpha_s[X]$ defined by

$$\alpha_s[X] \triangleq \sum_{i=1}^{n} x_i^s \mathcal{P}i$$

Initial meaning that the moment is calculated with respect to the origin of coordinates. In the case of a continuous random variable X, we have

$$\alpha s[X] \triangleq \int_{-\infty}^{+\infty} x^s f(x) dx$$

Therefore, the first-order moment, $\alpha_1[X]$, coincides with the mathematical expectation $E[X]$ of the random variable X

$$\alpha_1[X] \triangleq E[X]$$

The moments characterize the probability law $\mathcal{P}$, by giving its position, its degree of dispersion, and its shape.

Definition 1.5. We call the centered random variable associated with X, denoted $\mathring{X}$, the difference

$$\mathring{X} \triangleq X - E[X]$$

Definition 1.6. We call the centered moment of order s of a discrete random variable X the function $\mu_s[X]$ defined by

$$\mu_s[X] \triangleq \sum_{i=1}^{n} (x_i - m_x)^s \mathcal{P}i$$

In the case of a continuous random variable X, we have

$$\mu s[X] \triangleq \int_{-\infty}^{+\infty} (x_i - m_x)^s f(x) dx$$

Theorem 1.1. *For any random variable, the first-order centered moment is zero.*

Proof.

$$\begin{aligned} E[\mathring{X}] &= E[X - E[X]] \\ &= E[X] - E[X] \\ &= 0 \end{aligned}$$

□

Definition 1.7. We call the variance of the discrete random variable X its second-order centered moment. It is denoted $V[X]$ or V_x or $D[X]$ or D_x.

$$V[X] \triangleq \sum_{i=0}^{n} \mathring{x}_i^2 \, \mathcal{P}i$$

In the case of a continuous random variable X, we have

$$V[X] \triangleq \int_{-\infty}^{+\infty} \mathring{x}_i^2 f(x) dx$$

By its definition, the variance is the mathematical expectation of the square of the centered random variable associated with X.

$$V[X] \triangleq E[\mathring{X}^2]$$

2 Bienaymé-Chebyshev Inequality

Theorem 2.1. *Let X be a random variable with mathematical expectation $E[X]$ and variance $V[X]$. For any positive real α, the probability that X deviates from its mathematical expectation by an amount greater than or equal to α has an upper bound of $V[X]/\alpha^2$:*

$$P(|X - E[X]| \geqslant \alpha) \leqslant \frac{V[X]}{\alpha^2}$$

Proof. First, let's prove it for a discrete random variable X, whose distribution is given by:

$$\frac{x_1 \; x_2 \; x_3 \; \dots \; x_n}{p_1 \; p_2 \; p_3 \; \dots \; p_n}$$

where $p_i = P(X = x_i)$ is the probability that the random variable X takes the value x_i. Let α be a positive real,

$$P(|X - E[X]| \geqslant \alpha) = \sum_{|x_i - E[X]| \geqslant \alpha} p_i$$

Now,

$$\begin{aligned} V[X] &= \sum^{n} i = 1 (x_i - E[X])^2 p_i \\ &= \sum^{n} i = 1 |x_i - E[X]|^2 p_i \end{aligned}$$

The terms of the sum being positive or zero, we have

$$V[X] \geqslant \sum_{|x_i - E[X]| \geqslant \alpha} |x_i - E[X]|^2 p_i$$

and since $|x_i - E[X]| \geqslant \alpha$, we have

$$V[X] \geqslant \sum_{|x_i - E[X]| \geqslant \alpha} \alpha^2 p_i$$

$$\sum_{|x_i - E[X]| \geqslant \alpha} p_i \leqslant \frac{V[X]}{\alpha^2}$$

$$P(|x - E[X]| \geqslant \alpha) \leqslant \frac{V[X]}{\alpha^2}$$

□

Proof. In the case where X is continuous, let $f(x)$ be the probability density of X:

$$P(|X - E[X]| \geqslant \alpha) = \int_{|x_i - E[X]| \geqslant \alpha} f(x)dx$$

Now,

$$\begin{aligned} V[X] &= \int_{-\infty}^{+\infty} (x_i - E[X])^2 f(x)dx \\ &= \int_{-\infty}^{+\infty} |x_i - E[X]|^2 f(x)dx \end{aligned}$$

