الجمهورية الجزائرية الديمقراطية الشعبية
PEOPLE'S DEMOCRATIC REPUBLIC OF ALGERIA
وزارة التعليم العالي و البحث العلمي
MINISTRY OF HIGHER EDUCATION AND SCIENTIFIC RESEARCH
جامعة عباس لغرور خنشلة
ABBES LAGHROUR- KHENCHELA UNIVERSITY

Faculty of Sciences and Technology

Department of Mathematics and Computer Science

Serial number: ……………

THESIS Submitted for
obtain **Master** degree in mathematics

Field of study: **Mathematics**
Specialty: **Applied Mathematics**

Entitled by:

The Central Limit Theorem

Presented by**: Nouria MEDDOUR**

Directed by: **Dr. F. MERAHI**

Jury members:

Mr. S. BRAHIMI **President**

Mr. A. GUEMMAZ **Examiner**

2020-2021

The Central Limit Theorem

Meddour Nouria

June 2021

Contents

Dedications

This thesis is dedicated to:

The sake of Allah, my Creator and my Master,

My great teacher and messenger, Mohammed (May Allah bless and grant him), who taught us the purpose of life,

The **Abbes Laghrour University**; my second magnificent home,

My great parents, who never stop giving of themselves in countless ways,

My dear aunt Abla, who was my support and she is the light that precedes my steps and that leads me to success and excellence ,

My beloved brothers , who stand by me when things look bleak,

My friends who encourage and support me,

All the people in my life who touch my heart,

I dedicate this research.

Acknowledgements

All praises to Allah and his blessings for the completion of this thesis .

First and foremost , I would like to sincerely thank my supervisor **Dr.F. MERAHI** for his guidance understanding ,patience and most importantly , he has provided positive encouragement to finish this thesis .

It has been a great pleasure and honor to have him as my supervisor .

I also thank Dr.S. BRAHIMI and Dr.A. GUEMMAZ, members of the jury, for having done us the honor of accepting to evaluate this work .

My deepest gratitude goes to all of my family members for their great support and love .

I offer my special thanks to everyone who knows me and helped me even morally with a kind word .

In the end, I wish to acknowledge the University of Khenchela which hosted me throughout my master studies .

List of Figures

Introduction

in probability theory , the central limit theorem (CLT) establish that, in many situations, when independent random variables are added , their properly normalized sum tends toward a normal distribution even if the original variables themselves are not normally distributed . the theorem is a key concept in probability theory because it implies that probabilistic and statistical methods that work for normal distributions can be applicate to many problems involving others types of distributions.

if $X_{1,}X_{2,...,}X_n$ are n random samples drawn from a population with overall mean μ and finite variance σ^2 ,and if $\bar{X}_n$ is the sample mean , the limiting form of the distribution, $Z = \lim_{n\to\infty} \sqrt{n}\left(\frac{\bar{X}_n - \mu}{\sigma}\right)$ is the standard normal distribution .

for example , suppose that a sample is obtained containing many observations, each observation being randomly generated in a way that does not depend on the values of the other observations, and that the arithmetic mean of the observed values is computed . if this procedure is performed many times , the central limit theorem says that the probability distribution of the average will closely approximate a normal distribution . a simple example of this is that if one flips a coin many times , the probability of getting a given number of heads will approach a normal distribution , with the mean equal to half the total number of flips . at the limit of an infinite number of flips , it will equal a normal distribution .

the central limit theorem has several variants . in its common form , the random variables must be identically distributed . in variants , convergence of the mean to the normal distribution also occurs for non-identical distributions or for non-independent observations, if they comply with certain conditions.

Chapter 1

The Proof of Central Limit Theorem

1.1 Random Variable

Definition 1 *a random variable is a function X that assigns a rule of correspondence for every point ξ in the sample space S (called the domain) a unique real value $X(\xi)$.*

the rule of correspondence is given either by a probability mass function or the probability density function, depending on the type of random variable considred .

1.2 Probability Mass Function

Definition 2 *for a random variable that can take on at most a countable number of possible values, a probability mass function*

$p(a)$ *is defined by*

$$p(a) = P(X = a)$$

1.3 Probability Density Function

Definition 3 *for a random variable X that is continuously defined , a probability density function $f(x)$ is defined such that for a subset $B \in \mathbb{R}$*

$$P(X \in B) = \int_B f(x)dx$$

we might also be interested in the probability that the random variable is less than some value . for such cases we define the distribution function .

1.4 Distribution Function

Definition 4 *for a random variable X , the distribution function F is defined by $F(x) = P(X \leq x)$ for continuous random variables, the above equation can be represented as $F(x) = \int_{-\infty}^{\infty} f(t)dt$ where $f(t)$ is a probability density function .*

1.5 Normal Random Variable

Definition 5 *X is a normal random variable with parameters μ and σ^2 if the density of X is given by $f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2}$ when ever $\mu = 0$ and $\sigma^2 = 1$ we get a simplified equation :*

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

we can see that $f(x)$ *is indeed a distribution function since integrating it from* $-\infty$ *to* $+\infty$ *gives* 1 *and hence the sample space has probability* 1 *, as required by the second axiom of probability . having defined the random variable , we are now interested in its properties . we can discribe the whole distribution of probabilities through two qualities of a random variable : its average value and spread . these terms are called expected value and variance , respectively .*

1.6 Expected Value or Mean

Definition 6 *if* X *is a discrete random variable having the probability mass function* $p(x)$ *, the expected value , denoted by* $E[X]$ *, is defined as*

$$E[X] = \sum_{x:p(x)>0} xp(x)$$

if X *is a continuous random variable having the probability density function* $f(x)$ *, the expected value is defined as*

$$E[X] = \int_{-\infty}^{\infty} xf(x)dx$$

1.7 Variance

Definition 7 *if* X *is a random variable with mean* μ *, where* $\mu = E[X]$*, then the variance of* X *, denoted by* $Var(X)$ *,is defined as* $Var(X) = E\left[(X-\mu)^2\right]$

1.8 Sample Mean

Definition 8 *let* $X_{1,....,}X_n$ *be independent and identically distributed random variables having distribution function* F *and expected value* μ *. such a sequence constitutes a sample from the distribution* F *. given a sample , we define the sample mean ,*$\hat{X}$ *,as*

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

furthermore, $E\left[\hat{X}\right] = \frac{1}{n}\sum_{i=1}^{n} E\left[X_i\right]$

so we now know how to take the expected value of a random variable , but let's say we were interested in the expected value of the square of the random variable , or the cube , or so on . this brings us to the concept of the moments of a random variable .

1.9 Moments of a Random Variable

Definition 9 *the K-th moment of a random variable* X *is* $E\left[X^k\right] \quad \forall k \in \mathbb{N}$ *.*

to make the computation of the moments of a random variable easier, we define a special moment generating function .

1.10 Moment Generating Function

Definition 10 *the moment generating function* $M(t)$ *of a random variable* X *is defined for all real values of* t *by*

$$M(t) = E\left[e^{tX}\right] = \begin{cases} \sum_x e^{tX} p(x) & \text{if } X \text{ is discrete with mass function } p(x) , \\ \int_{-\infty}^{\infty} e^{tX} f(x)dx & \text{if } X \text{ is continuous with density } f(x) . \end{cases}$$

the moment generating function (MGF) has a few interesting properties which we will need to keep in mind throughout the paper .first , for independent random variables X *and* Y, *the MGF satisfies* $M_{X+Y}(t) = E\left[e^{t(X+Y)}\right] = E\left[e^{tX}.e^{tY}\right] = M_x(t).M_y(t)$ *second , all moments of a random variable can be obtained by differentiating the MGF and evaluating the derivative at* 0 *. more precisely ,*

$M^k(0) = E\left[X^k\right]$

Proof. $\acute{M}(t) = \frac{d}{dt} E\left[e^{tX}\right] = E\left[\frac{d}{dt}(e^{tX})\right] = E\left[Xe^{tX}\right]$ ■

and when we evaluate $\acute{M}(t)$ at $t = 0$, we get :

$$\acute{M}(0) = E\left[X.e^0\right] = E\left[X\right]$$

which is the first moment of the random variable X .

the use of MGFs, and in particular of the MGF of the standard normal distribution , will be the key of the proof of the central limit theorem .

Lemma 11 ***Markov's Inequality***

if X *is a random variable that takes only nonnegative values, then for any value* $a > 0$ *,*

$P(X \geq a) \leq \frac{E[X]}{a}$

Proof. *define a new random variable i such that*

$$I = \left\{ \begin{matrix} 1 & if\ X \geq a \\ 0 & otherwise \end{matrix} \right\}$$

case 1: if $I = 1$ *then* $X \geq a$ *; therfore* $I \leq \frac{X}{a}$ *since* $X \geq 0$ *and* $a \geq 0$ *.*

case 2: if $I = 0$ *then* $X < a$ *but* $\frac{X}{a} \geq 0$ *since* $X \geq 0$ *and* $a > 0$ *.*

thus, the inequality $I \leq \frac{X}{a}$ *holds in all cases .*

next , we take the expected value of both sides, i.e $E[I]$ *and* $E\left[\frac{X}{a}\right]$ *, s.t .*

$$E[I] = \int_{-\infty}^{\infty} p(x).I \qquad\qquad E\left[\frac{X}{a}\right] = \int_{-\infty}^{\infty} p(x).\frac{X}{a}$$

clearly $E[I] \leq E\left[\frac{X}{a}\right]$ *since* $I \leq \frac{X}{a}$

furthermore, $E[I] = P(X \geq a)$ *since* $E[I]$ *is just the sum the probabilities where* $X \geq a$ *.*

hence, $P(X \geq a) \leq \frac{E[X]}{a}$ *.*

using markov's inequality , we can create a bound on the probability of a certain outcome given only the distribution's mean . furthermore , this result is integral to deriving other bounds on the probability distribution of a random variable , such as the bound given by Chebyshev's inequality .

