

Colección  
**Estadística y Análisis de Datos**

**5**

*Dirigida por*  
Carles M. Cuadras

# CURS DE PROBABILITATS

DAVID NUJALART i MARTA SANZ  
Facultat de Matemàtiques, Universitat de Barcelona

PPU  
Barcelona, 1990

Aquest llibre presenta de manera autocontinguda les idees essencials de la Teoria de la Probabilitat, fonamentades en la Teoria de la Mesura, que podrien desenvolupar-se al llarg d'un curs quatrimestral. S'adreça fonamentalment a estudiants de Matemàtiques, o en general, a lectors amb bons coneixements de Càlcul diferencial i integral amb diverses variables.

Agraïm la col·laboració de tots aquells que han contribuït d'alguna manera en la seva realització.

Barcelona, Setembre 1990

Primera edició, 1990

La reproducció total o parcial d'aquesta obra per qualsevol procediment, compresos la reproductiva i el tractament informàtic i la distribució d'exemplars mitjançant lloguer o préstec públics, resta rigorosament prohibida sense l'autorització escrita dels titulars del "Copyright", i estarà sotmessa a les sancions estableties a la Llei.

© David Nualart i Marta Sanz

© PPU  
Promociones y Publicaciones Universitarias, S.A.  
Marques de Campo Sagrado, 16  
08015 Barcelona

I.S.B.N.: 84-7665-718-8  
D.L.: L-1918-90

Impres a: Pobligràfic, S.A.  
Av. Estació, s/n.  
Pobla de Segur (Lleida)

*... Il y a encore bien plus de liberté au-delà des  
mondes où les ponts sont coupés. Elle a le sens  
de la lumière gazeuse et du fijre.*

*Tristan Tzara*

## CONTINGUT

1. Introducció	1
2. Espais de probabilitat	4
3. Probabilitat condicionada i independència d'esdeveniments	7
4. Extensió de mesures	11
5. Probabilitats en la recta real i funcions de distribució	22
6. Variables aleatòries. Esperança matemàtica	30
7. Producte d'espais de probabilitat. Independència. Distribucions condicionades	47
8. La distribució binomial. Passejada aleatòria sobre els enters	66
9. Convergència de variables aleatòries	78
10. Lleis dels grans nombres	90
11. Convergència feble de probabilitats	100
12. Funcions característiques	110
13. El Teorema central del límit	123
Referències	131

## 1. INTRODUCCIÓ

L'objecte d'estudi de la teoria de la probabilitat són les *experiències aleatòries*. Considerem la realització d'una certa experiència, com per exemple, tirar un dau, una moneda, o efectuar un cert experiment físic o químic. L'experiència queda determinada per un conjunt de *condicions*. Direm que l'experiència és aleatòria si les condicions sota les quals la realitzem no ens permeten precisar "a priori" el seu resultat. Tindrem aleshores un conjunt  $\Omega$  de possibles resultats. Per exemple, si l'experiència consisteix en tirar un dau i observar el número que surt, el conjunt de resultats serà  $\Omega = \{1, 2, 3, 4, 5, 6\}$ .

D'altra banda a tota experiència aleatòria li podem associar un conjunt d'esdeveniments aleatoris que es poden produir (o no) en realitzar l'experiència. Per exemple, "que surti un tres" o "que surti un nombre parell" són esdeveniments aleatoris en el cas de la tirada d'un dau o "que el resultat numèric estigui comprès entre dos valors donats  $a$  i  $b$ " en el cas d'un experiment científic.

A cada esdeveniment aleatori  $\hat{A}$  associat a l'experiència que considerem, li correspon un subconjunt de resultats  $A$  d' $\Omega$  format pels resultats que produeixen l'esdeveniment  $\hat{A}$ . Per exemple, en el cas del llançament d'un dau, a l'esdeveniment  $\hat{A} =$  "que surti un nombre parell" li correspon el conjunt de resultats  $A = \{2, 4, 6\}$ .

Es diu que una família de subconjunts d' $\Omega$ ,  $\mathcal{A} \subset \mathcal{P}(\Omega)$  és una *àlgebra* si  $\Omega \in \mathcal{A}$  i si  $\mathcal{A}$  és estable per unions finites i per complementació.

Aleshores, si representem per  $\mathcal{A}$  la família dels conjunts associats a tots els esdeveniments aleatoris possibles en una certa experiència aleatòria, sembla raonable suposar que  $\mathcal{A}$  té estructura d'àlgebra. En efecte, si  $A$  i  $B$  són dos subconjunts d' $\Omega$  associats als esdeveniments  $\hat{A}$  i  $\hat{B}$ , aleshores  $A \cup B$  és el conjunt de resultats de l'esdeveniment " $\hat{A}$  ó  $\hat{B}$ " i  $A \cap B$  és el conjunt de resultats de l'esdeveniment " $\hat{A}$  i  $\hat{B}$ ". També,  $A^c$  està associat a l'esdeveniment "no  $\hat{A}$ " i els conjunts  $\Omega, \emptyset$  es corresponen amb els esdeveniments "segur" i "impossible", respectivament. A partir d'ara identificarem un esdeveniment  $\hat{A}$  amb el conjunt  $A \subset \Omega$  de resultats associat.

En aquests tipus d'experiències no tan sols podem dir que en cert esdeveniment  $\hat{A}$  és aleatori, sino que ens interessa poder donar una estimació quantitativa de la possibilitat que  $\hat{A}$  es produixi. Això ens porta a introduir la *probabilitat* d'un esdeveniment  $\hat{A}$  com un

cert valor numèric  $P(A)$ , compres entre 0 i 1, associat de forma objectiva a l'esdeveniment. Es planteja aleshores el problema de *definir i calcular la probabilitat d'un esdeveniment aleatori*.

La idea de que la probabilitat d'un esdeveniment aleatori  $A$ , admèti una evaluació quantitativa mitjançant un nombre  $p = P(A)$ , va ésser desenvolupada de forma sistemàtica per primera vegada, en el segle XVII, en els treballs de Fermat (1601–1665), Pascal (1623–1662), Huygens (1629–1695) i, en particular, Jakob Bernoulli (1654–1705). Les seves investigacions, que venien motivades essencialment pels jocs d'atzar de l'època, varen fomentar el càlcul de probabilitats.

Un cert nombre de definicions matemàtiques de probabilitat han estat proposades per diversos autors, al llarg de la història. Cal dir que, com és habitual en l'evolució de la ciència, l'interès per a una fonamentació lògica de la teoria de la probabilitat apareix històricament més tard que l'habilitat per a determinar probabilitats i dur a terme càlculs amb aquestes probabilitats, utilitzant els resultats d'aquests càlculs en problemes pràctics i en la investigació científica.

Deixant de banda les definicions "subjectives" de probabilitat com a mesura quantitativa del "grau de certesa" de l'observador, hi ha dues definicions que cal tenir en compte:

- (1) La *definició frequentista* de probabilitat com el límit de les freqüències relatives de l'esdeveniment aleatori  $A$  en una sèrie llarga de realitzacions de l'experiència. Suposem que  $\nu_n(A)$  és el nombre de vegades que s'ha produït  $A$  en una sèrie de  $n$  realitzacions de l'experiència repetides sota les mateixes condicions. Aleshores la *frequència relativa* es defineix com

$$f_n(A) = \frac{1}{n} \nu_n(A).$$

Quan  $n$  es fa gran, s'observa que  $f_n(A)$  s'aproxima a un valor constant, que per definició serà la probabilitat de l'esdeveniment  $A$ .

Aquest tipus de *regularitat estadística* va ésser observat per primera vegada en fenòmens demogràfics: A Xina, l'any 2238 A. C., en base als cens de població, es va veure que la proporció de nens nats en un any respecte al nombre total de naixements, en una gran ciutat, romania pròxima a 0.5 d'anys en any. Més tard, particularment en els segle XVII i XVIII, estudis més profunds sobre les estadístiques de poblacions permeteren retrobar regularitats estadístiques d'altres característiques. Aquesta definició empírica de probabilitat fou desenvolupada per Von Mises (1883–1953). La idea de Von Mises consisteix en considerar successions infinites de resultats

d'experiències aleatories (anomenades col·lectius), que presentin una certa regularitat estadística. La construcció d'una teoria matemàtica per aquests collectius va trobar dificultats lògiques insuperables.

- (2) La *definició clàssica* de probabilitat, introduida per Laplace l'any 1812, es basa en el concepte d'*equiprobabilitat* com a propietat objectiva dels diversos esdeveniments elementals associats a una experiència. Aquesta propietat d'equiprobabilitat es justifica per raons de simetria.