The terms of the integral being positive or zero, we have

$$V[X] \geqslant \int_{|x_i - E[X]| \geqslant \alpha} |x_i - E[X]|^2 f(x)dx$$

and since $|x_i - E[X]| \geqslant \alpha$, we have

$$\begin{aligned} V[X] &\geqslant \int_{|x_i - E[X]| \geqslant \alpha} \alpha^2 f(x)dx \\ \int_{|x_i - E[X]| \geqslant \alpha} f(x)dx &\leqslant \frac{V[X]}{\alpha^2} \\ P(|X - E[X]| \geqslant \alpha) &\leqslant \frac{V[X]}{\alpha^2} \end{aligned}$$

□

3 Functions of Random Variables

Let X be a discrete random variable with the following distribution table

x_1	x_2	x_3	$\ldots$	x_n
p_1	p_2	p_3	$\ldots$	p_n

Let Z be a function of the random variable X

$$Z = \varphi(X)$$

its distribution table is then the following

$\varphi(x_1)$	$\varphi(x_2)$	$\varphi(x_3)$	$\ldots$	$\varphi(x_n)$
p_1	p_2	p_3	$\ldots$	p_n

Its mathematical expectation $E[Z]$ is given by

$$\begin{aligned} E[Z] &= E[\varphi(X)] \\ &= \sum_{i=1}^{n} \varphi(x_i) p_i \end{aligned}$$

When the random variable X is continuous, we have

$$E[Z] = \int_{-\infty}^{+\infty} \varphi(x) f(x)dx$$

When Z is a function of two random variables, X and Y

$$Z = \varphi(X, Y)$$

its distribution table is the following

$$\begin{array}{ccccc}
\varphi(x_1,y_1) & \varphi(x_1,y_2) & \varphi(x_1,y_3) & \dots & \varphi(x_1,y_n) \\ \hline
p_{11} & p_{12} & p_{13} & \dots & p_{1n} \\
\varphi(x_2,y_1) & \varphi(x_2,y_2) & \varphi(x_2,y_3) & \dots & \varphi(x_2,y_n) \\ \hline
p_{21} & p_{22} & p_{23} & \dots & p_{2n} \\
\varphi(x_3,y_1) & \varphi(x_3,y_2) & \varphi(x_3,y_3) & \dots & \varphi(x_3,y_n) \\ \hline
p_{31} & p_{32} & p_{33} & \dots & p_{3n} \\
\dots & & & & \\
\varphi(x_n,y_1) & \varphi(x_n,y_2) & \varphi(x_n,y_3) & \dots & \varphi(x_n,y_n) \\ \hline
p_{n1} & p_{n2} & p_{n3} & \dots & p_{nn}
\end{array}$$

where $p_{ij} = P[(X = x_i)(Y = y_j)]$ is the probability that the system (X, Y) takes the value (x_i, y_j). Its mathematical expectation $E[Z]$ is given by

$$\begin{aligned} E[Z] &= E[\varphi(X,Y)] \\ &= \sum_{i=1}^{n}\sum_{j=1}^{n}\varphi(x_i,y_j)p_{ij} \end{aligned} \tag{1}$$

When the random variables X and Y are continuous, we have

$$E[Z] = \iint_{-\infty}^{+\infty}\varphi(x,y)f(x,y)dxdy \tag{2}$$

3.1 Sum of Two Random Variables

Theorem 3.1. *The mathematical expectation of the sum of two random variables is equal to the sum of their mathematical expectations.*

$$E[X+Y] = E[X] + E[Y]$$

Proof. Let X and Y be two discrete random variables. With equation (1) we have

$$\begin{aligned} E[X+Y] &= \sum_{i=1}^{n}\sum_{j=1}^{n}(x_i+y_j)p_{ij} \\ &= \sum_{i=1}^{n}\sum_{j=1}^{n}x_ip_{ij} + \sum_{i=1}^{n}\sum_{j=1}^{n}y_jp_{ij} \\ &= \sum_{i=1}^{n}\left(x_i\sum_{j=1}^{n}p_{ij}\right) + \sum_{j=1}^{n}\left(y_j\sum_{i=1}^{n}p_{ij}\right) \\ &= \sum_{i=1}^{n}x_ip_i + \sum_{j=1}^{n}y_jp_j \\ &= E[X] + E[Y] \end{aligned}$$