Lemma 12 Chebyshev's Inequality

if X *is finite random variable with finite mean* μ *and variance* σ^2 *,then for any value* $k > 0$ *,*

$$P(|X - \mu| \geq k) \leq \frac{\sigma^2}{k^2}$$

Proof. *from markov's inequality we know that*

$$P(X \geq a) \leq \frac{E[X]}{a}$$

where X *is any random variable that takes on only nonnegative values and* $a > 0$ *.*

we can apply markov' inequality to the random variable $(X-\mu)^2$ *, which satisfies the necessary conditions we get :*

$P((X-\mu)^2 \geq k^2) \leq \frac{E\left[(X-\mu)^2\right]}{k^2}$

next, notice that $(X-\mu^2) \geq k^2 \Longleftrightarrow |X-\mu| \geq k$ *.hence , whenever one of these inequalities is true so is the other inequality . in other words, the probability of either inequality being true is the same thus, we can swap one inequality for the other to get :*

$P(|X-\mu| \geq k) \leq \frac{E\left[(X-\mu)^2\right]}{k^2}$

lastly, by definition of variance , $E\left[(X-\mu)^2\right] = \sigma^2$ *, so we get :*

$P(|X-\mu| \geq k) \leq \frac{\sigma^2}{k^2}$ ■

chebyshev's inequality will be used to prove the weak law of large numbers that states the conditions under which the average of a sequence of random variables converges to the expected average . this result will rely primarily on chebyshev's inequality by allowing the random variable X *to be a sequence of random variables.*

■

1.11 Weak Law of Large Numbers

Theorem 13 *The Weak Law of Large Numbers*

let $X_1, X_2, \ldots$ *be a sequence of independent and identically distributed random variables, each having finite mean* $E[X_i] = \mu$ *and variance* σ^2 *. then, for any* $\varepsilon > 0$ *,*

$P\left(\left|\frac{X_1+\ldots+X_n}{n} - \mu\right| \geq \varepsilon\right) \to 0 \quad as \quad n \to \infty$

Proof. *we see that ,*

$E\left[\frac{X_1+...+X_n}{n}\right]=\frac{1}{n}E\left[X_1\right]+...+E\left[X_n\right]=\frac{n\mu}{n}=\mu$

furthermore,

$Var(\frac{X_1+...+X_n}{n})=Var(\frac{X_1}{n})+...+Var(\frac{X_n}{n})$

we can see that $Var(\frac{X_1}{n})=E\left[(\frac{X_1-\mu}{n})^2\right]=(\frac{1}{n^2})\cdot\ E\left[(X_1-\mu)^2\right]$ *, and therfore we get :*

$Var(\frac{X_1+...+Xn}{n})=\overbrace{\frac{\sigma^2}{n^2}+....+\frac{\sigma^2}{n^2}}=\frac{\sigma^2}{n^2}$

now we treat $(\frac{X_1+...+Xn}{n})$ *as a new random variable* X *. the new random variable* X *clearly satisfies the conditions for chebyshev's inequality .hence , we apply the lemma to get the following :*

$p(\,|X-\mu|\geq\epsilon)\leq\frac{Var(X)}{\epsilon^2}=\frac{\sigma^2}{n\cdot\epsilon^2}$

as $n\to\infty$ *,it then follows that*

$\lim_{n\to\infty}P(\,|X-\mu|\geq\varepsilon)=0$

the weak law of large numbers demonstrates that given a large aggregate of identical random variables, the average of the results obtained will approach the sample mean .

next, i will prove a restricted case of the central limit theorem that deals only with a standard normal random variable. this theorem is concerned with determining the conditions under which the sum of a large number of random variables has a probability distribution that is approximately normal . ■

the following lemma is integral to the proof of the central limit theorem.

Lemma 14 *let* $Z_1,Z_2,....$ *be a sequence of random variables having* *distribution functions* F_{Z_n} *and moment generating functions* M_{Z_n} *s.t .* $n \geq 1$ *. furthermore, let* Z *be a random variable having distribution function* F_Z *and moment generating functions* M_Z *. if* $M_{Z_n}(t) \to M_Z$ *for all* t *, then* $F_{Z_n}(t) \to F_Z(t)$ *for all* t *at which* $F_Z(t)$ *is continuous .*

to see the relevance of this lemma , let's set $Z_n = \sum_{i=1}^{n} \frac{X_i}{\sqrt{n}}$ *where* X_i *are independent and identicaly distributed random variables and let* Z *be a normal random variable . then, if we show that the MGF of* $\sum_{i=1}^{n} \frac{X_i}{\sqrt{n}}$ *approaches the MGF of* Z *(which we previously calculated to be* $e^{t^2/2}$ *) as* $n \to \infty$ *,we simultaneously show that the probability distribution of* $\sum_{i=1}^{n} \frac{X_i}{\sqrt{n}}$ *approaches the normal distribution as* $n \to \infty$ *. this is the method we will use to prove the central limit theorem .*

Theorem 15 *The Central Limit Theorem*

let $X_1, X_2, ...$ *be a sequence of independent and identically distributed random variables each having mean* μ *and variance* σ^2 *.then the distribution of tends to the standard normal as* $n \to \infty$ *. that is, for* $-\infty < a < \infty$ *,*

$$P\left(\frac{X_1 + ... + X_n - n\mu}{\sigma\sqrt{n}} \leq a\right) \longrightarrow \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{a} e^{-x^2/2} dx \qquad as \quad n \to \infty$$

1.12 The proof of Central Limit Theorem

Proof. we begin the the proof with the assumption that $\mu = 0$,$\sigma^2 = 1$ and that the MGF of the X_i exists and is finite .

we already know the MGF of a normal random variable , but we still need to compute the MGF of the sequence of random variables we are interested in : $\sum_{i=0}^{n} \frac{X_i}{\sqrt{n}}$.

by definition of MGF, we can see that $M(\frac{t}{\sqrt{n}}) = E\left[\exp(\frac{tX_i}{\sqrt{n}})\right]$.however, we are interested in the MGF of $E\left[\exp(t\sum_{i=1}^{n} \frac{X_i}{\sqrt{n}})\right]$ here is how we find the MGF :

$$E\left[\exp(t\sum_{i=1}^{n} \frac{X_i}{\sqrt{n}})\right] = E\left[\exp(\sum_{i=1}^{n} X_i \cdot \frac{t}{\sqrt{n}})\right]$$
$$= E\left[\Pi_{i=1}^{n} \exp(X_i \cdot \frac{t}{\sqrt{n}})\right]$$
$$= \Pi_{i=1}^{n} E\left[\exp(X_i \cdot \frac{t}{\sqrt{n}})\right]$$
$$= \Pi_{i=1}^{n} M(\frac{t}{\sqrt{n}})$$
$$= \left[M(\frac{t}{\sqrt{n}})\right]^n$$

step1 is accomplished by simple distribution . step 2 uses the rule $e^{x+y} = e^x \cdot e^y$. steps 3 relies on the fact that the X_is are independent and therefore the E and the Π operators are interchangeable .

in steps 4 i simply substitute a previously calculated identity . and in steps 5 i simply rewrite the equation into a more accessible format .

now we define $L(t) = \log M(t)$ and evaluate $L(0)$, $\acute{L}(0)$, $L"(0)$.

$$L(0) = \log M(0) = \log E\left[e^{0 \cdot X_i}\right] = \log E\left[1\right] = \log 1 = 0$$
$$\acute{L}(0) = \frac{\acute{M}(0)}{M(0)} = \acute{M}(0) = E\left[Xe^{0 \cdot X_i}\right] = E\left[X\right] = \mu = 0$$
$$L"(0) = \frac{M(0)M"(0) - \left[\acute{M}(0)\right]^2}{\left[M(0)\right]^2} = \frac{1 \cdot E\left[X^2\right] - 0^2}{1^2} = E\left[X^2\right] = \sigma^2 = 1$$

now we are ready to prove the central limit theorem by showing that $\left[M(\frac{t}{\sqrt{n}})\right]^n \to e^{t^2/2}$ as $n \to \infty$

by taking the log of both sides, we can see that this is equivalent to showing $nL(\frac{t}{\sqrt{n}}) \to t^2/2$ as $n \to \infty$ hence , we compute :

$\lim_{n\to\infty} \frac{L(\frac{t}{\sqrt{n}})}{n^{-1}} = \lim_{n\to\infty} \frac{-\acute{L}(\frac{t}{\sqrt{n}})n^{-\frac{3}{2}t}}{-2n^{-2}}$

$= \lim_{n\to\infty} \frac{-\acute{L}(\frac{t}{\sqrt{n}})t}{-2n^{-1/2}}$

$= \lim_{n\to\infty} \frac{-L"(\frac{t}{\sqrt{n}})n^{-3/2}t^2}{-2n^{-3/2}}$

$= \lim_{n\to\infty} L"(\frac{t}{\sqrt{n}})\frac{t^2}{2}$

$= \frac{t^2}{2}$

here, step1 was accomplished by L'Hopital's rule since both the top and the bottm of riginal fraction equaled 0 . step 2 simply reduces the fraction . step 3 is again accomplished by L'Hopital's rule since again both the top and the bottom of the reduced fraction equal 0 . step 4 reduces the equation . step 5 uses the previously calculated value of $L"(0)$ to give us the final result . having shown this, we can now apply lemma 3.3 to prove the central limit theorem for the case where $\mu = 0$ and $\sigma^2 = 1$.

i will briefly illustrate the theorem with a simple application from investment . consider an investor who chose a diversified portfolio with 100 stocks . we assume that possible yields of each stock are identically distributed(although in reality such an assumption would be difficult to satisfy)

then,using the central limit theorem, he can model the returns of this portfolio using the normal distribution. ∎

Chapter 2

Generalization of The Central Limit Theorem

2.1 The Central Limit Theorem

The X_k being real random variables defined on the same probability space and independent, we note

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

and we study the convergence in law of

$$S_n^* = \frac{S_n - ES_n}{\sqrt{VarS_n}}$$

when this quantity is defined. We will then consider the case where the X_k are random vectors of $\mathbb{R}^d$.