És a dir, si suposem que  $\Omega = \{\omega_1, \omega_2, \dots, \omega_n\}$  és finit i  $A = \mathcal{P}(\Omega)$ , els esdeveniments elementals (o signifiquen els singletons  $\{\omega_i\}$ ) tindran tots la mateixa probabilitat igual a  $\frac{1}{n}$ , i per definició, la probabilitat d'un esdeveniment  $A$  valdrà

$$P(A) = \frac{\text{card}(A)}{n} = \frac{\text{nombre de resultats "favorables"} \text{ a } A}{\text{nombre total de resultats}}$$

Per exemple, en el llançament d'un dau, la probabilitat de l'esdeveniment  $A =$  "que surti un nombre parell" serà  $P(A) = P(\{2, 4, 6\}) = \frac{1}{2}$ .

Aquesta definició no es pot considerar com una veritable definició matemàtica ja que comporta un "cercle viciós": el terme que es vol definir (probabilitat) apareix en la pròpia definició (resultats equiprobables).

- D'altra banda, es fàcil considerar experiències aleatories on els resultats elements no són equiprobables, com els dos exemples següents:
- Exemple 1.1.** Considerem una ruleta on el cercle giratori està dividit en un cert nombre de sectors circulars de diferents colors i amplituds:
- En aquest cas, la probabilitat de cada color és proporcional a l'angle del sector.

**Exemple 1.2.** Llançem dues monedes idèntiques i observem el nombre de cares que surten. El conjunt de resultats possibles és  $\Omega = \{0, 1, 2\}$  i, en aquest cas  $P(\{0\}) = P(\{2\}) = \frac{1}{4}$ ,  $P(\{1\}) = \frac{1}{2}$ .

La fonamentació teòrica de la probabilitat va haver de recórrer a una construcció axiomàtica, tal i com ja s'havia fet en d'altres parts de la matemàtica com la geometria, la topologia... L'*axiomàtica de Kolmogorov* va ésser publicada l'any 1933 i relaciona la teoria de la probabilitat amb la teoria de la mesura que havia estat recentment desenvolupada per Borel i Lebesgue. En el següent apartat introduirem els *espais de probabilitat* segons els axiomes de Kolmogorov.

## Propietats elementals de les $\sigma$ -àlgebres

- 1.- Una  $\sigma$ -àlgebra  $\mathcal{A}$  és estable per interseccions numerables.
- 2.- La  $\sigma$ -àlgebra més petita és  $\{\emptyset, \Omega\}$  i la més gran  $\mathcal{P}(\Omega)$ .
- 3.- La intersecció d'una família qualsevol de  $\sigma$ -àlgebres és també una  $\sigma$ -àlgebra.
- 4.- Donada una família  $\mathcal{C}$  de parts d' $\Omega$ , designarem per  $\sigma(\mathcal{C})$  la  $\sigma$ -àlgebra generada per  $\mathcal{C}$ , que per definició serà la intersecció de totes les  $\sigma$ -àlgebres que contenen  $\mathcal{C}$ .

**Definició 2.1.** Un *espai de probabilitat* és una terna  $(\Omega, \mathcal{A}, P)$ , tal que:

- (i)  $\Omega$  és un conjunt format per totes les possibles realitzacions o resultats del fenòmen aleatori que estudiem,
- (ii)  $\mathcal{A}$  és una família de parts d' $\Omega$  que té estructura de  $\sigma$ -àlgebra, és a dir que compleix les següents propietats:

$\Omega \in \mathcal{A}$ , i  $\mathcal{A}$  és estable per complementació i per unions numerables.

(iii)  $P$  és una aplicació,  $P : \mathcal{A} \rightarrow [0, 1]$  tal que  $P(\emptyset) = 0$ ,  $P(\Omega) = 1$  i  $P$  és  $\sigma$ -additiva, és a dir  $P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n)$ , si els conjunts  $A_n \in \mathcal{A}$  són disjunts dos a dos.

## Observacions

- 1.- Si el conjunt  $\Omega$  és finit, que  $\mathcal{A}$  sigui una  $\sigma$ -àlgebra és equivalent a dir que sigui una àlgebra. Així mateix la  $\sigma$ -additivitat de  $P$  és equivalent a dir que  $P$  sigui additiva:  $P(A \cup B) = P(A) + P(B)$  si  $A \cap B = \emptyset$ . En el cas d'un conjunt  $\Omega$  infinit, les hipòtesis de  $\sigma$ -àlgebra i  $\sigma$ -additivitat s'imposen per raons d'ordre tècnic. Per exemple, si volem comprovar que la probabilitat de l'esdeveniment "que surti alguna vegada cara al tirar repetidament una moneda" és igual a 1, hem de calcular  $\frac{1}{2} + \frac{1}{2^2} + \frac{1}{2^3} + \dots = 1$ , i necessitem la hipòtesi de  $\sigma$ -additivitat.
- 2.- L'adequació d'un determinat espai de probabilitat a una certa experiència no ho resol l'axiomàtica de Kolmogorov i aquest és un problema empíric que surt del marc de la teoria matemàtica. Per abordar aquest problema cal utilitzar arguments de simetria (com en els jocs d'atzar) o bé tècniques estadístiques.

Dins l'axiomàtica de Kolmogorov, el fet que les freqüències relatives convergeixin cap a la probabilitat  $P(A)$  d'un esdeveniment aleatori  $A$  es pot demostrar rigurosament i és un teorema bàsic conegut com a "Llei dels grans nombres".

- 3.- En el llenguatge de la teoria de la mesura.  $(\Omega, \mathcal{A})$  és un *espai measurable* i  $P$  és una *mesura* tal que  $P(\Omega) = 1$ .

## 2. ESPAIS DE PROBABILITAT

La següent definició precisa el model matemàtic que hom associarà a una experiència aleatòria.

**Definició 2.1.** Un *espai de probabilitat* és una terna  $(\Omega, \mathcal{A}, P)$ , tal que:

- (i)  $\Omega$  és un conjunt format per totes les possibles realitzacions o resultats del fenòmen aleatori que estudiem,
- (ii)  $\mathcal{A}$  és una família de parts d' $\Omega$  que té estructura de  $\sigma$ -àlgebra, és a dir que compleix les següents propietats:

Cal observar que aquest concepte de  $\sigma$ -àlgebra generada per una família de parts es pot introduir també en d'altres tipus d'estructures i aquí, hom parlarà per exemple, de l'àlgebra o de la classe monòtona generada per una família de parts d' $\Omega$ .

Direm que  $(\Omega, \mathcal{A}, P)$  és un *espai de probabilitat finit* si  $\mathcal{A}$  és una àlgebra finita (i per tant, serà  $\sigma$ -àlgebra). Es fàcil veure que tota partició finita  $\Omega = A_1 \cup A_2 \cup \dots \cup A_r$  del conjunt  $\Omega$  en  $r$  conjunts no buits genera una àlgebra de  $2^r$  elements. Tot esdeveniment no buit  $A$  és de la forma  $A = A_{i_1} \cup \dots \cup A_{i_p}$ ,  $1 \leq p \leq r$ , i la seva probabilitat queda determinada per les probabilitats dels elements de la partició:

$$P(A) = P(A_{i_1}) + \dots + P(A_{i_p}).$$

En aquest cas, donar una probabilitat  $P$  sobre l'espai mesurable  $(\Omega, \mathcal{A})$  equival a escollir  $r$  nombres  $p_1, \dots, p_r \in [0, 1]$  tals que  $p_2 + \dots + p_r = 1$ .

El següent resultat ens diu que, recíprocament, tota àlgebra finita és d'aquest tipus.

**Proposició 2.2.** Sigui  $\mathcal{A}$  una àlgebra finita de parts d'un conjunt  $\Omega$ . Aleshores  $\mathcal{A}$  està generada per una partició finita i, en conseqüència, el nombre d'elements d' $\mathcal{A}$  és una potència de 2.

*DemOSTRACIÓ:* Un element  $A \neq \emptyset$  d' $\mathcal{A}$  es diu que és un *átom* si cap subconjunt propi d' $A$  pertany a  $\mathcal{A}$ . D'aquesta definició es dedueix que si  $A$  és un àtom i  $B \in \mathcal{A}$ , aleshores  $A \cap B = \emptyset$  o bé  $A \subset B$ . Finalment és fàcil veure que els àtoms formen una partició finita d' $\Omega$  que genera l'àlgebra  $\mathcal{A}$ . ■

## Propietats elementals de les probabilitats

- 1.- Si  $A \subset B$ , aleshores  $P(A) \leq P(B)$ .

En efecte,  $P(B) = P[A \cup (B - A)] = P(A) + P(B - A) \geq P(A)$ .

$$2.- P(A \cup B) + P(A \cap B) = P(A) + P(B).$$

En efecte, només cal posar  $A = (A \cap B) \cup (A \cap B^c)$  i  $B = (B \cap A) \cup (B \cap A^c)$ .