□

Proof. Now let X and Y be two continuous variables. With equation (2) we have

$$\begin{aligned} E[X+Y] &= \iint_{-\infty}^{+\infty} (x+y) f(x,y) dx dy \\ &= \iint_{-\infty}^{+\infty} x f(x,y) dx dy + \iint_{-\infty}^{+\infty} y f(x,y) dx dy \\ &= \int_{-\infty}^{+\infty} x \left(\int_{-\infty}^{+\infty} f(x,y) dy \right) dx + \int_{-\infty}^{+\infty} y \left(\int_{-\infty}^{+\infty} f(x,y) dx \right) dy \\ &= \int_{-\infty}^{+\infty} x f_1(x) dx + \int_{-\infty}^{+\infty} y f_2(y) dy \\ &= E[X] + E[Y] \end{aligned}$$

□

Definition 3.1. We call the covariance of the random variables X and Y the mathematical expectation of the product of the two centered random variables $\mathring{X}$ and $\mathring{Y}$ associated with X and Y. It is denoted K_{xy}

$$K_{xy} \triangleq E[\mathring{X}\mathring{Y}]$$

Theorem 3.2. *The variance of the sum of two random variables is equal to the sum of their variances plus twice the covariance.*

$$V[X+Y] = V[X] + V[Y] + 2K_{xy}$$

Proof. Let the random variable Z be such that

$$Z = X + Y$$

Using theorem 3.1, we have

$$\begin{aligned} E[Z] &= E[X+Y] \\ &= E[X] + E[Y] \end{aligned}$$

Thus

$$\begin{aligned} Z - E[Z] &= X - E[X] + Y - E[Y] \\ \mathring{Z} &= \mathring{X} + \mathring{Y} \end{aligned}$$

By definition of variance

$$\begin{aligned} V[Z] &= E[\mathring{Z}^2] \\ &= E[(\mathring{X} + \mathring{Y})^2] \\ &= E[\mathring{X}^2 + \mathring{Y}^2 + 2\mathring{X}\mathring{Y}] \\ &= E[\mathring{X}^2] + E[\mathring{Y}^2] + 2E[\mathring{X}\mathring{Y}] \\ &= V[X] + V[Y] + 2K_{xy} \end{aligned}$$

□

Definition 3.2. The random variables X and Y are said to be independent if the distribution law of each does not depend on the values taken by the other.
The probability of a system of discrete independent random variables is equal to the product of the probabilities of the variables in the system

$$\begin{aligned} P_{ij} &= P[(X = x_i)(Y = y_j)] \\ &= P[(X = x_i)]P[(Y = y_j)] \\ &= P_i P_j \end{aligned}$$

The probability density of a system of continuous independent random variables is equal to the product of the densities of the variables in the system

$$f(x, y) = f_1(x) f_2(y)$$

Theorem 3.3. *The covariance K_{xy} of independent random variables X and Y is zero.*

Proof. Consider discrete random variables

$$K_{xy} = \sum_{i=1}^{n} \sum_{j=1}^{n} (x_i - m_x)(y_j - m_y) P_{ij}$$

From definition (3.2) of independent variables

$$P_{ij} = P_i P_j$$

Thus,

$$K_{xy} = \sum_{i=1}^{n} (x_i - m_x) P_i \sum_{j=1}^{n} (y_j - m_y) P_j$$

These sums are respectively the first-order centered moment of X and Y, which are zero from theorem 1.1, and

$$K_{xy} = 0$$ □

Proof. Now consider continuous random variables

$$K_{xy} = \iint_{-\infty}^{+\infty} (x - m_x)(y - m_y) f(x, y) dx dy$$

From definition (3.2) of independent variables

$$f(x, y) = f_1(x) f_2(y)$$

Thus,

$$K_{xy} = \int_{-\infty}^{+\infty} (x - m_x) f_1(x) dx \int_{-\infty}^{+\infty} (y - m_y) f_2(y) dy$$