2.1.1 Case of an I.I.D of sequence

De Moivre- Laplace Theorem

Theorem 16 *If the X_k are independent and have the same Bernoulli law with parameter $p \in]0,1[$, with $q = 1-p$,we have*

$S_n^* = \frac{S_n - np}{\sqrt{npq}} = \sqrt{\frac{n}{pq}} \left(\frac{S_n}{n} - p \right) \to_{n\to\infty}^{law} N(0,1)$

Since the distribution function Φ of $N(0,1)$ is continuous on $\mathbb{R}$, this is equivalent to

$P(S_n^* \leq x) \to_{n\to\infty} \Phi(x) = \int_{-\infty}^{x} \exp\left(\frac{-t^2}{2}\right) \frac{dt}{\sqrt{2\pi}}$

2.1.2 The Central Limit Theorem in the i.i.d case

Theorem 17 *Let $(X_i)_{i\geq 1}$ be a sequence of independent random variables,with the same law and square integral (and not constant). Note*

$\mu = E[X_1]$,$\sigma^2 = Var(X_1)$ *with* $\sigma > 0$

so

$S_n^* = \frac{S_n - E[S_n]}{\sqrt{Var(S_n)}} = \frac{S_n - n\mu}{\sigma\sqrt{n}} \to_{n\to\infty}^{law} N(0,1)$

Corollary 18 *If Y_n follows a Poisson distribution with parameter $\lambda = n\alpha$ (α constent)*

$$Y_n^* = \frac{Y_n - \lambda}{\sqrt{\lambda}} \to_{n\to\infty}^{law} N(0,1)$$

Proof. It suffices to notice that Y_n has the same law as $X_1 + ... + X_n$ where the X_i are

independent and with the same distribution $Pois(\alpha)$, the S_n^* build on the series of X_i converges in law to $N(0,1)$. As $E[Y_n] = \lambda$

and $Var(Y_n) = \lambda$, we see that Y_n^* has the same law .that S_n^* . It follows that Y_n^* converges in law to the same limit as S_n^* .

Before examining other applications of the central limit theorem, it is advisable to recall the following result .

Lemma 19 ***Slutsky***

Let (X_n) and (Y_n) be two sequences of real random variables defined on the same probabilized space and convergent in law respectively towards the random variable X and the constant C . Then the sequence of random vectors (X_n, Y_n) converges in law to (X, C). Here are three useful consequences:

- a) For any continuous function $g : \mathbb{R}^2 \to \mathbb{R}$ converges in law to $g(X, C)$.
- b) If X_n converges in law to X and Y_n converges in probability to 0 , then $X_n + Y_n$ converges in law to X .
- c) If X_n converges in law to X and Y_n converges in probability to C , then $X_n Y_n$ converges in law tawards CX .

■

Corollary 20 ***Convergence of Student's t_n statistic***

let $(X_i)_{i \geq 1}$ be a sequence of random variables i.i.d defined on the same space and having a moment of order 2 . We note

$E[X_1] = \mu$

and

$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$

so

$T_n = \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2}} \rightarrow_{n\rightarrow\infty}^{law} N(0,1)$.

Comment: When we want to construct confidence intervals for the estimate of an unknown expectation μ by the arithmetic mean $\bar{X}_n$ of the sample, the first idea is to use Theorem17 by noting that

$$S_n^* = \frac{S_n - E[S_n]}{\sqrt{Var(S_n)}} = \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sqrt{Var(X_1)}}$$

The problem is that in general $Var(X_1)$ is unknown. It is then replaced by an estimator without bias

$$W_n = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$$

which is none other than the σ_{n-1}^2 of calculators. Corollary 20 provides the legitimation theory of this recipe .

Proof. Before demonstrating $T_n = \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2}} \rightarrow_{n\rightarrow\infty}^{law} N(0,1)$. , let us verify that W_n is a strongly consistent estimator and without bias.

Interpreting

$V_n(w) = \frac{1}{n}\sum_{i=1}^{n}(X_i(w) - \bar{X}_n(w))^2$

as the variance of the measure $v_n(w) = n^{-1}\sum_{i=1}^{n}\delta_{X_i(w)}$ (empirical measure, whose esperance is $\int_{\mathbb{R}} x dv_n(w)(x) = \bar{X}_n(w)$, Koenig's formula gives us

$V_n(w) = \frac{1}{n}\sum_{i=1}^{n} X_i^2(w) - (\bar{X}_n(w))^2$.

By a double application of the strong law of large numbers,we see that

$V_n \to_{n\to\infty}^{A.s.} E\left[X_1^2\right] - (E\left[X_1\right])^2 = Var(X_1)$

using Koenig's formula again, but in the other direction and for the law of X_1 instead of $v_n(w)$. As $W_n = (\frac{n}{(n-1)})V_n$, we deduce

$W_n \to_{n\to\infty}^{A.s.} Var(X_1)$

Now let us verify that W_n is unbiased

$E(\sum_{i=1}^n (X_i - \bar{X}_n)^2) = nE\left(\frac{1}{n}\sum_{i=1}^n X_i^2 - (\bar{X}_n)^2\right)$

$= E(\sum_{i=1}^n X_i^2) - nE(\frac{1}{n^2}(\sum_{i=1}^n X_i)^2)$

$= nE\left[X_i^2\right] - \frac{1}{n}E(\sum_{i=1}^n X_i^2 + \sum_{1\leq i\neq j\leq n} X_i X_j)$

$= (n-1)E\left[X_1^2\right] - \frac{1}{n}\sum_{1\leq i\neq j\leq n} E\left[X_i X_j\right]$

$= (n-1)E\left[X_1^2\right] - \frac{n^2-n}{n}(E\left[X_1\right]^2)$ (because $E\left[X_1 X_2\right] = (E\left[X_1\right])^2$)

$= (n-1)(E\left[X_1^2\right] - E\left[X_1\right])^2)$

$= (n-1)Var(X_1)$

We therefore have $E\left[W_n\right] = Var\,(X_1)$

To establish the convergence in law of T_n , we write

$T_n = \frac{S_n - n\mu}{\sigma\sqrt{n}}\sigma W_n^{-1/2}$

where $\sigma = \sqrt{Var(X_1)}$, $\sigma W_n^{-1/2}$ Almost sure convergence and therefore a fortiori in law towards 1 . The Theorem 17 and Slutsky's lemma give us the convergence of T_n towards $N(0,1)$ ■

2.2 Reciprocal of the i.i.d CLT

Theorem 21 *Let $(X_i)_{i\geq 1}$ be a sequence of random variables i.i.d defined on the same probabilized space and such that $\frac{S_n}{\sqrt{n}}$ con-*

verges in law to $N(0,1)$ *. Then* X_1 *is square integrable,* $E[X_1] = 1$ *and* $E[X_1] = 0$

comment : This theorem essentially contains the converse of Theorem17.More precisely, we deduce that if there are $\mu \in \mathbb{R}$ and $\sigma > 0$ constants such that $(\sigma n)^{-1/2}(S_n - n\mu)$ converges in law to $N(0,1)$, then $X_1 \in L^2(\Omega)$, $E[X_1] = \mu$ and $Var(X_1) = \sigma^2$. Indeed, it suffices to apply Theorem to the random variables $\acute{X} = \frac{(X_i - \mu)}{\sigma}$.Thus Theorem 17 give us the equivalence between the convergence in law towards a Gaussian of $n^{-1/2}(S_n - n\mu)$ (for a certain constant μ) and the existence of a moment of order 2 for X_1.

Proof. The proof has 5 steps including two lemmas. We start with reduce the problem to the proof of the member ship of X_1 to $L^2(\Omega)$. Lemma 9 establishes then a link between the behavior of the characteristic function of X_1 in the neighborhood from 0 and $E[X_i^2]$. We can then complete the proof in the particular case where X_1 has a law symmetrical. Lemma 22 and 23 are useful for the transition from the symmetric case to the general case . ■

Reduction to the proof of $E[X_1] < +\infty$ .Suppose that X_1 belongs to $L^2(\Omega)$. We then have $E[X_1] < +\infty$ therefore by the strong law of large numbers, $n^{-1}S_n$ almost sur converges towards $E[X_1]$, therefore also in lawtowards the same limit. On the other hand, by writing $n^{-1}S_n = n^{-1/2}(n^{-1/2}S_n)$, the hypothesis of convergence of $n^{-1/2}S_n$ towards Z of law $N(0,1)$ a nd Slutsky's lemma (in a form degenerate) give us the convergence in law of $n^{-1}S_n$ towards $0 \times Z = 0$. By uniqueness ofthe limit law, the two constant random variables $E[X_1]$ and 0 must have the same law.This implies $E[X_1] = 0$ (for Dirac masses, $\delta_a = \delta_b$ is equivalent to $a = b$).Then by the CLT , $n^{-1/2}S_n$ converges in law to $N(0, E[X_1^2])$. By uniqueness of the limit law, we must therefore have $N(0,1) = N(0, E[X_1^2])$ hence $E[X_1^2] = 1$.

Lemma 22 *Let be X a real random variable with characteristic function φ so*

$$\lim_{u \to 0} \frac{2(1 - \operatorname{Re}\varphi(u))}{u^2} = \int_{\mathbb{R}} x^2 dP_X(x) = E[X^2] \in \bar{\mathbb{R}}_+$$

Lemma 23 *If X and Y are two independent random variables such that $X + Y \in L^2(\Omega)$, then $X \in L^2(\Omega)$ and $Y \in L^2(\Omega)$*

2.3 Convergence Speed

The speed of convergence in Theorem 17 is in the right cases in $O(n^{-1/2})$, as for the de Moivre-Laplace theorem. More precisely, we have the following result

Theorem 24 ***Berry-Esséen***

Let $(X_i)_{i \geq 1}$ be a sequence of random variables i.i.d such that $E[X_1] = 0$, $E|X_1|^3 < +\infty$ We denote by $\sigma^2 = E[X_1^2]$ $(\sigma > 0)$. universal constant $C > 0$ such that for all $n \geq 1$,

$$\Delta_n = \sup \left| P\left(\frac{S_n}{\sigma\sqrt{n}} \leq x\right) - \Phi(x) \right| \leq C \frac{E|X_1|^3}{\sigma^3} \frac{1}{\sqrt{n}}$$

The key to the problem is to obtain an inversion formula allowing to control the difference of the distribution functions using the difference of the characteristic functions ticks. Once this has been acquired, we carry out a third-order expansion as inthe proof of CLT is what explains the presence of $E|X_1|^3$ *in the upper bound . getting the best constant* C *has been the object of a long quest. The initial value of Esséen was* $C = 7,59$*. Feller proposes* $C = 3$ *. A value more modern and close to the optimum is* $C = 0.7975$ *(Van Beek (1972).*