3.-  $P$  és subadditiva:

$$P\left(\bigcup_{i=1}^n A_i\right) \leq \sum_{i=1}^n P(A_i),$$

En efecte, només cal comprovar la propietat per  $n = 2$  (per  $n > 2$  es faria per inducció) i en aquest cas és una conseqüència immediata de la propietat anterior.

Cal fer notar que aquestes propietats són certes si  $\mathcal{A}$  és un àlgebra de parts d' $\Omega$ : i  $P: \mathcal{A} \rightarrow [0, +\infty]$  és una funció additiva tal que  $P(\emptyset) = 0$ .

**Definició 3.1.** La probabilitat d'un conjunt  $A \in \mathcal{A}$  condicionada per  $B$  es defineix per

$$P(A/B) = \frac{P(A \cap B)}{P(B)}. \quad (3.1)$$

Es comprova fàcilment que  $P(\cdot/B) : \mathcal{A} \rightarrow [0, 1]$  és una nova probabilitat sobre l'espai mesurable  $(\Omega, \mathcal{A})$ . Aquesta probabilitat s'introdueix quan sabem "a priori" que l'esdeveniment  $B$  s'ha realitzat. És a dir, en rebre aquesta informació addicional sobre l'experiència que estem analitzant, cal modificar la probabilitat ja que, per exemple, l'esdeveniment  $B$  passa a tenir probabilitat 1.

**Proposició 3.2.** (Príncipi de les probabilitats compostes). Sigui  $A_1, \dots, A_n$  elements de  $\mathcal{A}$  tals que  $P(A_1 \cap \dots \cap A_{n-1}) > 0$ . Aleshores

$$P(A_1 \cap A_2 \cap \dots \cap A_n) = P(A_1)P(A_2/A_1)P(A_3/A_1 \cap A_2) \dots P(A_n/A_1 \cap A_2 \cap \dots \cap A_{n-1}), \quad (3.2)$$

**DemOSTRACIÓ:** Es demostra per inducció. Per  $n = 2$  s'obté  $P(A_1 \cap A_2) = P(A_1)P(A_2/A_1)$  que és una conseqüència immediata de la definició de probabilitat condicionada. Suposant que la fórmula és certa per  $n - 1$ , escrivim

$$P(A_1 \cap A_2 \cap \dots \cap A_n) = P(A_1 \cap A_2 \cap \dots \cap A_{n-1})P(A_n/A_1 \cap A_2 \cap \dots \cap A_{n-1})$$

i apliquem la hipòtesi d'inducció. ■

Considerem una partició finita del conjunt  $\Omega$ ,  $\mathcal{P}_1 = \{A_1, A_2, \dots, A_n\} \subset \mathcal{A}$  on  $P(A_i) > 0$ ,  $1 \leq i \leq n$ . Donat un esdeveniment  $A \in \mathcal{A}$  la coneixença de les probabilitats  $P(A_i)$  i de les probabilitats condicionades  $P(A/A_i)$  permet de calcular la probabilitat d' $A$  de la manera següent:

$$P(A) = \sum_{i=1}^n P(A \cap A_i)P(A/A_i). \quad (3.3)$$

### Fórmules d'inversió de les condicions

Si  $A \cap B$  són dos esdeveniments de probabilitat no nula, tindrem

$$P(A \cap B) = \frac{P(A \cap B)}{P(B)} = P(B/A) \frac{P(A)}{P(B)}, \quad (3.4)$$

fórmula que permet d'invertir les condicions en el càlcul de les probabilitats condicionades.

Més generalment, considerem dues particions del conjunt  $\Omega$ ,  $\mathcal{P}_1 = \{A_1, A_2, \dots, A_n\}$  i  $\mathcal{P}_2 = \{B_1, B_2, \dots, B_m\}$  donades per conjunts d' $\mathcal{A}$  de probabilitat no nula. Coneguda la probabilitat dels conjunts de la primera partició  $P(A_i)$  i la probabilitat dels conjunts de la segona partició condicionada pels conjunts de la primera  $P(B_j/A_i)$ , ens plantearem el problema de calcular la probabilitat dels conjunts de la primera condicionada per la dels de la segona  $P(A_i/B_j)$ . Utilitzant la fórmula d'inversió (3.4) i la fórmula (3.3) s'obté la fórmula de Bayes que resol aquest problema:

$$P(A_i/B_j) = \frac{P(A_i)P(B_j/A_i)}{P(B_j)} = \frac{P(A_i)P(B_j/A_i)}{\sum_{k=1}^n P(B_j/A_k)P(A_k)}. \quad (3.5)$$

**Exemple d'aplicació:** Tenim una urna amb 3 boles blanques i 2 de negres. Treiem dues boles successivament. Volem calcular la probabilitat de que la primera bola hagi sigut blanca, sabent que la segona és negra.

Podem prendre com a conjunt  $\Omega$  el producte cartesià  $\{(b, n) \times (b, n)\}$  on  $b$  representa treure una bola blanca i  $n$  treure una bola negra. Considerem els següents esdeveniments:

$$B_1 = \text{"primera bola blanca"} = \{(b, n), (b, b)\},$$

$$B_2 = \text{"segona bola blanca"} = \{(b, b), (n, b)\},$$

$$N_1 = \text{"primera bola negra"} = \{(n, b), (n, n)\},$$

$$N_2 = \text{"segona bola negra"} = \{(b, n), (n, n)\},$$

Les particions són  $\mathcal{P}_1 = \{B_1, N_1\}$ ,  $\mathcal{P}_2 = \{B_2, N_2\}$  i sabem que

$$P(B_1) = \frac{3}{5}, \quad P(N_1) = \frac{2}{5},$$

$$P(N_2/B_1) = \frac{1}{2}, \quad P(N_2/N_1) = \frac{1}{4}, \quad P(B_2/B_1) = \frac{1}{2}, \quad P(B_2/N_1) = \frac{3}{4}.$$

Aleshores, per una aplicació de la fórmula de Bayes

$$P(B_1/N_2) = \frac{P(B_1)P(N_2/B_1)}{P(B_1)P(N_2/B_1) + P(N_1)P(N_2/N_1)} = \frac{3}{4}.$$

**Definició 3.3.** Es diu que dos esdeveniments  $A, B \in \mathcal{A}$  són independents si

$$P(A \cap B) = P(A)P(B).$$

Si  $P(A) > 0$ , el fet de que  $A \cap B$  siguin independents és equivalent a dir que  $P(B/A) = P(B)$ . És a dir, si sabem que  $A$  s'ha realitzat, la probabilitat de  $B$  no queda modificada. En altres paraules, la realització de l'esdeveniment  $A$  no influix de cap manera sobre l'esdeveniment  $B$ . Si  $P(B) > 0$  podem fer un raonament simètric i en conclusió, la definició d'Independència que hem donat equival a la noció intuitiva d'Independència entre la realització dels esdeveniments.

Propietats dels esdeveniments independents

- 1.- Els esdeveniments  $\Omega \setminus \emptyset$  són independents de qualsevol esdeveniment  $A$ .
- 2.- Si un esdeveniment  $A$  és independent d'ell mateix, aleshores  $P(A) = 0$  ó  $P(A) = 1$ .
- 3.- Les següents afirmacions són equivalents:

" $A$  és independent de  $B$ ",

" $A^c$  és independent de  $B^c$ ",

" $A^c$  és independent de  $B$ ",

" $A$  és independent de  $B$ ".

**Definició 3.4.** Un conjunt d'esdeveniments  $\{A_i, i \in I\} \subset \mathcal{A}$  es diu independent si per qualsevol subconjunt finit  $\{A_{i_1}, \dots, A_{i_n}\}$  es verifica que:

$$P(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_n}) = P(A_{i_1})P(A_{i_2}) \dots P(A_{i_n}).$$

Aquesta definició d'Independentesa és més forta que la independència entre cada parella d'esdeveniments del conjunt  $\{A_i, i \in I\}$ . Per exemple, en el cas de tres esdeveniments  $A_1, A_2, A_3$ , pot ocurrir que

$$P(A_1 \cap A_2) = P(A_1)P(A_2),$$

$$P(A_1 \cap A_3) = P(A_1)P(A_3),$$

$$P(A_2 \cap A_3) = P(A_2)P(A_3),$$

però en canvi,

$$P(A_1 \cap A_2 \cap A_3) \neq P(A_1)P(A_2)P(A_3).$$

Això voldiria dir que els esdeveniments  $A_1, A_2, A_3$  són independents dos a dos, però  $A_1 \cap A_2$  no és independent de  $A_3$ .

Per exemple, en el llançament de dos daus simultàniament, els esdeveniments  $A =$  "el resultat del primer dau és 1, 2 ó 3",  $B =$  "el resultat del segon dau és 4, 5, ó 6" i  $C =$  "la suma dels resultats obtinguts en ambdós daus és 7", són independents dos a dos però no conjuntament, com pot comprovar-se fàcilment.