These integrals are respectively the first-order centered moment of X and Y, which are zero from theorem 1.1, and

$$K_{xy} = 0$$ □

3.2 Product of Two Random Variables

Theorem 3.4. *The mathematical expectation of the product of two random variables X and Y is equal to the product of their expectations plus the covariance*

$$E[XY] = E[X]E[Y] + K_{xy}$$

Proof. From definition (3.1) of covariance

$$\begin{aligned} K_{xy} &= E[\mathring{X}\mathring{Y}] \\ &= E[(X - E[X])(Y - E[Y])] \\ &= E[XY - XE[Y] - E[X]Y + E[X]E[Y]] \\ &= E[XY] - E[XE[Y]] - E[E[X]Y] + E[E[X]E[Y]] \\ &= E[XY] - E[X]E[Y] - E[X]E[Y] + E[X]E[Y] \\ &= E[XY] - E[X]E[Y] \end{aligned}$$

□

4 Central Limit Theorem

4.1 Characteristic Function

Definition 4.1. We call the characteristic function of the discrete random variable X the mathematical expectation of the function $\exp(itX)$

$$\begin{aligned} g(t) &= \sum_{k=1}^{n} e^{itx_k} p_k \\ &= E\left[e^{itX}\right] \end{aligned}$$

We call the characteristic function of the continuous random variable X the Fourier transform of its probability density $f(x)$

$$\begin{aligned} g(t) &= \int_{-\infty}^{+\infty} e^{itx} f(x) dx \\ &= E\left[e^{itX}\right] \end{aligned}$$

Theorem 4.1. *If two random variables X and Y are related by the relation*

$$Y = aX$$

where a is a non-random factor, their characteristic functions are related by the relation

$$g_y(t) = g_x(at)$$

Proof.

$$\begin{aligned} g_y(t) &= E\left[e^{itY}\right] \\ &= E\left[e^{itaX}\right] \\ &= E\left[e^{i(at)X}\right] \\ &= g_x(at) \end{aligned}$$

□

Theorem 4.2. *The characteristic function of a sum of independent random variables is equal to the product of the characteristic functions of the components.*

Proof. Let $X_1, X_2, \ldots, X_n$ be independent random variables with $g_{x_1}, g_{x_2}, \ldots, g_{x_n}$ for characteristic functions, and sum

$$Y = \sum_{k=1}^{n} X_k$$

The characteristic function of Y is written

$$\begin{aligned} g_y(t) &= E\left[e^{itY}\right] \\ &= E\left[e^{it\sum_{k=1}^{n} X_k}\right] \\ &= E\left[\prod_{k=1}^{n} e^{itX_k}\right] \end{aligned}$$

The random variables X_k being independent, the functions e^{itX_k} are also independent and we can use theorem 3.4

$$\begin{aligned} g_y(t) &= \prod_{k=1}^{n} E\left[e^{itX_k}\right] \\ &= \prod_{k=1}^{n} g_{x_k}(t) \end{aligned}$$

□

4.2 Characteristic Function of the Normal Distribution

Theorem 4.3. *The characteristic function of the normal distribution with mathematical expectation m_x zero and variance σ^2 is written*

$$g(t) = e^{-t^2\sigma^2/2}$$

Proof. Let a continuous random variable X be distributed according to the normal distribution, with mathematical expectation m_x zero, and variance σ^2. Its probability density $f(x)$ is written

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

Its characteristic function is written

$$\begin{aligned} g(t) &= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{itx} e^{-x^2/(2\sigma^2)} dx \\ &= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{itx - x^2/(2\sigma^2)} dx \end{aligned}$$

We use the following formula

$$\int_{-\infty}^{+\infty} e^{-Ax^2 \pm 2Bx - C}, dx = \sqrt{\frac{\pi}{A}}, e^{-(AC-B^2)/A}$$

We identify: $A = 1/(2\sigma^2)$, $B = it/2$, and $C = 0$. Hence

$$\begin{aligned} g(t) &= \frac{1}{\sigma\sqrt{2\pi}} \sqrt{2\pi\sigma^2}\, e^{-(t^2/4)2\sigma^2} \\ &= e^{-t^2\sigma^2/2} \end{aligned}$$

□

4.3 Sum of Two Normal Distributions

Theorem 4.4. *The sum of two normal distributions is a normal distribution.*

Proof. Let X and Y be two independent continuous random variables whose probability densities are distributed according to normal distributions

$$f(x) = \frac{1}{\sigma_x\sqrt{2\pi}}\, e^{-x^2/(2\sigma_x^2)}$$
$$f(y) = \frac{1}{\sigma_y\sqrt{2\pi}}\, e^{-y^2/(2\sigma_y^2)}$$

