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

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

Y \ X	x_1	...	x_i	...	x_k
y_1	p_{11}		p_{i1}		p_{k1}
...					
y_j	p_{1j}		$\boxed{p_{ij}}$		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 \ 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

$$f:(x,y) \in R^2 \to f(x,y) \in R$$

$$f(x,y) \geq 0 \qquad \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,dxdy = 1$$

Marginal density functions

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

Joint distribution function

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

Joint distribution function

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

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

Linear correlation

Covariance

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

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

Linear correlation coefficient

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

- 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	$\mu = E(X)$	$\sigma^2 = \mathrm{Var}(X)$
Bivariate (X,Y)	$\mu = \left(E(X), \mathrm{E}(Y)\right)$	$\begin{pmatrix} \sigma_X^2 & \sigma_{XY} \\ \sigma_{YX} & \sigma_Y^2 \end{pmatrix}$
n-variate $(X_1, X_2, \ldots, X_n)$	$\mu = \left(E(X_1), \ldots, \mathrm{E}(X_n)\right)$	$\begin{pmatrix} \sigma_1^2 & \sigma_{12} & \ldots & \sigma_{1n} \\ \sigma_{21} & \sigma_2^2 & \ldots & \sigma_{2n} \\ \ldots & \ldots & \ldots & \ldots \\ \sigma_{n1} & \sigma_{n2} & \ldots & \sigma_n^2 \end{pmatrix}$

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:

$$F(x,y) = F_X(x)\, F_Y(y) \qquad \forall (x,y) \in \Re^2$$

Joint distribution function of (X,Y)

Marginal distribution function of X

Marginal distribution function of Y

INDEPENDENCE CONDITION

(X,Y) DISCRETE $\quad P(x_i,y_j)=P(x_i)P(y_j) \qquad \forall (x_i,y_j) \in \Re^2$

(X,Y) CONTINUOUS $\quad f(x,y)=f_X(x)f_Y(y) \qquad \forall (x,y) \in \Re^2$

Properties of functions of independent RVs

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

- $E(XY) = E(X)E(Y)$.
- $Cov(X,Y) = 0$ and $\rho_{XY}=0$.
- $Var(X+Y) = Var(X) + Var(Y)$.
- $Var(aX+bY) = a^2Var(X) + b^2Var(Y), \quad \forall a,b \in IR$.

$Var(X-Y) = Var(X) + Var(Y)$.

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
$X \approx B(n_X, p)$ $Y \approx B(n_Y, p)$	$X+Y \approx B(n_X+n_Y, p)$	The binomial model is reproductive in n (for any fixed p)
$X \approx N(\mu_X, \sigma_X)$ $Y \approx N(\mu_Y, \sigma_Y)$	$X+Y \approx N(\mu_X+\mu_Y, \sqrt{\sigma_X^2+\sigma_Y^2})$	The normal models is reproductive in mean and variance
$X \approx P(\lambda_X)$ $Y \approx P(\lambda_Y)$	$X+Y \approx P(\lambda_X+\lambda_Y)$	The Poisson model is reproductive in λ

UNIT 5. Joint analysis and limit theorems.

5.3.- Sums of random variables

Sum of random variables

X_1 (Profits of firm 1)

X_2 (Profits of firm 2)

....

X_n (Profits of firm n)

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, …, X_n) with finite expectations and variances, namely:

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

The following holds:

Expectation

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

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, ..., X_n are independent:

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

Tchebyshev's Inequality

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

Tchebyshev's bound for the sum:

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

Tchebyshev's bound for the mean:

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

Independent identically distributed RVs

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

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

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, ..., X_n, iid random variables with finite $E(X_i) = \mu$ and $Var(X_i) = \sigma^2$, i = 1, 2, …, n, the following holds as $n \to \infty$:

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

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

This approximation is accurate for $n \geq 30$

Central Limit Theorem

X_1

X_2

...

X_n

INDIVIDUAL CAUSES

TOTAL EFFECT

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

MEAN EFFECT

$$\overline{X}_n = \frac{\sum_{i=1}^{n} X_i}{n} \rightarrow N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

Central Limit Theorem

De Moivre's Theorem:

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

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

➡Consequence:

$$X_i \sim Be(p)$$

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

A **continuity correction** improves the accuracy of this approximation

Approximation BINOMIAL → NORMAL