#### 4. EXTENSIÓ DE MESURES

En aquest apartat exposarem alguns resultats sobre les propietats i la construcció de mesures que ens seran d'utilitat en l'estudi de les probabilitats. Introduirem primer algunes definicions.

**Definició 4.1.** Sigui  $\Omega$  un conjunt i  $\mathcal{C}$  una col·lecció de parts d' $\Omega$  tal que  $\emptyset \in \mathcal{C}$ . Considerem una funció  $\mu : \mathcal{C} \rightarrow [0, +\infty]$ .

(i) Direm que  $\mu$  és *additiva* si  $\mu(A \cup B) = \mu(A) + \mu(B)$ , sempre que  $A, B, A \cup B \in \mathcal{C}$  i  $A \cap B = \emptyset$ .

(ii) Direm que  $\mu$  és  *$\sigma$ -additiva* si  $\mu(\bigcup_{n=1}^{\infty} A_i) = \sum_{n=1}^{\infty} \mu(A_i)$ , sempre que  $A_n \in \mathcal{C}$  per a tot  $n \geq 1$ ,  $\bigcup_{n=1}^{\infty} A_n \in \mathcal{C}$  i  $A_n \cap A_m = \emptyset$  si  $n \neq m$ .

Es pot veure que una funció additiva que no sigui idènticament infinita compleix  $\mu(\emptyset) = 0$ .

Una terna  $(\Omega, \mathcal{A}, \mu)$  on  $\mathcal{A}$  és una  $\sigma$ -àlgebra de parts d' $\Omega$  i  $\mu$  és  $\sigma$ -additiva (amb  $\mu(\emptyset) = 0$ ) s'anomena un *espai de mesura*.

Si  $\mu(\Omega) < \infty$ , direm que l'espai de mesura és *finit*.  
Si  $\mu(\Omega) = 1$ , aleshores  $(\Omega, \mathcal{A}, \mu)$  és un *espai de probabilitat*.

##### 4.1. Caracterització de la $\sigma$ -additivitat

El següents resultats ens serviran per a comprovar si una funció de conjunt  $\mu$  és  $\sigma$ -additiva. Si  $\mu$  és una funció additiva en una àlgebra, aleshores la  $\sigma$ -additivitat equival a la continuïtat de  $\mu$  sobre successions creixents o decreixents (en aquest cas, cal afegir la condició  $\mu(\Omega) < \infty$ ).

**Proposició 4.2.** Suposem que  $\mu : \mathcal{A} \rightarrow [0, +\infty]$  és additiva,  $\mu(\emptyset) = 0$  i  $\mathcal{A}$  és una àlgebra de parts d' $\Omega$ . Aleshores

(i)  $\mu$  és  $\sigma$ -additiva si i únicament si per a tota successió creixent  $\{A_n, n \geq 1\} \subset \mathcal{A}$  tal que  $A = \bigcup_{n=1}^{\infty} A_n \in \mathcal{A}$  es compleix  $\mu(A) = \mu(A_n)$ .

(ii) Si  $\mu$  és  $\sigma$ -additiva, aleshores per a tota successió decreixent  $\{A_n, n \geq 1\} \subset \mathcal{A}$  tal que  $A = \bigcap_{n=1}^{\infty} A_n \in \mathcal{A}$  i  $\mu(A_n) < \infty$  es compleix  $\mu(A) = \mu(A_n)$ .

(iii) Si per a tota successió decreixent  $\{A_n, n \geq 1\} \subset \mathcal{A}$  tal que  $\cap_{n=1}^{\infty} A_n = \emptyset$  es compleix  $\mu(A_n) \downarrow 0$ , aleshores  $\mu$  és  $\sigma$ -additiva.

*Demonstració:*

(i) Suposem primer que  $\mu$  és  $\sigma$ -additiva i sigui  $\{A_n, n \geq 1\} \subset \mathcal{A}$  una successió creixent tal que  $A = \cup_{n=1}^{\infty} A_n \in \mathcal{A}$ . Aleshores

$$\begin{aligned}\mu(A) &= \mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \mu\left(\bigcup_{n=1}^{\infty} (A_n - A_{n-1})\right) = \sum_{n=1}^{\infty} \mu(A_n - A_{n-1}) = \\ &= \lim_{n \rightarrow \infty} \sum_{k=1}^n \mu(A_k - A_{k-1}) = \lim_{n \rightarrow \infty} \mu(A_n),\end{aligned}$$

on hem suposat  $A_0 = \emptyset$ .

Suposem ara que es compleix la propietat de continuitat seqüencial en  $\mathcal{A}$  i fixem una família  $\{A_n, n \geq 1\}$  de conjunts d' $\mathcal{A}$ , disjunts dos a dos i tal que  $\cup_{n=1}^{\infty} A_n = A$ . Aleshores tindrem

$$\mu(A) = \mu\left(\bigcup_{n=1}^{\infty} \left(\bigcup_{k=1}^n A_k\right)\right) = \lim_n \mu\left(\bigcup_{k=1}^n A_k\right) = \lim_n \sum_{k=1}^n \mu(A_k) = \sum_{k=1}^{\infty} \mu(A_k).$$

(ii) Suposem que  $\mu$  és  $\sigma$ -additiva i sigui  $\{A_n, n \geq 1\}$  una successió decreixent d'elements de  $\mathcal{A}$  tal que  $A = \cap_{n=1}^{\infty} A_n \in \mathcal{A}$ , amb  $\mu(A_1) < \infty$ . Aleshores

$$\begin{aligned}\mu(A_1) - \mu(A) &= \mu(A_1 - A) = \mu\left(\bigcup_{n=1}^{\infty} (A_1 - A_n)\right) = \lim_n \mu(A_1 - A_n) = \\ &= \mu(A_1) - \lim_n \mu(A_n).\end{aligned}$$

(iii) Fixem una família  $\{A_n, n \geq 1\}$  de conjunts d' $\mathcal{A}$ , disjunts dos a dos i tals que  $\cup_{n=1}^{\infty} A_n = A \in \mathcal{A}$ . Aleshores,  $\cup_{k=n}^{\infty} A_k \downarrow \emptyset$  quan  $n$  tendeix a infinit, i en conseqüència  $\mu(\cup_{k=n}^{\infty} A_k) \downarrow 0$ . Tindrem doncs

$$\mu(A) = \mu\left[\left(\bigcup_{k=1}^{n-1} A_k\right) \cup \left(\bigcup_{k=n}^{\infty} A_k\right)\right] = \sum_{k=1}^{n-1} \mu(A_k) + \mu\left(\bigcup_{k=n}^{\infty} A_k\right).$$

D'on en resulta

$$\mu(A) = \lim_n \sum_{k=1}^{n-1} \mu(A_k) = \sum_{k=1}^{\infty} \mu(A_k). \quad \blacksquare$$

### Observacions

- 1.- Hem vist que una funció additiva  $\mu$  sobre una àlgebra és subadditiva. Si  $\mu$  és  $\sigma$ -additiva, l'apartat (i) de la Proposició 4.2 implica que  $\mu$  és numerablement subadditiva, és a dir  $\mu(\cup_{n=1}^{\infty} A_n) \leq \sum_{n=1}^{\infty} \mu(A_n)$ , per a tota col·lecció de conjunts  $\{A_n, n \geq 1\} \subset \mathcal{A}$  tal que  $A = \cup_{n=1}^{\infty} A_n \in \mathcal{A}$ .

En efecte,

$$\mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \lim_n \mu\left(\bigcup_{k=1}^n A_k\right) \leq \lim_n \sum_{k=1}^n \mu(A_k) = \sum_{k=1}^{\infty} \mu(A_k).$$

- 2.- La condició  $\mu(A_1) < \infty$  és clarament necessària en la propietat (ii) de la Proposició 4.2. En efecte, considerem l'exemple  $\Omega = \mathbb{R}$  i  $\mu$  la mesura de Lebesgue. La successió de conjunts  $A_n = [n, +\infty)$  compleix  $\mu(A_n) = \infty$  per a tot  $n$ , és decreixent, i en canvi  $\mu(\cap_{n=1}^{\infty} A_n) = \mu(\emptyset) = 0$ .

## 4.2. Exemples de mesures

- 1.- Considerem un espai measurable  $(\Omega, \mathcal{A})$ . Fixat un element  $x \in \Omega$ , la mesura definida per

$$\delta_x(A) = 1_A(x), \quad A \in \mathcal{A},$$

- 1.- Considerem un espai measurable  $(\Omega, \mathcal{A})$ . Fixat un element  $x \in \Omega$ , la mesura definida s'anomena "delta de Dirac en el punt  $x$ " i representa una mesura de massa total igual a 1 concentrada en el punt  $x$ . Observi's que  $\delta_x$  és una probabilitat que correspon a la realització segura del resultat  $x$ .
- 2.- Si  $\mu_1$  i  $\mu_2$  són dues mesures i  $\alpha, \beta > 0$  aleshores  $\alpha\mu_1 + \beta\mu_2$  és una mesura. Si  $\{\mu_n, n \geq 1\}$  és una successió creixent de mesures,  $\sup_n \mu_n$  és una mesura.