From definition 4.1, their characteristic functions are written

$$g_x(t) = e^{-t^2\sigma_x^2/2}$$
$$g_y(t) = e^{-t^2\sigma_y^2/2}$$

From theorem 4.2, the characteristic function $g_z(t)$ of their sum is equal to the product of the characteristic functions

$$\begin{aligned} g_z(t) &= g_x(t) \times g_y(t) \\ &= e^{-t^2\sigma_x^2/2} \times e^{-t^2\sigma_y^2/2} \\ &= e^{-t^2(\sigma_x^2+\sigma_y^2)/2} \end{aligned}$$

which, from theorem 4.3, is the characteristic function of a normal distribution with mathematical expectation m_z zero and variance $\sigma_x^2 + \sigma_y^2$. □

4.4 Central Limit Theorem

Theorem 4.5. *Let $X_1, X_2, \ldots, X_n$ be independent random variables with the same distribution law, with mathematical expectation m_x and variance σ^2. As n increases indefinitely, the distribution law of the sum*

$$Y_n = \frac{1}{\sqrt{n}} \sum_{k=1}^{n} X_k$$

tends toward the normal distribution.

Proof. Note that the factor $1/\sqrt{n}$ is necessary for the proof, but the convergence of Y_n toward the normal distribution implies that of the sum of the random variables $\sum_{k=1}^{n} X_k$. We will use the characteristic functions and admit without proof that the convergence of the characteristic functions implies that of the laws. From theorem 4.1, the characteristic function of Y_n is equal to

$$\begin{aligned} g_{y_n}(t) &= g_{1/\sqrt{n}\sum_{k=1}^{n} X_k} \\ &= g_{\sum_{k=1}^{n} X_k}\left(\frac{t}{\sqrt{n}}\right) \end{aligned}$$

From theorem 4.2, the characteristic function of Y_n is equal to the product of the characteristic functions of the X_k

$$g_{y_n}(t) = \left[g_x\left(\frac{t}{\sqrt{n}}\right)\right]^n \tag{3}$$

The characteristic function of each X_k is written

$$g_x\left(\frac{t}{\sqrt{n}}\right) = \int_{-\infty}^{+\infty} e^{ixt/\sqrt{n}} f(x)dx$$

Perform a Taylor expansion around zero, up to order 3

$$g_x\left(\frac{t}{\sqrt{n}}\right) = g_x(0) + g_x'(0)\left(\frac{t}{\sqrt{n}}\right) + \left[\frac{g_x''(0)}{2} + o\left(\frac{t}{\sqrt{n}}\right)\right]\frac{t^2}{n} \tag{4}$$

$$\text{where } \lim_{t/\sqrt{n}\to 0} o\left(\frac{t}{\sqrt{n}}\right) = 0$$

Calculate each term

$$\begin{aligned} g_x(0) &= \int_{-\infty}^{+\infty} f(x)dx \\ &= 1 \end{aligned}$$

By differentiating we have

$$\begin{aligned} g_x'\left(\frac{t}{\sqrt{n}}\right) &= \int_{-\infty}^{+\infty} ixe^{ixt/\sqrt{n}} f(x)dx \\ &= i\int_{-\infty}^{+\infty} xe^{ixt/\sqrt{n}} f(x)dx \end{aligned}$$

setting $t/\sqrt{n} = 0$ we have

$$\begin{aligned} g_x'(0) &= i\int_{-\infty}^{+\infty} xf(x)dx \\ &= im_x \end{aligned}$$

We do not restrict generality by shifting the origin to the point m_x, then we have $m_x = 0$, and

$$g_x'(0) = 0$$

Differentiating a second time

$$g_x''\left(\frac{t}{\sqrt{n}}\right) = -\int_{-\infty}^{+\infty} x^2 e^{ixt/\sqrt{n}} f(x)dx$$

setting $t/\sqrt{n} = 0$

$$g_x''(0) = -\int_{-\infty}^{+\infty} x^2 f(x)dx$$

and since we set $E[X] = 0$, from definition 1.7 the previous expression is nothing other than the negative of the variance

$$g_x''(0) = -\sigma^2$$

Substituting these results into equation (4), when $t/\sqrt{n} \to 0$

$$g_x\left(\frac{t}{\sqrt{n}}\right) = 1 - \left[\frac{\sigma^2}{2} - o\left(\frac{t}{\sqrt{n}}\right)\right]\frac{t^2}{n}$$

then into equation (3)

$$g_{y_n}(t) = \left\{1 - \left[\frac{\sigma^2}{2} - o\left(\frac{t}{\sqrt{n}}\right)\right]\frac{t^2}{n}\right\}^n$$

