

UNIT 5. Joint analysis and limit theorems.

5.1.- n-dimension distributions. Marginal and conditional distributions

5.2.- Sequences of independent random variables. Properties

5.3.- Sums of random variables

5.4.- Central Limit Theorem and its applications



UNIT 5. OBJECTIVES

- To apply the main properties derived from independence of random variables.
- To calculate probabilities for the main aggregates of independent random variables.
- To apply and interpret the Central Limit Theorem.



UNIT 5. Joint analysis and limit theorems.

5.1.- n-dimension distributions. Marginal and conditional distributions.



Joint distribution of two random variables

CARS The variables *number of children in the family (X)* and *number of cars in the family (Y)* have been analysed jointly:

Y \ X	1	2	3
1	0.19	0.16	0.09
2	0.06	0.19	0.31

❖ Pairs (x_i, y_j)

❖ Joint probabilities p_{ij}

$$P(X=2, Y=1)=0.16$$



Discrete bidimensional random variables

Joint observation of two discrete random variables X and Y .

(X,Y) DISCRETE RVs

$\{(x_i, y_j) / i, j = 1, 2, \dots\}$

$$P:(x,y) \in \mathbb{R}^2 \rightarrow P(x,y) = P(X=x, Y=y) \in [0,1]$$

$Y \setminus X$	x_1	...	x_i	...	x_k
y_1	p_{11}		p_{i1}		p_{k1}
...					
y_j	p_{1j}		p_{ij}		p_{kj}
...					
y_l	p_{1l}		p_{il}		p_{kl}

Joint probability
 $p_{ij} = P(X=x_i, Y=y_j)$

$$P(X = x_i, Y = y_j) = p_{ij}$$

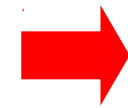
$$\sum_i \sum_j p_{ij} = 1$$

$$p_{ij} \geq 0$$

Marginal distributions

$Y \setminus X$	x_1	...	x_i	...	x_k
y_1	p_{11}		p_{i1}		p_{k1}
...					
y_j	p_{1j}		p_{ij}		p_{kj}
...					
y_l	p_{1l}		p_{il}		p_{kl}

**Marginal distribution
of Y: $(y_j, p_{.j})$**



$$p_{.j} = \sum_i p_{ij}$$



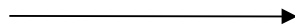
$$p_{i.} = \sum_j p_{ij}$$

**Marginal distribution
of X: $(x_i, p_{i.})$**



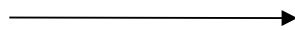
Marginal distributions

The marginal distributions are univariate distributions,
so

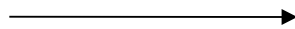


PROBABILITIES

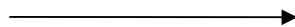
and all the characteristics of X and Y may be computed
in a straightforward way:



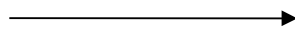
$E(X)$



$Var(X)$



$E(Y)$



$Var(Y)$



Continuous bidimensional variables

Joint observation of two continuous Rvs, X and Y.

(X,Y) CONTINUOUS RVs

Joint density function

$$f:(x,y) \in \mathbb{R}^2 \rightarrow f(x,y) \in \mathbb{R}$$

$$f(x,y) \geq 0 \quad \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x,y) dx dy = 1$$

Marginal density functions

$$f_Y(y) = \int_{-\infty}^{+\infty} f(x,y) dx$$

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y) dy$$



Joint distribution function

For any bidimensional RV (X,Y) , its distribution function is defined as follows:

**Joint
distribution
function**

$$F:(x,y) \in \mathbb{R}^2 \rightarrow F(x,y) = P(X \leq x, Y \leq y) \in [0,1]$$

[The value of the joint distribution function at (x,y) is the total probability accumulated up to point (x,y) .]