Si  $\{\mu_n, n \geq 1\}$  és una successió de mesures, la sèrie  $\sum_{n=1}^{\infty} \mu_n$  és una mesura.

- 3.- Considerem una successió  $\{x_n, n \geq 1\} \subset \Omega$  i nombre reals positius  $\alpha_n$ . Aleshores  $\sum_{n=1}^{\infty} \alpha_n \delta_{x_n}$  és una mesura. Tota mesura d'aquest tipus s'anomena *discreta*. Si  $\Omega$  és un conjunt numerable o finit,  $\Omega = \{x_i, i \in I\} \subset \mathcal{A} = \mathcal{P}(\Omega)$ , aleshores totes les mesures són discretes. En efecte, una mesura qualsevol  $\mu$  es pot escriure com  $\mu = \sum_{i \in I} \alpha_i \delta_{x_i}$ , amb  $\alpha_i = \mu(\{x_i\})$ .

- 4.- La mesura  $\mu(A) = \text{card}(A)$  si el conjunt  $A$  és finit i  $\mu(A) = \infty$  si  $A$  és infinit s'anomena *mesura comptadora*.

5.- Considerem un espai de mesura  $(\Omega, \mathcal{A}, \mu)$  i un conjunt  $A \in \mathcal{A}$ . La funció  $\mu_A(B) = \mu(A \cap B)$ ,  $B \in \mathcal{A}$  és una mesura que s'anomena la restricció de  $\mu$  al conjunt  $A$ .

Si  $\Omega = \bigcup_{n=1}^{\infty} A_n$ , on els conjunts  $A_n \in \mathcal{A}$  són disjunts dos a dos, aleshores  $\mu = \sum_{n=1}^{\infty} \mu_{A_n}$ :

**4.3. Classes monòtones i unicitat de l'extensió**

En aquest apartat ens proposarem analitzar l'extensió d'una funció  $\sigma$ -additiva definida en una àlgebra a la  $\sigma$ -àlgebra generada per l'àlgebra. Primer estudiarem la unicitat i després l'existeï�性 d'aquesta extensió. Aquests resultats seran utilitzats per a determinar la llei d'una variable aleatòria a partir de la seva funció de distribució. Per tractar la unicitat necessitem introduir els conceptes de classe monòtona i mesures  $\sigma$ -finites.

**Definició 4.3.** Una col·lecció no buida  $\mathcal{C}$  de subconjunts d' $\Omega$  és una classe monòtona si per a tota successió monòtona (creixent o decreixent)  $\{A_n, n \geq 1\} \subset \mathcal{C}$  es compleix

$$\lim_n A_n \in \mathcal{C}.$$

Obviament, tota  $\sigma$ -àlgebra és una classe monòtona. Recíprocament, una àlgebra  $\mathcal{A}$  que sigui classe monòtona és  $\sigma$ -àlgebra. En efecte si  $A_n \in \mathcal{A}$ , per  $n \geq 1$ , aleshores  $\bigcup_{n=1}^{\infty} A_n = \lim_n \left( \bigcup_{k=1}^n A_k \right) \in \mathcal{A}$ .

**Teorema 4.4.** (Teorema de les classes monòtones). Si  $\mathcal{A}$  és una àlgebra de parts d' $\Omega$ ;  $\mathcal{C}$  és una classe monòtona tal que  $\mathcal{C} \supseteq \mathcal{A}$ , aleshores  $\mathcal{C} \supseteq \sigma(\mathcal{A})$ .

**Demostració:** Designarem per  $\mathfrak{m}(\mathcal{A})$  la classe monòtona engendrada per  $\mathcal{A}$ . Llavors  $\mathfrak{m}(\mathcal{A}) \subset \mathcal{C}$ . Escriurem  $\mathfrak{m} = \mathfrak{m}(\mathcal{A})$ . Només cal venire que  $\mathfrak{m} \supseteq \sigma(\mathcal{A})$  (en realitat tindrem  $\mathfrak{m} = \sigma(\mathcal{A})$ ).

Fixat un conjunt  $A \in \mathfrak{m}$ , definim

$$\mathfrak{m}_{\mathcal{A}} = \{B \in \mathfrak{m} : A \cap B, A \cap B^c, A^c \cap B \in \mathfrak{m}\}.$$

Es comprova fàcilment que  $\mathfrak{m}_{\mathcal{A}}$  és una classe monòtona. Si  $A \in \mathcal{A}$ , aleshores  $A \subset \mathfrak{m}_{\mathcal{A}}$  ja que  $\mathcal{A}$  és una àlgebra i, en conseqüència,  $\mathfrak{m} \subset \mathfrak{m}_{\mathcal{A}}$ , és a dir,  $\mathfrak{m} = \mathfrak{m}_{\mathcal{A}}$ .

Suposem ara que el conjunt  $A$  pertany a  $\mathfrak{m}$ . Si  $B \in \mathcal{A}$ , com que  $A \in \mathfrak{m}_B = \mathfrak{m}$  tindrem que  $B \cap A, B \cap A^c, B^c \cap A \in \mathfrak{m}$ . Això implica que  $B \in \mathfrak{m}_{\mathcal{A}}$ . Per tant,  $\mathcal{A} \subset \mathfrak{m}_{\mathcal{A}}$ , i, com abans, això ens diu que  $\mathfrak{m} \subset \mathfrak{m}_{\mathcal{A}}$  i  $\mathfrak{m} = \mathfrak{m}_{\mathcal{A}}$ .

Finalment, si  $A \cap B$  són dos conjunts de  $\mathfrak{m}$ , les relacions  $A \in \mathfrak{m}_B$  i  $B \in \mathfrak{m}_A$  ens diuen que  $A \cap B, A \cap B^c, A^c \cap B \in \mathfrak{m}$ . Per tant,  $\mathfrak{m}$  és una àlgebra i com que és classe monòtona, és una  $\sigma$ -àlgebra. En conclusió  $\mathfrak{m} \supseteq \sigma(\mathcal{A})$ . ■

**Definició 4.5.** Sigui  $\Omega$  un conjunt,  $\mathcal{A}$  un subconjunt de parts de  $\Omega$  i  $\mu$  una funció  $\sigma$ -additiva sobre  $\mathcal{A}$ . Direm que  $\mu$  és  $\sigma$ -finita si existeix una descomposició  $\Omega = \bigcup_{n=1}^{\infty} A_n$  del conjunt  $\Omega$  en conjunts  $A_n \in \mathcal{A}$  tals que  $\mu(A_n) < \infty$  per a tot  $n$ . La successió  $\{A_n, n \geq 1\}$  es pot agafar creixent (només cal prendre  $A'_n = A_1 \cup A_2 \cup \dots \cup A_n$ ) o formada per conjunts disjunts dos a dos (prenen en compte de  $A_n$ ,  $A''_n = A'_n - (A_1 \cup A_2 \cup \dots \cup A_{n-1})$ ,  $n > 1$ ). La mesura de Lebesgue en  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$  és un exemple de mesura  $\sigma$ -finita. La mesura compactadora en un conjunt  $\Omega$  no numerable és un exemple de mesura que no és  $\sigma$ -finita.

**Teorema 4.6.** (Unicitat de l'extensió). Sigui  $\mathcal{A}$  una àlgebra de parts d'un conjunt  $\Omega$ . Si  $\mu_1, \mu_2$ , són dues mesures sobre la  $\sigma$ -àlgebra  $\sigma(\mathcal{A})$  tals que coincideixen sobre  $\mathcal{A}$  i són  $\sigma$ -finites sobre  $\mathcal{A}$ , aleshores  $\mu_1 = \mu_2$ .

**Demostració:** Suposem primer que  $\mu_1 \neq \mu_2$ , són finites. Definim  $\mathcal{C} = \{A \in \mathcal{A} : \mu_1(A) = \mu_2(A)\}$ .  $\mathcal{C}$  és una classe monòtona, per tant  $\mathcal{C} \supseteq \sigma(\mathcal{A})$ , i el teorema queda demostrat. En el cas  $\sigma$ -finit tindrem  $\Omega = \bigcup_{n=1}^{\infty} A_n$ ,  $A_n \in \mathcal{A}$ ,  $\mu_1(A_n) < \infty$  i  $\mu_2(A_n) < \infty$  per a tot  $n$  i els  $A_n$  es poden prendre disjunts dos a dos. Aleshores per cada  $n$ , les mesures  $\mu_{1,A_n}(B) = \mu_1(B \cap A_n)$ ,  $\mu_{2,A_n}(B) = \mu_2(B \cap A_n)$ , coincideixen sobre  $A$  i són finites. Per tant,  $\mu_{1,A_n} = \mu_{2,A_n}$ , i en conseqüència  $\mu_1 = \sum_{n=1}^{\infty} \mu_{1,A_n} = \sum_{n=1}^{\infty} \mu_{2,A_n} = \mu_2$ . ■

#### 4.4. Mesures extiors. Teoremes de Carathéodory

Per tal de demostrar els teoremes d'estensió de mesures cal introduir el concepte de mesura extior.