It is interesting to look at what the Berry-Esséen theorem gives for the case of de Moivre-Laplace, therefore with $X_1 = Y_1 - p$, where Y_1 follows a Bernoulli law with parameter p. We then find

$$\Delta_n \leq C\frac{p^2+q^2}{\sqrt{pq}}\frac{1}{\sqrt{n}}$$

Here is a very elementary example allowing to understand that there is no instead of hoping for a speed of convergence better than $O(n^{-1/2})$ for Δ_n Take X_1 of Bernoulli's law of parameter $\frac{1}{2}$. We then have

$$S_n \sim Bin(2n, \tfrac{1}{2}), \qquad E(S_{2n}) = 2n\tfrac{1}{2} = n.$$

We are looking for an equivalent of $P(S^*_{2n} \leq 0) - \Phi(0)$. Note first that

$$\{S^*_{2n} < 0\} = \{0 \leq S_{2n} < n\} \quad \text{and} \quad \{S^*_{2n} > 0\} = \{n < S_{2n} \leq 2n\}$$

Due to the symmetry of the binomial coefficients $(C_{2n}^{k} = C_{2n}^{2n-k})$,

$$P(S_{2n}^{*} < 0) = \sum_{k=0}^{n-1} C_{2n}^{k} 2^{-2n} = \sum_{j=n+1}^{2n} C_{2n}^{j} 2^{-2n} = P(S_{2n}^{*} > 0)$$

We thus have $2P(S_{2n}^{*} < 0) + P(S_{2n}^{*} = 0) = 1$ from which we take $P(S_{2n}^{*} < 0) = \frac{1}{2} - \frac{1}{2} P(S_{2n}^{*} = 0)$ and $P(S_{2n}^{*} \leq 0) = \frac{1}{2} + \frac{1}{2} P(S_{2n}^{*} = 0)$. Recalling that $\Phi(0) = \frac{1}{2}$, we end up with

$$P(S_{2n}^{*} \leq 0) - \Phi(0) = \frac{1}{2} P(S_{2n}^{*} = 0) = \frac{1}{2} P(S_{2n} = n) = C_{2n}^{n} 2^{-2n-1}$$

By the Stirling formula $(n! \sim \sqrt{2\pi} n^{n+1/2} e^{-n})$, we get the equivalent

$$P(S_{2n}^{*} \leq 0) - \Phi(0) \sim \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{2n}}$$

Since $(2\pi)^{-1/2} > 0.3989$, we have for $n \geq n_0$,$|P(S_{2n}^{*} \leq 0) - \Phi(0)| \geq 0.398(2n)^{-1/2}$,lower bound to be compared with the uniform upper bound $\Delta_{2n} \leq 0.798(2n)^{-1/2}$ provided in this case by the Berry-Esséen theorem

Let us return to the general situation of Theorem 11. What is happening in the area intermediate where X_1 satisfies the CLT $(i.e\ E[X_1^2] < +\infty)$ but has no order moment 3 ? We always have a speed of convergence, provided that the integrability of X_1 is a little stronger than the only existence of a moment of order 2 .

Theorem 25 *(Katz 1963, Petrov 1965). Let* $(X_i)_{i\geq 1}$ *be a sequence of random variables i.i.d such that* $E[X_1] = 0,\ E[X_1^2]\, g(X_1) < +\infty$

where g *is a positive, pair, increasing on* $]0,+\infty[$ *and such that* $\frac{x}{g(x)}$ *is increasing on* $]0,+\infty[$. *There is then a universal constant* $A>0$ *such that for all* $n\geq 1$

$$\Delta_n = \sup\left|P(\tfrac{S_n}{\sqrt{n}}\leq x)-\Phi(x)\right|\leq A\tfrac{E(X_1^2 g(X_1))}{\sigma^2 g(\sigma\sqrt{n})}$$

in particular if $E\left|X_1\right|^{2+\delta}<+\infty$ for a $\delta\in]0,1]$,

$$\Delta_n \leq A\tfrac{E|X_1|^{2+\delta}}{\sigma^{2+\delta}}\tfrac{1}{n^{\delta/2}}$$

2.4 CLT without Equidistribution

We now consider the question of CLT for independent random variables not necessarily having the same law. The essential result is Lindeberg's theorem

Theorem 26 ***lindeberg***

Consider the ‹triangular array › of random variables of row number $n(n\in\mathbb{N}^*)$

$X_{n,1},\ldots,X_{n,i},\ldots X_{n,k_n}$,

where these k_n *variables are defined on the same* $(\Omega_n,\mathcal{F}_n,P_n)$, *independent, centered, of integrable squares. We notice*

$\sigma^2_{n,i}=Var(X_{n,i})$ $\qquad S^2_n=\sum_{i=1}^{k_n}\sigma^2_{n,i}$ $\qquad S_n=\sum_{i=1}^{k_n}X_{n,i}$

We assume $S_n>0$ *for all* n *If moreover the table satisfies the condition of Lindeberg*

$\forall\varepsilon>0,\frac{1}{S^2_n}\sum_{i=1}^{k_n}\int_{\{|X_{n,i}|>\epsilon s_n\}}X^2_{n,i}dP_n\rightarrow_{n\rightarrow+\infty}0$

then $\frac{S_n}{s_n}$ *converges in law to* $N(0,1)$

Theorem 27 *We suppose that* X *and* Y *are square integrable (with* $\sigma_X, \sigma_Y > 0$ *) and that* $l = l(n)$ *and* $m = m(n)$ *tend to infinity with* n*. Then under* (H_0) *the statistic of test* $T = T_n$ *converges in law to* $N(0,1)$ *as* n *approaches infinity. Through against, under* (H_1)*,* $|T_n|$ *almost sure tends towards infinity*

Proof. Whether we are under (H_0) or (H_1), the independence and equidistribution of the X_i the one hand and Yj on the other hand give us by the strong law of large numbers the almost sure convergences from $\bar{X}$ to $E[X]$, from S_X^2 to σ_X^2 , from $\bar{Y}$ to $E[Y]$ and from S_Y^2 towards σ_Y^2 .

Thus the numerator of T_n almost sure converges towards $E[X]-E[Y]$ and the almost sure denominator towards 0 . It is then clear that under (H_1) , $|T_n|$ tends almost sure towards infinity.

Under (H_0) we have an indeterminate form of the type $\langle 0/0 \rangle$ for the almost sure convergence and we study the convergence in law. Consider for this the triangular array of nth line

$\frac{-X_1}{l}, \ldots, \frac{-X_l}{l}, \frac{Y_1}{m}, \ldots, \frac{Y_m}{m}$

Here $k_n = l + m$ and the sum of the row is $S_n = \bar{Y} - \bar{X}$ Thanks to independence, its variance is

$s_n^2 = Var(S_n) = \frac{1}{l} \sum_{i=1}^{l} Var(X_i) + \frac{1}{m^2} \sum_{j=1}^{m} Var(Y_j) = \frac{1}{l}\sigma_X^2 + \frac{1}{m}\sigma_Y^2$

By the strong law of large numbers, it is easy to verify that

$\frac{\frac{1}{l(n)}\sigma_X^2 + \frac{1}{m(n)}\sigma_Y^2}{\frac{1}{l(n)}S_X^2 + \frac{1}{m(n)}S_y^2} \xrightarrow[n \to +\infty]{A.S} 1.$

by writing T_n in the form

$T_n = \left(\frac{\frac{1}{l}\sigma_X^2 + \frac{1}{m}\sigma_Y^2}{\frac{1}{l}S_X^2 + \frac{1}{m}S_y^2} \right)^{1/2} \times \frac{S_n}{s_n}$

and by invoking Slutsky's lemma, we reduce the proof of the theorem to the verification of the Lindeberg condition for the triangular array. It is therefore a question of showing for all $\varepsilon > 0$ the convergence towards 0 of

$A_n(\varepsilon) = \frac{1}{s_n^2} \left[\sum_{i=1}^{l} \int_{\{|u| \geq \varepsilon s_n\}} u^2 P_{(-X_i/l)}(du) + \sum_{j=1}^{m} \int_{\{|v| \geq \varepsilon s_n\}} v^2 P_{(Y/m)}(dv) \right]$

$= \frac{1}{s_n^2} \int_{\{|u| \geq \varepsilon s_n\}} u^2 P_{(-X_1/l)}(du) + \frac{m}{s_n^2} \int_{\{|v| \geq \varepsilon s_n\}} v^2 P_{(Y/m)}(dv)$

The changes of variable $u = \frac{-x}{l}$ and $v = \frac{y}{m}$ give us as measures respective images P_{X_1} and P_{Y_1} and as new integration sets $\{|x| \geq \varepsilon l s_n\}$and $\{|y| \geq \varepsilon m s_n\}$. We thus see that

$A_n(\varepsilon) = \frac{1}{l s_n^2} \int_{\{|x| \geq \varepsilon l s_n\}} x^2 P_{X_1}(dx) + \frac{1}{m s_n^2} \int_{\{|y| \geq \varepsilon m s_n\}} y^2 P_{Y_1}(dy)$

To increase $A_n(\varepsilon)$, we notice that $l s_n^2 > \sigma_X^2$ hence the inclusion

$\{|x| \geq \varepsilon l s_n\} = \{x^2 \geq \varepsilon^2 l^2 s_n^2\} \subset \left\{x^2 \geq l \varepsilon^2 \sigma_X^2\right\}$

by doing the same with m and σ_Y^2 , we get

$A_n(\varepsilon) \leq \frac{1}{\sigma_X^2} \int_{\left\{x^2 \geq l\varepsilon^2 \sigma_X^2\right\}} x^2 P_{X_1}(dx) + \frac{1}{\sigma_Y^2} \int_{\left\{y^2 \geq m\varepsilon^2 \sigma_Y^2\right\}}$

As $l = l(n)$ and $m = m(n)$ tend towards $+\infty$, this upper bound tends towards 0 by convergence dominated since X_1 and Y_1 are square integrable

Theorem 28 ***Lyapounov***

Let $(X_k)_{k \geq 1}$ be a sequence of random variables defined on the same probabilized space, independent (but not necessarily of the same law), centered,all having a moment of order $2+\delta$ $(\delta > 0)$. *We notice $s_n^2 = Var(S_n)$ and we assumes $s_n > 0$. We further assume that Lyapunov's condition is verified*