Linear correlation

Covariance

$$\sigma_{X,Y} = \text{Cov}(X,Y) = E[(X - \mu_X)(Y - \mu_Y)]$$

$$\sigma_{X,Y} = E(XY) - E(X)E(Y)$$

Linear correlation coefficient

$$\rho_{X,Y} = \frac{\sigma_{X,Y}}{\sigma_X \sigma_Y} \quad -1 \leq \rho_{X,Y} \leq 1$$

- The sign of $\rho_{X,Y}$ indicates the type of correlation (direct, inverse).
- The absolute value of $\rho_{X,Y}$ indicates the intensity of correlation.



Random vectors. Characteristics

Vector

Expectation

Variances-
Covariances

Univariate
 X

$$\mu = E(X)$$

$$\sigma^2 = \text{Var}(X)$$

Bivariate
 (X, Y)

$$\mu = (E(X), E(Y))$$

$$\begin{pmatrix} \sigma_X^2 & \sigma_{XY} \\ \sigma_{YX} & \sigma_Y^2 \end{pmatrix}$$

n-variate
 $(X_1, X_2, \dots, X_n)$

$$\mu = (E(X_1), \dots, E(X_n))$$

$$\begin{pmatrix} \sigma_1^2 & \sigma_{12} & \dots & \sigma_{1n} \\ \sigma_{21} & \sigma_2^2 & \dots & \sigma_{2n} \\ \dots & \dots & \dots & \dots \\ \sigma_{n1} & \sigma_{n2} & \dots & \sigma_n^2 \end{pmatrix}$$



UNIT 5. Joint analysis and limit theorems.

5.2.- Sequences of independent random variables. Properties



Independent random variables

Two RVs X and Y are **independent** iff:

$$F(x,y) = F_X(x) F_Y(y) \quad \forall (x,y) \in \mathbb{R}^2$$

Joint distribution
function of (X,Y)

Marginal distribution
function of X

Marginal
distribution
function of Y

INDEPENDENCE CONDITION

(X,Y) DISCRETE

$$P(x_i, y_j) = P(x_i) P(y_j) \quad \forall (x_i, y_j) \in \mathbb{R}^2$$

(X,Y) CONTINUOUS

$$f(x,y) = f_X(x) f_Y(y) \quad \forall (x,y) \in \mathbb{R}^2$$



Properties of functions of independent RVs

Given X and Y , **independent** RVs, it holds:

□ $E(XY) = E(X)E(Y)$.

□ $\text{Cov}(X, Y) = 0$ and $\rho_{XY} = 0$.

□ $\text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y)$.

□ $\text{Var}(aX+bY) = a^2\text{Var}(X) + b^2\text{Var}(Y), \quad \forall a, b \in \mathbb{R}$.

□ $\text{Var}(X-Y) = \text{Var}(X) + \text{Var}(Y)$.

If X and Y are **uncorrelated normal RVs** ($\rho_{XY} = 0$), then

X and Y are **independent**



Reproductivity

A class ψ of RVs is reproductive if for every independent $X_1, X_2 \in \psi$ it holds $X_1 + X_2 \in \psi$

Independent RVs

Sum

Reproductivity

$$X \approx \mathbf{B}(\mathbf{n}_X, p)$$

$$Y \approx \mathbf{B}(\mathbf{n}_Y, p)$$

$$X + Y \approx \mathbf{B}(\mathbf{n}_X + \mathbf{n}_Y, p)$$

The binomial model is reproductive in n (for any fixed p)

$$X \approx \mathbf{N}(\mu_X, \sigma_X)$$

$$Y \approx \mathbf{N}(\mu_Y, \sigma_Y)$$

$$X + Y \approx \mathbf{N}(\mu_X + \mu_Y, \sqrt{\sigma_X^2 + \sigma_Y^2})$$

The normal models is reproductive in mean and variance

$$X \approx \mathbf{P}(\lambda_X)$$

$$Y \approx \mathbf{P}(\lambda_Y)$$

$$X + Y \approx \mathbf{P}(\lambda_X + \lambda_Y)$$

The Poisson model is reproductive in λ

UNIT 5. Joint analysis and limit theorems.