**Definició 4.7.** Sigui  $\Omega$  un conjunt. Una aplicació  $\mu^* : \mathcal{P}(\Omega) \rightarrow [0, +\infty]$  s'anomena una mesura extior si compleix les següents propietats:

- (i)  $\mu^*(\emptyset) = 0$ .
- (ii)  $\mu^*$  és creixent.
- (iii)  $\mu^*$  és subadditiva, és a dir, per a qualsevol successió de subconjunts de  $\Omega$ ,  $\{A_n, n \geq 1\}$ , es compleix  $\mu^*\left(\bigcup_{n=1}^{\infty} A_n\right) \leq \sum_{n=1}^{\infty} \mu^*(A_n)$ .

Per exemple, si  $\varphi : \Omega \rightarrow [0, +\infty]$  és una funció arbitrària, aleshores  $\mu^*(A) = \sup_{x \in A} \varphi(x)$  és una mesura exterior.

Donada una mesura exterior  $\mu^*$ , direm que un conjunt  $A \subset \Omega$  és  $\mu^*$ -mesurable si  $\mu^*(B) \geq \mu^*(A \cap B) + \mu^*(A^c \cap B)$ , per a tot  $B \subset \Omega$ .

En realitat es compleix la igualtat ja que la desigualtat contrària sempre és certa, degut a la subadditivitat.

**Teorema 4.8.** (Primer teorema de Carathéodory).

Signi  $\mathcal{A}_{\mu^*}$  la família dels conjunts  $\mu^*$ -mesurables. Aleshores,  $\mathcal{A}_{\mu^*}$  és una  $\sigma$ -àlgebra i  $\mu^*$  és  $\sigma$ -additiva sobre  $\mathcal{A}_{\mu^*}$ .

*Demostració:*

1) Veurem primer que  $\mathcal{A}_{\mu^*}$  és una àlgebra. El conjunt  $\emptyset$  és de  $\mathcal{A}_{\mu^*}$ , ja que  $\mu^*(B) = \mu^*(\emptyset \cap B) + \mu^*(\Omega \cap B)$ , per a tot  $B \subset \Omega$ .

Clarament  $\mathcal{A}_{\mu^*}$  és estable per complementació.

Provem finalment l'estabilitat per reunions finites. Signin  $A_1, A_2 \in \mathcal{A}_{\mu^*}$ . Fixem  $B \subset \Omega$ .

Tindrem

$$\begin{aligned}\mu^*(B) &= \mu^*(B \cap A_1) + \mu^*(B \cap A_1^c) \\ &= \mu^*(B \cap A_1 \cap A_2) + \mu^*(B \cap A_1 \cap A_2^c) \\ &\quad + \mu^*(B \cap A_1^c \cap A_2) + \mu^*(B \cap A_1^c \cap A_2^c).\end{aligned}$$

Si apliquem aquesta fórmula general al conjunt  $B \cap (A_1 \cup A_2)$  obtenim

$$\mu^*(B \cap (A_1 \cap A_2)) = \mu^*(B \cap A_1 \cap A_2) + \mu^*(B \cap A_1 \cap A_2^c) + \mu^*(B \cap A_1^c \cap A_2),$$

ja que  $A_1 \cap A_2^c \subset A_1 \cup A_2$  i  $A_1^c \cap A_2 \subset A_1 \cup A_2$ .

En conclusió

$$\mu^*(B) = \mu^*(B \cap (A_1 \cup A_2)) + \mu^*(B \cap (A_1 \cup A_2)^c),$$

per a tot  $B \subset \Omega$ . La qual cosa implica que  $A_1 \cup A_2 \in \mathcal{A}_{\mu^*}$ .

2) Demostrem finalment que  $\mathcal{A}_{\mu^*}$  és una  $\sigma$ -àlgebra i que  $\mu^*$  és  $\sigma$ -additiva sobre  $\mathcal{A}_{\mu^*}$ .

Considerem una família de conjunts de  $\mathcal{A}_{\mu^*}$ ,  $\{A_n, n \geq 1\}$  que, sense pèrdua de generalitat podem suposar disjunts dos a dos. Possem  $B_n = \bigcup_{k=1}^n A_k$  per  $n \geq 1$ ,  $B_0 = \emptyset$  i  $A = \bigcup_{n=1}^{\infty} A_n$ . Sabem que  $B_n \in \mathcal{A}_{\mu^*}$  per a tot  $n$ . Signi  $C \subset \Omega$  i  $p \geq 1$ . Es compleix

$$\begin{aligned}\mu^*(C \cap B_p) &= \mu^*(C \cap B_p \cap B_{p-1}) + \mu^*(C \cap B_p \cap B_{p-1}^c) = \\ &= \mu^*(C \cap B_{p-1}) + \mu^*(C \cap A_p).\end{aligned}$$

Sumant per  $p = 1$  fins a  $n$  obremim

$$\mu^*(C \cap B_n) = \sum_{p=1}^n \mu^*(C \cap A_p).$$

En conseqüència

$$\begin{aligned}\mu^*(C) &= \mu^*(C \cap B_n) + \mu^*(C \cap B_n^c) = \sum_{p=1}^n \mu^*(C \cap A_p) + \mu^*(C \cap A^c), \\ &\geq \sum_{p=1}^n \mu^*(C \cap A_p) + \mu^*(C \cap A^c),\end{aligned}$$

ja que  $B_n \subset A$ .

Fent  $n \rightarrow \infty$  i degut a la subadditivitat de  $\mu^*$ , tindrem

$$\mu^*(C) \geq \sum_{p=1}^{\infty} \mu^*(C \cap A_p) + \mu^*(C \cap A^c) \geq \mu^*(C \cap A) + \mu^*(C \cap A^c),$$

Ia qual cosa implica  $A \in \mathcal{A}_{\mu^*}$  i, per tant,  $\mathcal{A}_{\mu^*}$  és  $\sigma$ -àlgebra.

Finalment, posant  $C = A$  en l'expressió anterior arribem a

$$\mu^*(A) \geq \sum_{p=1}^{\infty} \mu^*(A_p). \quad \blacksquare$$

**Teorema 4.9.** (Segon teorema de Carathéodory). Signi  $\mathcal{A}$  una àlgebra de parts d'un conjunt  $\Omega$  i  $\mu$  una funció  $\sigma$ -additiva sobre  $\mathcal{A}$ . Per a tot  $B \subset \Omega$  definim

$$\mu^*(B) = \inf \sum_{n=1}^{\infty} \mu(A_n),$$

on l'infinim es pren sobre tots els recobriments numerables  $\{A_n, n \geq 1\}$  del conjunt  $B$  per conjunts  $A_n \in \mathcal{A}$ . Aleshores es compleix que  $\mu^*$  és una mesura exterior.  $A \subset \mathcal{A}_{\mu^*}$  i  $\mu^*|_A = \mu$ .

*Demostració:*

1) Veurem primer que  $\mu^*|_A = \mu$ . En efecte, fixem  $A \in \mathcal{A}$  i un recobriment  $\{A_n, n \geq 1\} \subset \mathcal{A}$  de  $A$ . Tindrem que  $\{A_n \cap A, n \geq 1\}$  és una família de conjunts tal que  $A = \bigcup_{n=1}^{\infty} (A_n \cap A)$ . Com que  $\mu$  és subadditiva podem escriure,

$$\mu(A) = \mu\left(\bigcup_{n=1}^{\infty} (A_n \cap A)\right) \leq \sum_{n=1}^{\infty} \mu(A_n \cap A) \leq \sum_{n=1}^{\infty} \mu(A_n).$$

la qual cosa implica que  $\mu(A) \leq \mu^*(A)$ . La desigualtat contrària és immediata.

2)  $\mu^*$  és una mesura exterior. En efecte,

$$(i) \quad \mu^*(\emptyset) = 0.$$

- (ii) Si  $A \subset B$ , tot recobriment numerable de  $B$  és també un recobriment d' $A$ . don en resulta  $\mu^*(A) \leq \mu^*(B)$ .

(iii) Si  $B = \bigcup_{n=1}^{\infty} B_n$ , tenim que per a tot  $\varepsilon > 0$  i per a tot  $n \geq 1$  existeix un recobriment  $\{B_n^k, k \geq 1\} \subset A$  de  $B_n$  tal que  $\sum_{k=1}^{\infty} \mu(B_n^k) \leq \mu^*(B_n) + \frac{\varepsilon}{2^n}$ .