Take the logarithm of this expression

$$\ln g_{y_n}(t) = n\ln\left\{1 - \left[\frac{\sigma^2}{2} - o\left(\frac{t}{\sqrt{n}}\right)\right]\frac{t^2}{n}\right\}$$

Set

$$\chi = \left[\frac{\sigma^2}{2} - o\left(\frac{t}{\sqrt{n}}\right)\right]\frac{t^2}{n}$$

and we have

$$\ln g_{y_n}(t) = n \ln\{1-\chi\}$$

As n tends to infinity, χ tends to 0, and $\ln(1-\chi)$ tends to $-\chi$

$$\begin{aligned}\lim_{n\to+\infty} \ln g_{y_n}(t) &= \lim_{n\to+\infty} n(-\chi) \\ &= \lim_{n\to+\infty} n\left[-\frac{\sigma^2}{2} + o\left(\frac{t}{\sqrt{n}}\right)\right]\frac{t^2}{n} \\ &= -\frac{t^2\sigma^2}{2} + \lim_{n\to+\infty} t^2, o\left(\frac{t}{\sqrt{n}}\right)\end{aligned}$$

Now $\lim_{t/\sqrt{n}\to 0} o(t) = 0$, so

$$\lim_{n\to+\infty} o\left(\frac{t}{\sqrt{n}}\right) = 0$$

thus

$$\lim_{n\to+\infty} \ln g_{y_n}(t) = -\frac{t^2\sigma^2}{2}$$

hence

$$\lim_{n\to+\infty} g_{y_n}(t) = e^{-t^2\sigma^2/2}$$

which, from theorem 4.3, is the characteristic function of the normal distribution with mathematical expectation zero, and variance σ^2. □

5 Laws of Large Numbers

The laws of large numbers assert the convergence in probability of random variables toward constant quantities, as the number of experiments increases indefinitely.

5.1 Chebyshev's Theorem

Let a random variable X provide, during n independent experiments, the values $x_1, x_2, \ldots, x_n$. To use theorem 3.1, we will consider that these values are the results of a single experiment performed on each of the n independent random variables $X_1, X_2, \ldots, X_n$, with the same distribution law as the random variable X.
Rather than considering the sum of these random variables, we will consider the arithmetic mean of these variables, in other words the sum divided by n. Let the random variable Y be defined as the arithmetic mean of these n random variables

$$Y = \frac{1}{n}\sum_{i=1}^{n} X_i$$

Its mathematical expectation is written

$$\begin{aligned}E[Y] &= E\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] \\ &= \frac{1}{n} E\left[\sum_{i=1}^{n} X_i\right]\end{aligned}$$

With theorem 3.1 we have

$$\begin{aligned} E[Y] &= \frac{1}{n}\sum_{i=1}^{n} E[X_i] \\ &= \frac{1}{n}\, n\, E[X] \\ &= E[X] \end{aligned}$$

The variance of Y is written

$$\begin{aligned} V[Y] &= E[\mathring{Y}^2] \\ &= E\left[\left(\frac{1}{n}\sum_{i=1}^{n} \mathring{X}_i\right)^2\right] \\ &= \frac{1}{n^2}\, E\left[\left(\sum_{i=1}^{n} \mathring{X}_i\right)^2\right] \\ &= \frac{1}{n^2}\, V\left[\sum_{i=1}^{n} X_i\right] \end{aligned}$$

With theorem 3.2 and theorem 3.3 we have

$$\begin{aligned} V[Y] &= \frac{1}{n^2}\sum_{i=1}^{n} V[X_i] \\ &= \frac{1}{n}\, V[X] \end{aligned}$$