5.3.- Sums of random variables



Sum of random variables

$$\left\{ \begin{array}{l} X_1 \text{ (Profits of firm 1)} \\ X_2 \text{ (Profits of firm 2)} \\ \dots\dots\dots \\ X_n \text{ (Profits of firm } n) \end{array} \right.$$

Sum

$$S_n = \sum_{i=1}^n X_i$$

Total profit of the firms

Mean

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

Mean profit of the firms



Characteristics

Given an n-dimension RV $(X_1, X_2, \dots, X_n)$ with finite expectations and variances, namely:

X_1	X_2	...	X_n
$E(X_1)=\mu_1$	$E(X_2)=\mu_2$	...	$E(X_n)=\mu_n$
$Var(X_1)=\sigma_1^2$	$Var(X_2)=\sigma_2^2$	...	$Var(X_n)=\sigma_n^2$

The following holds:

Expectation

$$E(S_n) = \sum_{i=1}^n \mu_i$$

$$E(\bar{X}_n) = \frac{\sum_{i=1}^n \mu_i}{n}$$

Variance

$$Var(S_n) = Var\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \sigma_i^2 + \sum_{i \neq j} \sigma_{ij}$$

If $X_1, X_2, \dots, X_n$ are independent:

$$Var(S_n) = \sum_{i=1}^n \sigma_i^2$$

$$Var(\bar{X}_n) = \frac{\sum_{i=1}^n \sigma_i^2}{n}$$



Tchebyshev's Inequality

Given $(X_1, X_2, \dots, X_n)$, an n -dimension RV with finite expectations and variances, it holds:

Tchebyshev's bound for the sum:

$$P\left(|S_n - E(S_N)| \geq \varepsilon\right) \leq \frac{Var(S_n)}{\varepsilon^2}$$

Tchebyshev's bound for the mean:

$$P\left(|\bar{X}_n - E(\bar{X}_n)| \geq \varepsilon\right) \leq \frac{Var(\bar{X}_n)}{\varepsilon^2}$$



Independent identically distributed RVs

Given $X_1, X_2, \dots, X_n$ **independent identically distributed (iid) RVs**
with $E(X_i) = \mu$ **and** $\text{Var}(X_i) = \sigma^2, i = 1, \dots, n$, **we have:**

RV	EXPECTED VALUE	VARIANCE	TCHEBYSHEV'S BOUND
Sum (S_n)	$E(S_n) = n\mu$	$\text{Var}(S_n) = n\sigma^2$	$P(S_n - n\mu \geq \varepsilon) \leq \frac{n\sigma^2}{\varepsilon^2}$
Mean ($\bar{X}_n$)	$E(\bar{X}_n) = \mu$	$\text{Var}(\bar{X}_n) = \frac{\sigma^2}{n}$	$P(\bar{X}_n - \mu \geq \varepsilon) \leq \frac{\sigma^2}{n\varepsilon^2}$



UNIT 5. Joint analysis and limit theorems.

5.4.- The Central Limit Theorem and its applications



Central Limit Theorem (CLT)

Levy-Lindeberg's CLT :

Given $X_1, X_2, \dots, X_n$, iid random variables with finite $E(X_i) = \mu$ and $\text{Var}(X_i) = \sigma^2$, $i = 1, 2, \dots, n$, the following holds as $n \rightarrow \infty$:

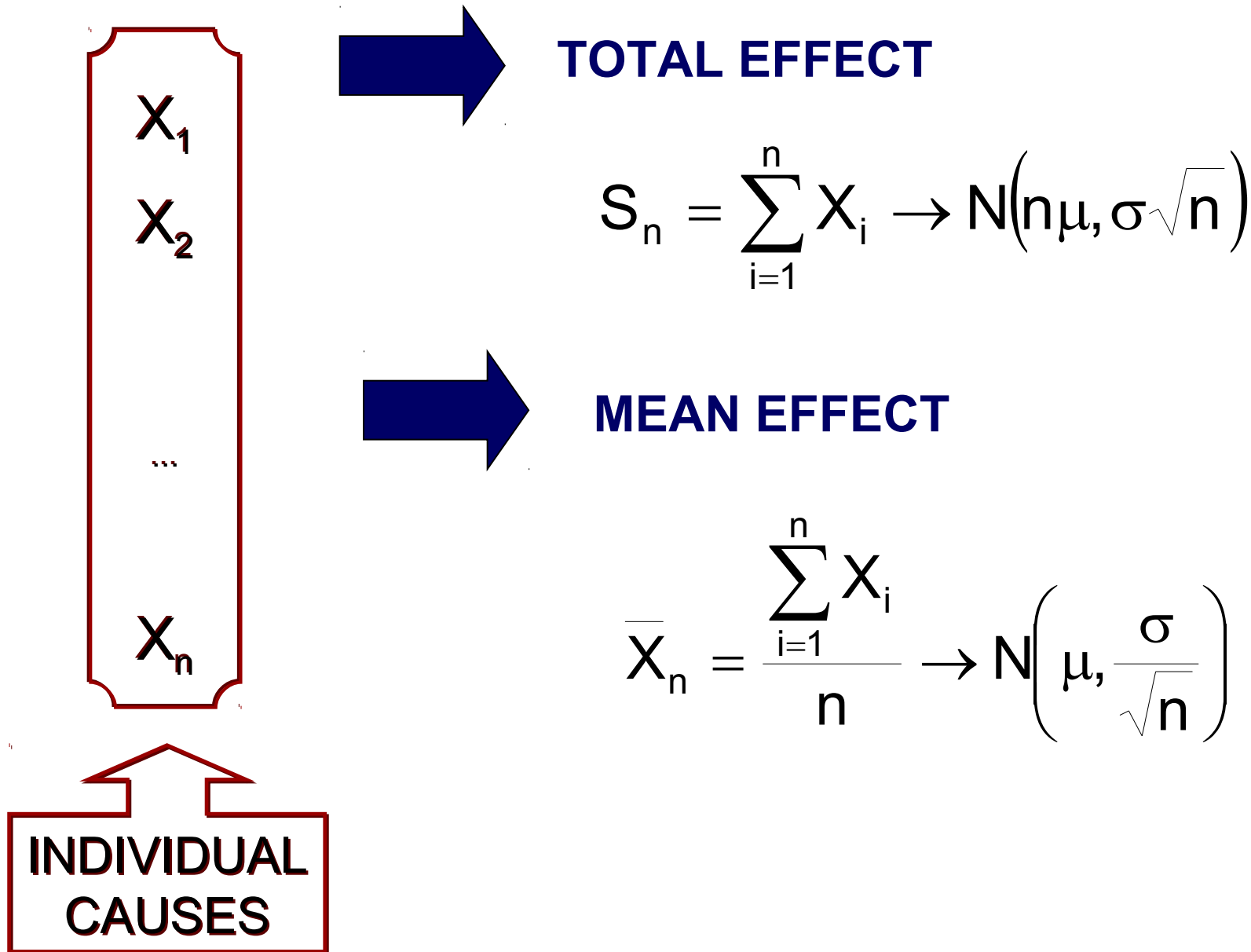
$$S_n = \sum_{i=1}^n X_i \longrightarrow N(n\mu, \sigma\sqrt{n})$$

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} \longrightarrow N(0, 1)$$

This approximation is accurate for $n \geq 30$



Central Limit Theorem



Central Limit Theorem

De Moivre's Theorem:

Given $X_1, X_2, \dots, X_n$, independent **Bernoulli** RVs with parameter p , the following holds as $n \rightarrow \infty$:

$$\frac{\sum_{i=1}^n X_i - np}{\sqrt{npq}} \rightarrow N(0, 1)$$

➔ Consequence:

$$X_i \sim \text{Be}(p)$$

$$\sum_{i=1}^n X_i \sim B(n, p) \rightarrow N(np, \sqrt{npq})$$

A continuity correction improves the accuracy of this approximation



Approximation BINOMIAL $\rightarrow$ NORMAL