$\frac{1}{s_n^{2+\delta}} \sum_{k=1}^{n} E\left|X_k\right|^{2+\delta} \rightarrow_{n\rightarrow+\infty} 0$

then S_n/s_n *converges in law to* $N(0,1)$ *.*

■

Proof. We apply Lindeberg's theorem to the triangular array $X_{n,k}= X_k$, $1 \leq k \leq n$. It is then sufficient to verify that the Lyapunov condition results in that by Lindeberg . By noting that on the event $\{|X_k| \geq \varepsilon s_n\}$ we have $\frac{|X_k|^\delta}{(\varepsilon s_n)^\delta} \geq 1$, the increase

$\int_{\{|X_k|\geq \varepsilon s_n\}} X_k^2 dP \leq \int_{\{|X_k|\geq \varepsilon s_n\}} X_k^2 (\frac{|X_k|}{\varepsilon s_n})^\delta dP \leq \frac{1}{\varepsilon^\delta s_n^\delta} \int_\Omega |X_k|^{2+\delta} dP$

gives the expected conclusion via inequality

$\frac{1}{s_n^2} \sum_{k=1}^{n} \int_{\{|X_k|\geq \varepsilon s_n\}} X_k^2 dP \leq \frac{1}{\varepsilon^\delta s_n^{2+\delta}} \sum_{k=1}^{n} E\left|X_k\right|^{2+\delta}$ ■

2.5 The CLT in $\mathbb{R}^d$

We study in this section the CLT in finite dimension. For the sake of simplicity, we will be limited mainly to CLt i.i.d in $\mathbb{R}^d$. We will see that in fact there is always a way to reduce to dimension 1 thanks to the Cramér-Wold lemma (Cramér-Wold device) (in Anglo-Saxon literature). However, this does not exempt the possession of a minimum knowledge on random vectors (functional characteristics,covariance structure, Gaussian laws in dimension d) . It is convenient to present these "reminders" in the more abstract framework of a vector space E , on the one hand to avoid to prematurely play a role in the Euclidean structure of $\mathbb{R}^d$, on the other hand for pave the way in the direction of Gaussian laws and CLT in infinite dimension

2.5.1 Random Vectors

Let E therefore be a finite dimensional $\mathbb{R}$-vector space d and E its dual. Recall that there is only one topology of vector space on E (i.e. making the addition continuous vectors and the multiplication of a vector by a scalar) and that all the norms on E are equivalent and each metrize this topology.

In the following, E will be provided of its Borelian tribe, that is to say the tribe generated by the openings of E. A vector random X in E is a measurable application of a probability space (Ω, F, P) in E.

The characteristic functional of X (or of its law) is the application

$$\varphi_X \colon \acute{E} \to \mathbb{C} \ , \ u \to \varphi_X(u) = E\exp\left(iu(X)\right)$$

The link with the characteristic functions in dimension 1 is given by

$$\forall u \in \acute{E} \ , \ \forall t \in \mathbb{R} \ , \ \varphi_X(tu) = E\exp\left(itu(X)\right) = \varphi_{u(X)}(t)\text{.}$$

As in dimension 1 , the characteristic functional characterizes the law and the convergence point of the characteristic functionals is equivalent to the convergence in law. The lemma of Cramér-Wold is an immediate consequence of these remarks

Lemma 29 ***Cramér-Wold***

The sequence $(X_n)_{n\geq 1}$ *of random vectors in vector spacefinite dimensional toriel* E *converges in law in* E *towards the random vector* X *if and only if for any linear form* $u \in \acute{E}$ *, the sequence of real random variables* $(u(X_n))_{n\geq 1}$ *converges in law in* $\mathbb{R}$ *towards* $u(X)$*.*

Let us move on to the definition of the expectation of a random vector X *. We suppose for this that for all* $u \in \acute{E}$*,* $E\,|u(X)|$ *is finite and we consider the application*

$\theta : \acute{E} \to \mathbb{R}$, $\quad u \to E(u(X))$

Due to the linearity of the expectation, θ *is a linear form on* E *therefore an element of bidual* E"*. There is then a unique element* $x_0 \in E$ *such that* $\theta(u) = u(x_0)$*. This element which is defined as the* $E\,[X]$ *deterministic vector. Thus* $E\,[X]$ *is the unique element of* E *verifying*

$\forall u \in \acute{E}$, $u(E\,[X]\,) = E\,[u(X)]$

To give a more familiar expression of $E\,[X]$ *, let us choose a basis* $(e_1, ..., e_d)$ *of* E *and denote by* $\acute{e}_k$ *the associated coordinate forms (*$\acute{e}_k(e_j) = \delta_{j,k}$ *, Kronecker symbol) so that* $(\acute{e}_1, ..., \acute{e}_d)$ *is a basis of* $\acute{E}$ *and all* $X \in E$ *is written* $x = \sum_{k=1}^{d} \acute{e}_k(x)e_k$ *. In applying* $\forall u \in \acute{E}$, $u(E\,[X]\,) = E\,[u(X)]$ *with* $u =_{\acute{e}_k} (k = 1, ..., d)$ *, we obtain by noting that* $X = \sum_{k=1}^{d} \acute{e}_k(X)e_k$*,*

$E\,[X] = \sum_{k=1}^{d} (E[\acute{e}_k(X)])e_k = (E\,[\acute{e}_1(X)]\,, ..., E\,[\acute{e}_d(X)]\,)$

Suppose now that for all $u \in \acute{E}$ *,*$u(X)$ *has an integrable square and define*

$B : \acute{E} \times \acute{E} \to \mathbb{R}$, $(u, v) \to B(u, v) = Cov(u(X), v(X))$

B is clearly a bilinear form on $\acute{E} \times \acute{E}$ and using the identification of E" with E , we have for each $v \in \acute{E}$ a unique element $y = Kv \in E$ such that $u(y) = B(u, v)$ .The operator $K : \acute{E} \to E$ is linear, autoadjoint $(u(Kv) = v(Ku))$ and positive $(u(Ku) \geq 0)$.

We call it the covariance operator of X . It is intrinsically defined by

$\forall u, v \in \acute{E}$, $u(Kv) = Cov(u(X), v(X))$.

We can give a matrix representation of K by providing the arrival space E of a basis $(e_1, ..., e_d)$ and the starting space E of the dual basis $(\acute{e}_1, ..., \acute{e}_d)$ defined as above. By choosing $u = \acute{e}_i$ and $v = \acute{e}_j$ in $\forall u, v \in \acute{E}$, $u(Kv) = Cov(u(X), v(X))$. we see that the matrix of K has a general term

$K_{i,j} = Cov(\acute{e}_i(X), \acute{e}_j(X))$, $1 \leq i, j \leq d$

The covariance operator gives information on the support of the law of X. Let us indeed define $H = \left\{u \in \acute{E}; u(Ku) = 0\right\}$. As the belonging of u to his equivalent to $Var(u(X)) = 0$, it is easy to verify that H is a vector subspace of $\acute{E}$.

Consider then $H^0 = \{x \in E; \forall u \in H, u(x) = 0\}$. It is clear that H^0 is a v.s of E .

The law of X is supported by the affine subspace $E[X] + H^0$, which means that $P(X - E[X] \in H^0) = 1$. To verify this, note d_0 the dimension of H and choose any basis $(f_i)_{1 \leq i \leq d_0}$ of H. We then have

$x \notin H \iff \exists i \in \{1, ..., d_0\}, f_i(x) \neq 0.$

from where

$P(X - E[X] \notin H^0) = P(\cup_{i=1}^{d} \{f_i(X - E[X]) \neq 0\}) \leq \sum_{i=1}^{d} P(f_i(X - E[X]) \neq 0)$.

Now by definition of H *, for all* $i = 1, ..., d_0$, $Var(f_i(X - E[X]) = 0$ *, which implies that the real random variable* $f_i(X - E[X])$ *which has zero expectation is equal to* 0 *with probability* 1 *.The right-hand member of*

$P(X - E[X] \notin H^0) = P(\cup_{i=1}^{d} \{f_i(X - E[X]) \neq 0\})$
$\leq \sum_{i=1}^{d} P(f_i(X - E[X]) \neq 0)$.

is therefore zero and we have $X \in E[X] + H^0$ *almost surely .*

We say that the random vector X *is Gaussian if for all* $u \in \acute{E}$ *, the variable real random* $u(X)$ *is Gaussian . This implies that* $u(X)$ *is square integrablefor all* u *and according to the above,* X *therefore has an expectation* $m \in E$ *and an operator of covariance* $K : \acute{E} \to E$ *.*

The law of X *is then noted* $N(m, K)$ *. As* $E[u(x)] = u(m)$ *and* $Var(u(X)) = u(Ku)$ *, the Gaussian law of* $u(X)$ *is* $N(u(m), u(Ku))$ *,. We deduce that the characteristic functional of* $N(m, K)$ *is written*

$\varphi_{X(u)} = \exp\left(iu(m) - \frac{u(Ku)}{2}\right)$, $\quad u \in \acute{E}$

An immediate consequence of this definition is that the image $T(X)$ *of a gaussian vector of* E *by a linear application* T *from* E *to* F *is a Gaussian vector of* F *having for covariance operator* T K $\acute{T}$ *where* $\acute{T} : \acute{F} \to \acute{E}$ *is the adjunct of* T*, defined by*

$\forall x \in E$, $\quad \forall v \in \acute{F}$, $\quad v(T(x)) = (\acute{T}(v))(x)$

In the usual case $E = \mathbb{R}^d$ *, we identify* E *and* $\acute{E}$ *and we notice* *by* $\sum$ *the matrix of* K *for the canonical basis of* $\mathbb{R}^d$. *When* K *is positive (the subspace* H *of* $\acute{E}$ *defined above is therefore reduced to the zero vector and* K *is invertible) the Gaussian law* $N(m, \sum)$ *then has for density with respect to the canonical Lebesgue measure of* $\mathbb{R}^d$

$$x \to \frac{1}{(2\pi)^{d/2}(\det \sum)^{1/2}} \exp\left(-\frac{1}{2}(x-m)^t \sum^{-1} (x-m)\right)$$

In this case, the support of the Gaussian law $N(m, K)$ *is the whole space* $\mathbb{R}^d$. *If* K *is only positive semi-definite, the law* $N(m, K)$ *still exists, but is supported by the sub-affine space* $m + H^0$ *, where* H^0 *is now the orthogonal of* H *(after identification of* $E = \mathbb{R}^d$ *and* $\acute{E}$*) and has a density with respect to the canonical Lebesgue measure of this subspace, except in the completely degenerate case where* $H = \mathbb{R}^d$ *and where* $N(m, K)$ *is the Dirac mass at point* m .