Theorem 5.1. *For a sufficiently large number of experiments, the arithmetic mean of the observed values of a random variable converges in probability toward its mathematical expectation*

$$\forall \delta > 0,\ \forall \varepsilon > 0,\ \exists n \text{ such that } P\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - E[X]\right| \geqslant \varepsilon\right) < \delta$$

Proof. Apply to Y the Bienaymé-Chebyshev inequality by setting $\alpha = \varepsilon$

$$P(|Y - E[Y]| \geqslant \varepsilon) \leqslant \frac{V[Y]}{\varepsilon^2}$$

Applying the previous results on the expectation and variance of Y

$$P\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - E[X]\right| \geqslant \varepsilon\right) \leqslant \frac{V[X]}{n\varepsilon^2}$$

and set $\delta = V[X]/(n\varepsilon^2)$. We can always choose n sufficiently large to have the above inequality, no matter how small ε is. □

5.2 Generalized Chebyshev Theorem

We can generalize the law of large numbers to the case of random variables with different distribution laws.

Theorem 5.2. *Let there be n independent random variables X_1, X_2, $\ldots$, X_n, with different distribution laws, with mathematical expectations $E[X_1]$, $E[X_2]$, $\ldots$, $E[X_n]$, and variances $V[X_1], V[X_2], \ldots, V[X_n]$. If all variances have the same upper bound L*

$$\forall i,\ V[X_i] < L$$

for a sufficiently large number of experiments, the arithmetic mean of the observed values of the random variables $X_1, X_2, \ldots, X_n$, converges in probability toward the arithmetic mean of their mathematical expectations

$$\forall \delta > 0, \forall \varepsilon > 0,\ \exists n \text{ such that} \quad P\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n} E[X_i]\right| \geqslant \varepsilon\right) < \delta$$

Proof. Let the random variable Y be such that

$$Y = \frac{1}{n}\sum_{i=1}^{n} X_i$$

Apply to Y the Chebyshev inequality:

$$\forall \delta > 0, \forall \varepsilon > 0,\ \exists n \text{ such that} \quad P\left(|Y - E[Y]| \geqslant \varepsilon\right) < \frac{V[Y]}{\varepsilon^2}$$

$$P\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n} E[X_i]\right| \geqslant \varepsilon\right) < \frac{\sum_{i=1}^{n} V[X_i]}{n^2\varepsilon^2}$$

$$P\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n} E[X_i]\right| \geqslant \varepsilon\right) < \frac{L}{n\varepsilon^2}$$

and set $\delta = L/(n\varepsilon^2)$. We can always choose n sufficiently large to have the above inequality, no matter how small ε is. ☐

5.3 Markov's Theorem

We can generalize the law of large numbers to the case of dependent random variables.

Theorem 5.3. *Markov's Theorem Let there be n dependent random variables X_1, X_2, $\ldots$, X_n, with different distribution laws, with mathematical expectations $E[X_1], E[X_2], \ldots, E[X_n]$, and variances $V[X_1], V[X_2], \ldots, V[X_n]$. If*

$$\lim_{n\to\infty} \frac{1}{n^2}\sum_{i=1}^{n} V[X_i] = 0$$

the arithmetic mean of the observed values of the random variables $X_1, X_2, \ldots, X_n$, converges in probability toward the arithmetic mean of their mathematical expectations

$$\forall \delta > 0, \forall \varepsilon > 0,\ \exists n \text{ such that} \quad P\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n} E[X_i]\right| \geqslant \varepsilon\right) < \delta$$

Proof. Let the random variable Y be such that

$$Y = \frac{1}{n}\sum_{i=1}^{n} X_i$$

Apply to Y the Chebyshev inequality:

$$\forall\delta > 0,\, \forall\varepsilon > 0,\; \exists n \text{ such that} \quad P\left(|Y - E[Y]| \geqslant \varepsilon\right) < \frac{V[Y]}{\varepsilon^2}$$
$$P\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n} E[X_i]\right| \geqslant \varepsilon\right) < \frac{\sum_{i=1}^{n} V[X_i]}{n^2\varepsilon^2}$$

Set $\delta = \dfrac{1}{n^2\varepsilon^2}\sum_{i=1}^{n} V[X_i]$. By hypothesis, we can always choose n sufficiently large to have $\lim_{n\to\infty}\delta = 0$, no matter how small ε is. □

Email address : o.castera@free.fr
Website : https://sciences-physiques.neocities.org