La col·lecció de conjunts  $\{B_n^k, k, n \geq 1\}$  és un recobriment numerable de  $B$  i per a tot  $\varepsilon > 0$ ,  $\mu^*(B) \leq \sum_{n,k} \mu^*(B_n^k) \leq \sum_{n=1}^{\infty} \mu^*(B_n) + \varepsilon$ . En conseqüència,

$$\mu^*(B) \leq \sum_{n=1}^{\infty} \mu^*(B_n).$$

3)  $\mathcal{A} \subset \mathcal{A}_{\mu^*}$ . En efecte, fixem  $A \in \mathcal{A}$ ;  $B \subset \Omega$ . Per a tot  $\varepsilon > 0$  existeix un recobriment de  $B$ ,  $\{A_n, n \geq 1\} \subset \mathcal{A}$ , tal que  $\sum_{n=1}^{\infty} \mu(A_n) \leq \mu^*(B) + \varepsilon$ . Ara bé

$$\sum_{n=1}^{\infty} \mu(A_n) = \sum_{n=1}^{\infty} [\mu(A \cap A_n) + \mu(A^c \cap A_n)] \geq \mu^*(A \cap B) + \mu^*(A^c \cap B).$$

Per tant,  $\mu^*(B) \geq \mu^*(A \cap B) + \mu^*(A^c \cap B)$ , la qual cosa implica  $A \in \mathcal{A}_{\mu^*}$ . ■

El teorema anterior ens diu que existeix una extensió  $\sigma$ -additiva de  $\mu$  a la  $\sigma$ -àlgebra  $\sigma(\mathcal{A}) \subset \mathcal{A}_{\mu^*}$ , que serà única si  $\mu$  és  $\sigma$ -finita sobre  $\mathcal{A}$ , tal i com hem provat en el Teorema 4.6.

Un espai de mesura qualsevol  $(\Omega, \mathcal{A}, \mu)$  es pot completar de la forma següent:

Definim

$$(i) \quad \bar{\mathcal{A}} = \{A \cup N, A \in \mathcal{A}, N \text{ negligible}\}$$

$$(ii) \quad \bar{\mu}: \bar{\mathcal{A}} \rightarrow [0, +\infty], \quad \bar{\mu}(A \cup N) = \mu(A).$$

$$(iii) \quad \bar{\mu}(A) = [0, +\infty], \quad \bar{\mu}(A \cup N) = \mu(A).$$

i

Aleshores:

(i)  $\bar{\mathcal{A}}$  és una  $\sigma$ -àlgebra que conté  $\mathcal{A}$ . En efecte,  $\bar{\mathcal{A}}$  és estable per unions numerables. ja que

$$\bigcup_{n=1}^{\infty} (\mathcal{A}_n \cup N_n) = \left( \bigcup_{n=1}^{\infty} \mathcal{A}_n \right) \cup \left( \bigcup_{n=1}^{\infty} N_n \right) \in \bar{\mathcal{A}}$$

degot a que  $\bigcup_{n=1}^{\infty} A_n \in \mathcal{A}$  i  $\bigcup_{n=1}^{\infty} N_n$  és negligible, com es comprueba fàcilment. D'altra banda,  $\bar{\mathcal{A}}$  és estable per pas al complementari, ja que  $(A \cup N)^c = A^c \cap B^c = (\mathcal{A}^c \cap B^c) \cup [\mathcal{A}^c \cap (N^c - B^c)]$ , on  $N \subset B$ ,  $B \in \mathcal{A}$  i  $\mu(B) = 0$ . Aleshores  $(A \cup N)^c \in \mathcal{A}$  degut a que  $\mathcal{A}^c \cap B^c \in \mathcal{A}$ , i  $\mathcal{A}^c \cap (N^c - B^c)$  és negligible donat que està contingut en  $B$ .

(ii)  $\bar{\mu}$  està ben definida. En efecte, si  $A_1 \cup N_1 = A_2 \cup N_2$ , tindrem

$$\bar{\mu}(A_1) = \mu(A_1 \cap A_2) + \mu(A_1 - A_2) = \mu(A_1 \cap A_2),$$

ja que  $A_1 - A_2 \subset N_2$ , i per simetria,  $\bar{\mu}(A_2) = \mu(A_1 \cap A_2)$ .

(iii)  $\bar{\mu}$  és  $\sigma$ -additiva i  $\bar{\mu}|_{\mathcal{A}} = \mu$ .

(iv) L'espai de mesura  $(\Omega, \bar{\mathcal{A}}, \bar{\mu})$  és complet. En efecte, si  $M$  és negligible en  $(\Omega, \bar{\mathcal{A}}, \bar{\mu})$  tindrem  $M \subset A \cup N$ ,  $A \in \mathcal{A}$ ,  $\mu(A) = 0$  i  $N$  negligible en  $(\Omega, \mathcal{A}, \mu)$  és a dir  $N \subset B$ ,  $B \in \mathcal{A}$ ,  $\mu(B) = 0$ . Aleshores  $M \subset A \cup B$ ,  $A \cup B \in \mathcal{A}$ ,  $\mu(A \cup B) = 0$ , és a dir,  $M$  és negligible en  $(\Omega, \mathcal{A}, \mu)$ , i per tant  $M \in \bar{\mathcal{A}}$ .

L'espai  $(\Omega, \bar{\mathcal{A}}, \bar{\mu})$  s'anomena la *compleció* de  $(\Omega, \mathcal{A}, \mu)$ .

**Teorema 4.11.** Sigui  $\mu$  una funció  $\sigma$ -additiva i  $\sigma$ -finita en una àlgebra  $\mathcal{A}$  de parts d'un conjunt  $\Omega$ . Aleshores, l'espai de mesura  $(\Omega, \mathcal{A}_{\mu^*}, \mu^*)$  és la compleció de  $(\Omega, \sigma(\mathcal{A}), \mu^*)$ , on hem utilitzat les notacions introduïdes en els teoremes de Carathéodory.

*DemOSTRació:*

En primer lloc, veurem que  $(\Omega, \mathcal{A}_{\mu^*}, \mu^*)$  és complet. En efecte, si  $N$  és negligible, existeix  $A \in \mathcal{A}_{\mu^*}$  tal que  $N \subset A$ ,  $\mu^*(A) = 0$ . Això implica que per tot  $B \subset \Omega$ ,  $\mu^*(B) \geq \mu^*(B \cap N) + \mu^*(B \cap N^c)$  ja que  $\mu^*(B \cap N) = 0$ . Per tant,  $N \in \mathcal{A}_{\mu^*}$ .

**Definició 4.10.** Considerarem un espai de mesura  $(\Omega, \mathcal{A}, \mu)$ . Direm que un conjunt  $N \subset \Omega$  (que pot no pertànyer a  $\mathcal{A}$ ) és *negligible* si existeix un  $A \in \mathcal{A}$  tal que  $N \subset A$  i  $\mu(A) = 0$ . Direm que l'espai de mesura és *complet* si la  $\sigma$ -àlgebra  $\mathcal{A}$  conté tots els conjunts negligibles.

Es clar que  $\overline{\sigma(\mathcal{A})} \subseteq \mathcal{A}_\mu$ , ja que  $(\Omega, \mathcal{A}_\mu, \mu^*)$  és complet.

Per simplificar la notació, designarem per  $\mu$  la mesura  $\mu^*$  sobre  $\mathcal{A}_\mu$ . Fixem un  $A \in \mathcal{A}_\mu$ . Volem veure que  $A \in \overline{\sigma(\mathcal{A})}$ .

Per cada  $n \geq 1$  existeix un recobriment  $\{A_n^k, k \geq 1\}$  de  $A$ , format per conjunts de  $\mathcal{A}$  tal que

$$\frac{1}{n} + \mu^*(A) \geq \sum_{k=1}^{\infty} \mu(A_n^k) \geq \mu(\bigcup_{k=1}^{\infty} A_n^k) = \mu(G_n) \geq \mu^*(A),$$

on  $G_n = \bigcup_{k=1}^{\infty} A_n^k \in \sigma(\mathcal{A})$ . Definim  $G = \bigcap_{n=1}^{\infty} G_n \in \sigma(\mathcal{A})$ . Es compleix  $A \subset G$  i  $\frac{1}{n} + \mu^*(A) \geq \mu(G_n) \geq \mu^*(A)$ , per a tot  $n \geq 1$ . En conseqüència,  $\mu(G) = \mu(A)$ .