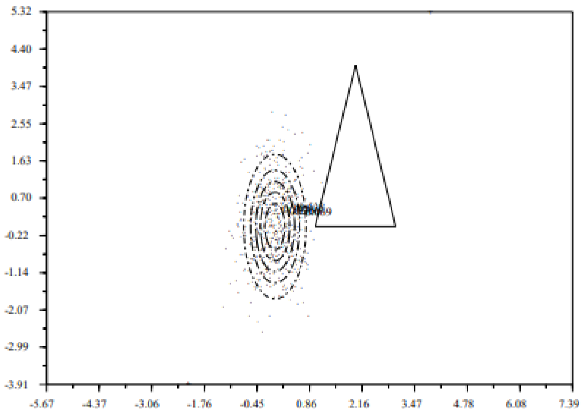
2.5.2 CLT i.i.d

Theorem 30 *Let* (X_k) *be a sequence of independent random vectors of* $\mathbb{R}^d$ *,of the same law and square integrable* $(i.e\ E\|X_1\|^2 < +\infty)$ *and* $S_n = \sum_{k=1}^n X_k$ *so*

$$\frac{S_n - E[S_n]}{\sqrt{n}} \xrightarrow[n\to+\infty]{law} N(0, K)$$

where K *is the covariance operator of* X_1

Proof. By the Cramér-Wold lemma, it suffices to verify that for any linear form u on $\mathbb{R}^d$, $u(n^{-1/2}(S_n - E[S_n]))$ converges in law to $u(Z)$ where Z denotes a random vector Gaussian with law $N(0, K)$



Let us note straight away that $E\left[u(Z)\right] = u(E\left[Z\right]) = u(0) = 0$ and $Var(u(Z)) = u(Ku) = Var(u(X_1))$

By linearity of u

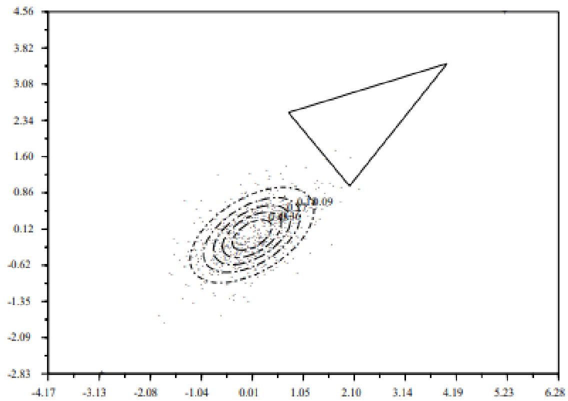
$u(\frac{S_n - E[S_n]}{\sqrt{n}}) = \frac{1}{\sqrt{n}} \sum_{k=1}^{n} (u(X_k) - E\left[u(X_k)\right])$

The $u(X_k)$ are independent real random variables, of the same law, of squared integrable $(E\left|u(X_1)\right|^2 \leq \|u\|_{E'}^2 E\left\|X_1\right\|_E^2)$.

$u(\frac{S_n - E[S_n]}{\sqrt{n}}) \rightarrow_{n\to+\infty}^{law} N(0, Var\ \ u(X_1))$

The limit law is indeed that of $u(Z)$

To illustrate the theorem graphically in $\mathbb{R}^2$, we have chosen to generate 600 samples.independent of size 500 of a uniform law over a triangle. For each of these samples have been calculated $n^{-1/2}(S_n - E\left[S_n\right])$ and plotted the corresponding point. We obtain thus a cloud of 600 points which behaves approximately like a sample of size 600 of the limit Gaussian law. The level lines drawn on this cloud of points are those of the density of this Gaussian law (figures 1.1and 1.2) . ■



1.png

figure 1.1 $-n^{-1/2}(S_n - E[S_n])$ for 600 samples of the uniform law on the triangle

figure 1.2 $-n^{-1/2}(S_n - E[S_n])$ for 600 samples of the uniform law on the triangle

2.6 Local Limit Theorems

The CLT on $\mathbb{R}$ converges the mass of an interval for the law of S_n^* towards the mass of same interval for the standard Gaussian law $N(0,1)$. Local limit theorems relate to a uniform convergence to the density of $N(0,1)$.

The supposed quantity converge towards this density will be either the density of S_n^* wh en it exists, let the mass suitably normalized of atoms of the law of S_n^* when the law of X_1 is supported by a discrete lattice of $\mathbb{R}$.

Theorem 31 *Let $(X_{k)})_{k\geq 1}$ be an i.i.d sequence of centered real* *random variables and of variance* 1*. We suppose that the characteristic function* $\varphi(t) = E\left[\exp(itX_1)\right]$ *satisfies*

$\int_{\mathbb{R}} |\varphi(t)|\, dt < +\infty$

Then $\frac{S_n}{\sqrt{n}}$ *has a density which converges uniformly on* $\mathbb{R}$ *towards that of* $N(0,1)$

Theorem 32 *Let* $(X_k)_{k\geq 1}$ *be an i.i.d sequence of centered real random variables and variance* 1 *. We further assume that the law of* X_1 *is supported by a network* $b + h\mathbb{Z}$ *(*$b \in \mathbb{R}$ *and* $h \in \mathbb{R}^{+*}$ *fixed) In the sequence* h *denote the greatest real such that* $P(X_1 \in b + h\mathbb{Z}) = 1$*.Atoms of the law of* $\frac{S_n}{\sqrt{n}}$ *are then among the reals* $\frac{(nb+kh)}{\sqrt{n}}, k \in \mathbb{Z}$ *We note* A_n *the set of these atoms and we define*

$P_n(x) = P(\frac{S_n}{\sqrt{n}} = x), x \in A_n$

Then we note $f(x) = (2\pi)^{-1/2}\exp\left(-x^2/2\right)$

$sup_{x\in A_n}\left|\frac{\sqrt{n}}{h}p_n(x) - f(x)\right| \to_{n\to+\infty} 0$

comments:

The conclusion is also written

$$P_n(x) = \frac{hf(x)}{\sqrt{n}} + \frac{\varepsilon_n(x)}{\sqrt{n}}, \sup_{x\in A_n}|\varepsilon_n(x)| \to_{n\to+\infty} 0$$

Let us examine the application of this result to the Gaussian approximation of a binomial distribution.Let $(X_k)_{k\geq 1}$ therefore be an i.i.d sequence of Bernoulli random variables with parameter p , $S_n = X_1 + ... + X_n$ then follows the $Bin(n,p)$ law By noting $q = 1 - p$, we have

$$S_n^*=\frac{S_n-E[S_n]}{\sqrt{Var(S_n)}}=\frac{S_n-np}{\sqrt{npq}}=\frac{1}{\sqrt{n}}\sum_{k=1}^{n}\acute{X}_k=\frac{\acute{S}_n}{\sqrt{n}}$$

where $\acute{X}_k=\frac{(X_k-p)}{\sqrt{pq}}$, atoms of the law of $\frac{\acute{S}_n}{\sqrt{n}}$ are the x such that $P(\acute{S}_n= x\sqrt{n}) > 0$ By setting $x =\frac{y}{\sqrt{npq}}$, we have

$$P(\acute{S}_n= x\sqrt{n}) = P(S_n-np = y)$$

and as the point masses of the binomial distribution of S_n are the integers $K = 0, 1, 2, ..., n$,we see that

$$A_n=\left\{\frac{k-np}{\sqrt{npq}}; k = 0, 1, 2, ..., n\right\}$$

So $b = -\sqrt{\frac{p}{q}}, h =\frac{1}{\sqrt{pq}}$ and is written

$$P(S_n= k) =\frac{1}{\sqrt{2\pi npq}}\exp\left(\frac{-(k-np)^2}{2npq}\right)+\frac{\varepsilon_{n,k}}{\sqrt{n}}, \quad \max_{0\leq k\leq n}|\varepsilon_{n,k}| \to_{n\to+\infty}0$$

2.7 CLT for The Sums of a Random Number of Terms

Theorem 33 *All the considered random variables being defined on the same probabilized space, we suppose that*

1. $(X_k)_{k\geq1}$ *is a sequence of random variables such that*

   $\frac{1}{\sqrt{n}}\sum_{k=1}^{n}X_k\to^{law}_{n\to+\infty}N(0,\sigma^2), (\sigma > 0)$

2. $(N_n)_{n\geq1}$ *is a sequence of random variables with values in* $\mathbb{N}^*$ *tending to* $+\infty$ *in probability,*

   which means : $\forall A > 0, \lim_{n\to+\infty} P(N_n> A) = 1$

3. *For all* $j \geq 1, N_j$ *is independent of the sequence* $(X_k)_{k\geq 1}$ *in these conditions ,*

$\frac{1}{\sqrt{N_n}} \sum_{k=1}^{N_n} X_k \rightarrow_{n\rightarrow+\infty}^{law} N(0, \sigma^2)$

comments: Here is an illustration of this theorem in the field of insurance N_n represents the number of claims received by a company during n units of time (say during the n working days since an origin date) X_k is the amount of compensation paid by the company for the kth claim so stated. The total amount of reimbursements for the period considered is therefore $\sum_{k=1}^{N_n} X_k$

Theorem 34 *All the considered random variables being defined on the same probabilized space, we suppose that*

1. $Y_n \rightarrow_{n\rightarrow+\infty}^{law} Y$

2. $N_n \rightarrow_{n\rightarrow+\infty}^{pr} +\infty$

3. *For all* $j, n \in \mathbb{N}^*, Y_j$ *and* N_n *are independent*

so

$Y_{N_n} \rightarrow_{n\rightarrow+\infty}^{law} Y$

Proof. It suffices to show that at any point of continuity X of the distribution function F of Y ,

$P(Y_{N_n} \leq x) \rightarrow_{n\rightarrow+\infty} F(x)$

To do this, we start by partitioning according to the possible values of N_n to obtain from 3

$P\left(Y_{N_n}\leq x\right)=\sum_{j=1}^{+\infty}P\left(Y_{N_n}\leq x,N_n=j\right)$

$=\sum_{j=1}^{+\infty}P\left(Y_j\leq x,N_n=j\right)$

$=\sum_{j=1}^{+\infty}P\left(Y_j\leq x\right)P\left(N_n=j\right)$

By 1, we have for all $\varepsilon\in]0,1[$ a $j_0=j_0(\varepsilon,x)$ such that ,

$\forall j>j_0,\left|P\left(Y_j\leq x\right)-F\left(x\right)\right|\leq\varepsilon$

By cutting then

$P\left(Y_{N_n}\leq x\right)-F\left(x\right)=\left\{\sum_{j=1}^{j_0}+\sum_{j=j_0+1}^{+\infty}\right\}\left(P\left(Y_j\leq x\right)-F\left(x\right)\right)P\left(N_n=j\right)$

and by increasing $\left|P\left(Y_j\leq x\right)-F\left(x\right)\right|$ by 1 for $j\leq j_0$ and by ε for $j>j_0$, we end up with

$\left|P\left(Y_{N_n}\leq x\right)-F\left(x\right)\right|\leq P\left(N_n\leq j_0\right)+\varepsilon P(N_n>j_0)$

Due to2 , the two terms of this upper bound tend towards 0 and towards ε respectively when n tends to $+\infty$. We can deduce

$\forall\varepsilon\in]0,1[$, upper $\lim_{n\to+\infty}\left|P\left(Y_{N_n}\leq x\right)-F\left(x\right)\right|\leq\varepsilon$,which give

$P\left(Y_{N_n}\leq x\right)\to_{n\to+\infty}F\left(x\right)$ and ends the proof . ■

Chapter 3

Applications of The Central Limit Theorem

we state the central limit theorem

Theorem 35 *Suppose that* $X_1, X_2...$ *is an infinite sequence of independent, identically distributed random variables with common mean* $\mu = E(X_1)$ *and finite variance* $\sigma^2 = Var(x_1)$*. Then, if we let* $S_n = X_1 + ... + X_n$ *we have that*

$$\lim_{n\to+\infty} P\left(\frac{S_n - n\mu}{\sigma\sqrt{n}} \leq c\right) = \Phi(c) = \int_{-\infty}^{c} \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx$$

There are many applications of this theorem to real-world problems, and we will give two: An application to hypothesis testing, and an application to noise cancellation.

3.1 Hypothesis Testing

Here we will give an example of how to use the CLT to test hypotheses. We have already seen how to do this using a chi-square test to determine whether to reject a hypothesized population distribution (with finitely many classes) as being false. Here we will do this for when the population breaks down into two classes, smokers and non-smokers.

3.1.1 The Main Problem

Problem 36 *You read in a newspaper that* 20% *of Georgians smoke, and you decide to test this hypothesis by doing a poll on* $1,000$ *randomly selected Georgians with replacement (if the population you are testing is very large, then you would not need to test with replacement). Suppose that* 205 *of the responses are "smoker", while* 795*are "non-smoker". Is the claim "*20% *of Georgians smoke" unreasonable? (Obviously not, but let's see what the math tells us...)*

Well, in order to answer this question we would need more information; we would need to know what we mean by "unreasonable". Here we will mean "unreasonable" with respect to a certain statistical test which we presently describe:

Let $X_i = 1$ if respondant i says he/she is a smoker, and let $X_i = 0$ if he/she is not a smoker. These X_i's are independent Bernoulli random variables. Let $S_n = X_1 + ... + X_{1,000}$. If our hypothesis that 20% of Georgians smoke were correct, then $\mu = E(X_i) = 0.2$, and $V(X_i) = \mu(1-\mu) = 0.16$; and so, the Central Limit Theorem would tell us that

$$\frac{S_{1,000}-200}{\sqrt{1,000 \cdot 0.16}} \text{ is approximately } N(0,1)$$

in the sense that

$$P\left(\frac{S_{1,000}-200}{12.64911} \leq c\right) \approx \Phi(c)$$

Now, if $S^*_{1,000}$ is the observed value of $S_{1,000}$, and if

$$\gamma = \frac{S^*_{1,000}-200}{12.64911}$$

then, on the basis of the central limit theorem and (1)we wouldn't expect that ■ is an atypical value for $N(0,1)$. In particular, we wouldn't expect that $|\gamma|$ is too big ; that is, we wouldn't expect that

$$P(|N(0,1)| \geq |\gamma|) < 0.05$$

if $\mu = 0.2$ is the true mean

Thus, we have the following basic statistical test

Statistical Test .Fix an $\alpha > 0$, typically $\alpha = 0.05$ or 0.01. Compute γ as in $\gamma = \frac{S^*_{1,000}-200}{12.64911}$. if

$$P(|N(0,1)| \geq |\gamma|) = 2\Phi(-|\gamma|) < \alpha$$

then we reject the hypothesis that the mean value of X_1 is μ ; and, if this inequality is not satisfied, we do not reject it, which is not the same as saying we accept it.

In the example given above we have that

$$\gamma = \frac{205-200}{12.64911} = 0.39528$$

and one can readily compute that

$$2\Phi\left(-0.39528\right) > 0.05$$

Thus, we do not reject the hypothesis that 20% of Georgians smoke

3.2 Noise Cancellation

Suppose that a man is driving through the desert, and runs out of gas. He grabs his cellphone to make a call for help, dialing 911, but he is just at the edge of the broadcast range for his cellphone, and so his call to the 911 dispatcher is somewhat noisy and garbled. Suppose that the 911 dispatcher has the ability to use several cellphone towers to clean up the signal. Suppose that there are about 100 towers near to the stranded driver, and suppose that the signals they each receive at a particular instant in time is given by

$$X_1, ..., X_{100}$$

where

$$Xi = S + Y_i$$

where S is the true signal being sent to the towers, and where Y_i is the noise. Suppose that all the noises $Y_1, ..., Y_{100}$ are independent and identically distributed, and further suppose they all have mean 0 and variance σ^2. Further, it is not unreasonable to assume that the noises are all normally distributed .

– i.e. they are all $N(0, \sigma^2)$ – though we will not need this assumption for what follows.

The 911 dispatcher cleans up the signal by computing the average

$$\bar{X} = \frac{X_1 + ... + X_{100}}{100} = S + \frac{Y_1 + ... + Y_{100}}{100}$$

Now, by the Central Limit Theorem, we would expect that

$$\frac{Y_1 + ... + Y_{100}}{100} \text{ is approximately } N(0, \sigma^2/100)$$

Of course we need to be careful here – the central limit theorem only applies for n large, and just how large depends on the underyling distribution of the random variables Y_i. There are more powerful versions of the central limit theorem, which give conditions on n under which $\frac{Y_1 + ... + Y_{100}}{100}$ is approximately $N(0, \sigma^2/100)$ holds under a precise notion of "is approximately". At any rate, if we assume that the Y_i's are all independent normal random variables, then we don't even need the central limit theorem, because in that case we have that $\bar{X} - S$ is exactly $N\left(0, \sigma^2/100\right)$

Now, suppose that, in fact, all the noises Y_i's have variance $\sigma^2 = 1$. Then, the central limit theorem in the guise $\frac{Y_1 + ... + Y_{100}}{100}$ is approximately $N(0, \sigma^2/100)$ would be telling us that the new noise $\bar{X} - S$ is approximately normal with variance $1/100$, a 100-fold improvement in the noise variance gotten just using one tower!

3.2.1 Just How Good is the Average Method for Noise Cancellation - Can we Do Better ?

It turns out that not only does averaging give us a pretty good way to cancel noise, but it is, in some sense, the best thing we could possibly try. The proper language is that taking the average $\bar{X}$ gives us a maximum liklihood estimate for the signal S, which is the same as the expected value of X_i for all $i = 1, ..., 100$. Let us make this more precise:

Maximum Liklihood Estimates. Suppose X is some random variable having a distribution that depends on a list of unknown, underlying parameters $\theta_1, ..., \theta_k$. Let $f\left(x; \theta_1, ..., \theta_k\right)$ denote the pdf for X , given the parameters $\theta_1, ..., \theta_k$. Suppose we make n independent observations of our random source X , and suppose these observations are the values $x_1, ..., x_n$. Then, the liklihood value of this observation is

$$L(\ x_1,...,x_n;\ \ \theta_1,...,\theta_k) = \Pi_{i=1}^n f\left(x_i;\theta_1,...,\theta_k\right)$$

We say that $\hat{\theta}_1,...,\hat{\theta}_k$ are maximum liklihood estimates for $\theta_1,...,\theta_k$, given the observations $x_1,...,x_n$, if these values $\hat{\theta}_i$ maximize $L(\ x_1,...,x_n;\theta_1,...,\theta_k)$. Note that the $\hat{\theta}_i$s which maximize L may not be unique (there may be more than one global max).

Question. Why the word 'liklihood', and not, say, 'probability'? To answer this, note that in the discrete setting it is easy to describe what the liklihood function computes: It is just the probability that given particular values for $\theta_1,...,\theta_k$, the observed values for some random variable X were $x_1,...,x_n$ So, in this case 'liklihood' and 'probability' coincide. However, as we well know, the pdf for a continuous random variable X does not give us probability values when we plug in varlues for x , and hence the use of the word 'liklihood'.

In our case, let us suppose that the received signals X_i are, in fact, normal, with mean S, and variance σ^2 ; that is to say, the noises Y_i are $N\left(0,\sigma^2\right)$. Now suppose that we have definite values for these observations (that is, our observed signals are 'instantiated'), and suppose that those values are $x_1,...,x_{100}$. The liklihood function here is

$$L\left(x_1,...,x_n;S,\sigma^2\right) = \frac{1}{(2\pi)^{50}\sigma^{100}}\exp\left(-\left[(x_1-S)^2+...+(x_{100}-S)^2\right]/2\sigma^2\right)$$

If we seek S which maximizes this (for any given value for σ^2), we can ignore the factor $(2\pi)^{50}\sigma^{100}$, and we maximize the log of the remaining exponential factor; thus, we just need to maximize

$$-\frac{(x_1-S)^2+...+(x_{100}-S)^2}{2\sigma^2}$$

We can ignore the σ^2; so, we maximize

$$-\frac{(x_1-S)^2+...+(x_{100}-S)^2}{2}$$

Taking a derivative with respect to S and setting equal to 0 we have

$$(x_1+...+x_{100})-100S=0$$

so

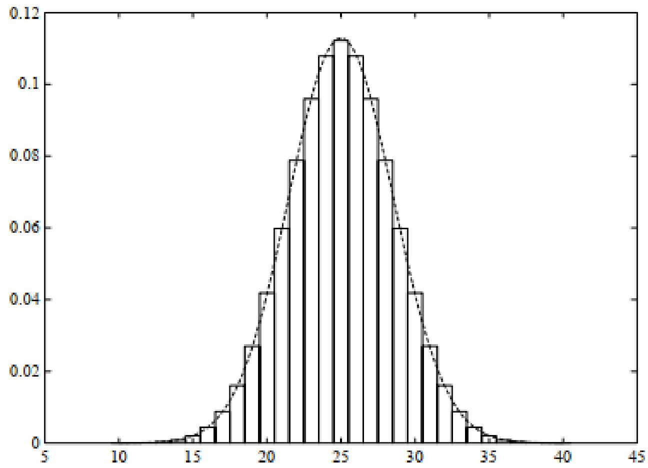
$$S=\frac{x_1+...+x_{100}}{100}$$

which is our sample mean. The fact that the expression $-\frac{(x_1-S)^2+...+(x_{100}-S)^2}{2}$ is a down-turning parabola means that, indeed, this is a maximum.

Thus, we see that by averaging we obtain a maximum liklihood estimate for S, and therefore, in some sense, this is the best we could hope to do to recover S.

3.3 Graphic Study

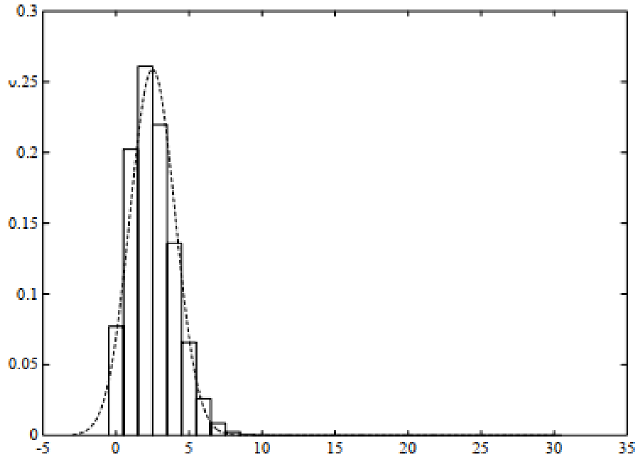
Let $X_1,...,X_n$ be n independent Bernoulli random variable with the same parameter p. We pose $S_n= X_1+....X_n$. The law of S_n is then a binomial $B(n,p)$:



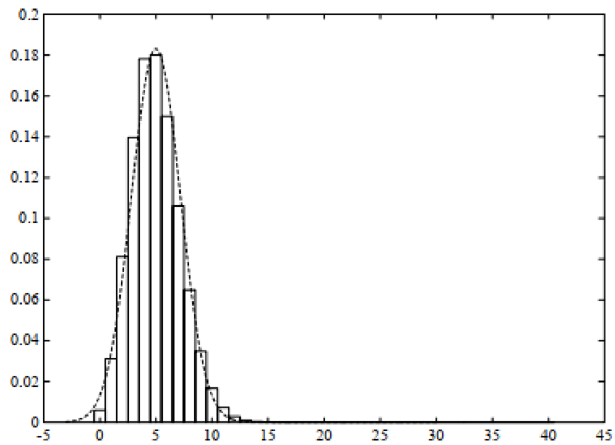
Binomial $n = 50$, $p = 0{,}5$ ($\varepsilon = 3{,}1 \times 10^{-4}$)

$$P\left(S_n = k\right) = C_n^k p^k \left(1-p\right)^{n-k} \qquad (0 \leq k \leq n)$$

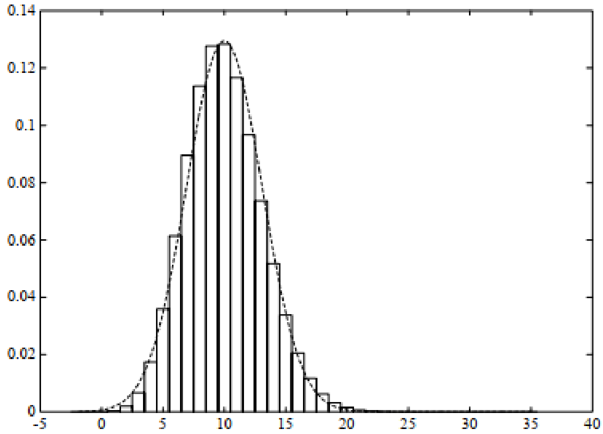
The histograms below represent this law for different values of the parameters n and p . Although the histogram of the distribution $B\left(n,p\right)$ theoretically consists of $n+1$ rectangles, only a part of them is visible in the drawing, the others correspond giving too small probabilities. The number ε represents for each figure aincreasing by the probability corresponding to the union of these"invisible" rectangles.In each figure, the dotted curve represents the density $f_{m,\sigma}$ whose parameters are given by: $m = E\left[S_n\right] = np$ and $\sigma^2 = Var\left(S_n\right) = npq$



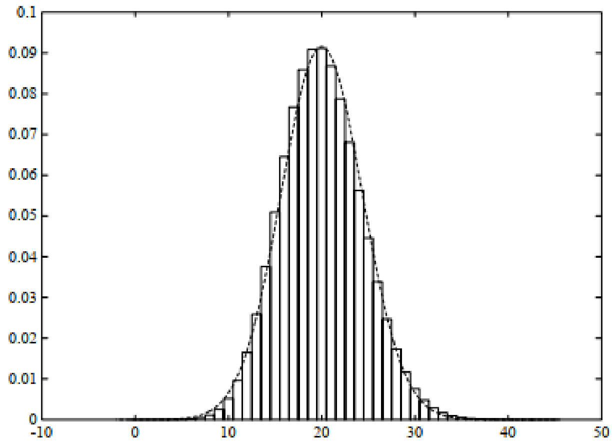
Binomial $n = 50,\ p = 0{,}05\ (\varepsilon = 1{,}6 \times 10^{-4})$



Binomial $n = 100,\ p = 0{,}05\ (\varepsilon = 1{,}4 \times 10^{-4})$



Binomial $n = 200$, $p = 0{,}05$ ($\varepsilon = 2{,}3 \times 10^{-4}$)



Binomial $n = 400$, $p = 0{,}05$ ($\varepsilon = 4{,}9 \times 10^{-4}$)

Appendix A

Abbreviations and Notations

X,Y,...random variables

x,y,... values of the random variables **X,Y,...**

f(x) value of the probability density function of a continuous random variable **X**

F(x) value of the cumulative distribution function of a continuous random variable **X**

E[X] expectation of the random variable **X**

E[f(X)] expectation of **f(X)**

Var(X) variance of the random variable **X**

$\mathbf{e}^x$,exp**(x)** exponential function of **x**

n! n factorial

Bern(p) bernoulli distribution with parameter **p**

$\mathbf{M}_X\mathbf{(t)}$ moment genereting function for the random variable **X**

$\mathbf{N}(\boldsymbol{\mu}, \boldsymbol{\sigma}^2)$ normal distribution with mean $\boldsymbol{\mu}$ and variance $\boldsymbol{\sigma}^2$

$\boldsymbol{\mu}$ population mean

$\boldsymbol{\sigma}^2$ population variance

$\boldsymbol{\sigma}$ population standard deviation

$\bar{\mathbf{x}}$ sample mean ,$\bar{\mathbf{x}} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i$

$\mathbf{s}^2$ unbiased estimate of population variance from a sample, $\mathbf{s}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$

$\boldsymbol{\phi}$ probability density function of the standardised normal variable $Z \sim N(0,1)$

$\boldsymbol{\Phi}$ cumulative distribution function of the standardised normal variable $Z \sim N(0,1)$

$\hat{\boldsymbol{\theta}}$ is an estimator for $\boldsymbol{\theta}$

$\mathbf{H}_0, \mathbf{H}_1$ null and alternative hypotheses for a hypothesis test

a.s. almost surely

i.i.d. independent and identically distributed

pdf probability density function

pmf probability mass function

r.v. random variable

CLT central limit theorem

Conclusion

The advantage of the CLT is that it is powerful, meaning implying that regardless of whether the data originates from an assortment of distributions if their mean and variance are the equivalent, the theorem can even now be utilized.

The CLT notes that the sample means converge on the population means and the distance between them converges to be normally distributed with a variance equal to the population variance as the sample size increases. It is important in the application of statistics and in the understanding of nature.

References

[1] Introductory Probability and the Central Limit Theorem Vlad Krokhmal 07/29/2011

[2] Probability and Random Processes by Venkatarama Krishnan .2006 Edition

[3] A First Course in Probability by Sheldon Ross . Sixth Edition

[4] Théorème Limite Central . Charles Suquet

Lille University of Science and Technology ,Faculty of Pure and Applied Mathematics ,Build. M2, F-59655 Villeneuve d'Ascq Cedex

[5] P.BILLINGSLEY. Probability and measure . Wiley, third edition 1995

[6] W.FELLER. An Introduction to Probability Theory and its Applications,Vol . I.

[7] W.FELLER. An Introduction to Probability Theory and its Applications,Vol . II.

[8] V. V. PETROV . Sums of Independent Random Variables. Spring,1975

[9] P.S. TOULOUSE. Thèmes de Probabilités et statistique . Dunod, 1999.

[10] J. V. USPENSKY,Introduction to mathematical probability. McGraw-Hill,1937

[11] http://people.math.gatech.educ/~ecroot/3770/central limit apps.pdf

[12] http://towardsdatascience.com

Abstract

The Central Limit Theorem establishes the convergence in law of the sum of a series of random variables to the normal distribution. Intuitively, this result asserts that a sum of identical and independent random variables tends towards a Gaussian random variable.

Keywords: random variables, normal distribution, Bernoulli's law.

Résumé

Le théorème de limite centrale établit la convergence en loi de la somme d'une série de variables aléatoires à la distribution normale. Intuitivement, ce résultat affirme qu'une somme de variables aléatoires identiques et indépendantes tend vers une variable aléatoire gaussienne.

Mots clés : variables aléatoires, distribution normale, loi de Bernoulli.

ملخص

تؤسس نظرية الحد المركزي التقارب في القانون (التقارب في التوزيع) لمجموع سلسلة من المتغيرات العشوائية للتوزيع الطبيعي. بشكل حدسي ، تؤكد هذه النتيجة أن مجموع المتغيرات العشوائية المتطابقة والمستقلة يميل نحو متغير عشوائي غوسي .