Acabem de veure que per tot conjunt  $B \in \mathcal{A}_\mu$  existeix un  $H \in \sigma(\mathcal{A})$  tal que  $B \subset H$  i  $\mu(B) = \mu(H)$ . Si apliquem aquest resultat a  $G - A$  tindrem  $G - A \subset F$ ,  $F \in \sigma(\mathcal{A})$ ,  $\mu(F) = 0$ . Finalment,  $A = (G - F) \cup (A \cap F)$  on  $G - F \in \sigma(\mathcal{A})$  i  $\mu(A \cap F) = 0$ , és a dir,  $A \in \overline{\sigma(\mathcal{A})}$ . ■

#### 4.6. Semialgebres

Algunes vegades la funció de conjunt  $\mu$  que volem estendre no estarà definida inicialment en una àlgebra sinó en una família de conjunts que no és estable per unions. Aquest és el cas, per exemple, de la família de tots els intervals de la recta.

Per tal de tractar aquest exemple i també el cas de l'espai producte, introduirem la noció de semialgebra i veurem com es pot estendre una funció  $\mu$  definida inicialment en una semialgebra a una funció definida en una àlgebra.

**Definició 4.12.** Direm que una col·lecció  $\mathcal{S}$  de parts d'un conjunt  $\Omega$  és una semialgebra si

- (i)  $\Omega \in \mathcal{S}$ ,
- (ii) Si  $A, B$  són conjunts de  $\mathcal{S}$ , aleshores  $A \cap B$  és també de  $\mathcal{S}$ .
- (iii) Si  $A$  és de  $\mathcal{S}$ , el conjunt  $A^n$  és una unió finita de conjunts de  $\mathcal{S}$  disjunts dos a dos.

**Proposició 4.13.** L'àlgebra generada per una semialgebra  $\mathcal{S}$  coincideix amb la família de les unions finites de conjunts de  $\mathcal{S}$  disjunts dos a dos.

**Demostració:** Sigui  $\mathcal{G}(\mathcal{S})$  l'àlgebra generada per  $\mathcal{S}$  i  $\mathcal{G}$  la família de les unions finites de

conjunts de  $\mathcal{S}$  disjunts dos a dos. És clar que  $\mathcal{S} \subset \mathcal{G} \subset a(\mathcal{S})$ . Es doncs suficient demostrar que  $\mathcal{G}$  és una àlgebra, per tal d'obtenir  $\mathcal{G} = a(\mathcal{S})$ :

$\mathcal{G}$  és estable per interseccions finites degut a la propietat (ii). En efecte, sigüen  $A_1, A_2 \in \mathcal{G}$  i suposem que  $A_1 = B_1^1 \cup \dots \cup B_1^n$ ,  $A_2 = B_2^1 \cup \dots \cup B_2^m$ . Aleshores

$$A_1 \cap A_2 = [B_1^1 \cup \dots \cup B_1^n] \cap [B_2^1 \cup \dots \cup B_2^m] = \bigcup_{i=1}^n \bigcup_{j=1}^m (B_1^i \cap B_2^j) \in \mathcal{G},$$

on els  $B_1^i$ ,  $i = 1, \dots, n$ ,  $B_2^j$ ,  $j = 1, \dots, m$ , són de  $\mathcal{S}$  i disjunts dos a dos.

$\mathcal{G}$  és estable per pas al complementari, ja que si  $B_1, \dots, B_n$  són conjunts de  $\mathcal{S}$  disjunts dos a dos, tindrem  $(B_1 \cup \dots \cup B_n)^c = B_1^c \cap \dots \cap B_n^c \in \mathcal{G}$ , degut a que  $\mathcal{G}$  és estable per interseccions finites i els  $B_i^c$  són de  $\mathcal{G}$  per la propietat (iii). ■

**Teorema 4.14.** Sigui  $\mu$  una funció additiva definida en una semialgebra  $\mathcal{S}$ . Existeix una única extensió additiva de  $\mu$  a l'àlgebra  $\mathcal{A} = a(\mathcal{S})$ . Endemés si  $\mu$  és  $\sigma$ -additiva en  $\mathcal{S}$ , també ho serà en  $a(\mathcal{S})$ .

**Demostració:** Com que  $\mathcal{A}$  és la família de les unions finites de conjunts de  $\mathcal{S}$  disjunts dos a dos, tot conjunt  $A \in \mathcal{A}$  és de la forma  $A = \bigcup_{i=1}^n A_i$ , on  $A_i \in \mathcal{S}$  i  $A_i \cap A_j = \emptyset$  si  $i \neq j$ . Posem per definició  $\mu(A) = \sum_{i=1}^n \mu(A_i)$ .

Aquesta definició és consistent ja que donades dues particions diferents del conjunt  $A = \bigcup_{i=1}^n A_i = \bigcup_{j=1}^m B_j$ , on  $B_j \in \mathcal{S}$  i  $B_i \cap B_k = \emptyset$  per  $j \neq k$ , tindrem

$$\mu(A) = \sum_{i=1}^n \mu(A_i) = \sum_{i=1}^n \sum_{j=1}^m \mu(A_i \cap B_j) = \sum_{j=1}^m \mu(B_j).$$

Es facilt veure que  $\mu$  és additiva en  $\mathcal{A}$  i que és la única extensió additiva.

Finalment, suposem que  $\mu$  és  $\sigma$ -additiva en  $\mathcal{S}$  i considerem una família  $\{A_n, n \geq 1\}$  de conjunts de  $\mathcal{A}$  disjunts dos a dos. Posem  $A = \bigcup_{n=1}^{\infty} A_n$ , i suposem també que  $A \in \mathcal{A}$ . Aleshores, si  $A = \bigcup_{j=1}^m C_j$ ,  $C_j \in \mathcal{S}$ ,  $C_j \cap C_k = \emptyset$  per  $j \neq k$ , i  $A_n = \bigcup_{j=1}^{m_n} B_j^n$ ,  $B_j^n \in \mathcal{S}$ ,  $B_j^n \cap B_k^n = \emptyset$  per  $j \neq k$ , tindrem

$$\mu(A) = \sum_{j=1}^m \mu(C_j) = \sum_{j=1}^m \sum_{n=1}^{\infty} \sum_{k=1}^{m_n} \mu(C_j \cap B_k^n) = \sum_{n=1}^{\infty} \mu(A_n).$$

Observem que aquest teorema, juntement amb el segon teorema de Carathéodory, ens diu que existeix una extensió  $\sigma$ -additiva de  $\mu$  a la  $\sigma$ -àlgebra generada per  $\mathcal{S}$ , que serà única si  $\mu$  és  $\sigma$ -finita.

## 5. PROBABILITATS EN LA RECTA REAL I FUNCIONS DE DISTRIBUCIÓ

En moltes experiències aleatòries el conjunt de resultats és un conjunt de nombres reals. per exemple, un conjunt finit o numerable, o bé un interval. Això ens porta a estudiar les probabilitats sobre la recta real. Veurem que hi ha una forma molt senzilla de donar una probabilitat sobre  $\mathbb{R}$ , mitjançant les funcions de distribució.

Signi  $\mathcal{S}$  la semialgebra de parts de  $\mathbb{R}$  formada pels intervals  $(a, b]$ ,  $a < b$ ; les semirectes  $(a, +\infty)$ ,  $(-\infty, a]$ ,  $a \in \mathbb{R}$ ;  $\emptyset$  i  $\mathbb{R}$ . La  $\sigma$ -àlgebra generada per  $\mathcal{S}$  s'anomena la  $\sigma$ -àlgebra de Borel  $\mathcal{B}(\mathbb{R})$  i coincideix amb la  $\sigma$ -àlgebra generada pels conjunts oberts ( $\circ$  tancats) de  $\mathbb{R}$ .

Si  $\mu$  és una probabilitat en  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$  es defineix la seva funció de distribució per  $F : \mathbb{R} \rightarrow [0, 1]$ ,  $F(x) = \mu((-\infty, x])$ . Aquesta funció té les següents propietats:

- (i)  $F$  és creixent. En efecte, si  $x \leq y$ , és té  $(-\infty, x] \subset (-\infty, y]$ . Per tant hom dedueix  $F(x) = \mu((-\infty, x]) \leq \mu((-\infty, y]) = F(y)$ .
- (ii)  $F$  és contínua per la dreta. En efecte, donada una successió  $y_n \downarrow x$  es compleix que  $(-\infty, y_n] \downarrow (-\infty, x]$ , i per les propietats de continuitat de les mesures (Proposició 4.2) obtenim  $F(y_n) = \mu((-\infty, y_n]) \downarrow \mu((-\infty, x]) = F(x)$ .

- (iii) Es compleix  $\lim_{x \rightarrow -\infty} F(x) = 0$ . En efecte, per a tota successió  $x_n \uparrow \infty$  tenim  $(-\infty, x_n] \uparrow \mathbb{R}$ , i per tant  $F(x_n) = \mu((-\infty, x_n]) \uparrow \mu(\mathbb{R}) = 1$ .
- (iv)  $\lim_{x \rightarrow -\infty} F(x) = 1$ . En efecte, per a tota successió  $x_n \downarrow -\infty$  tenim  $(-\infty, x_n] \downarrow \emptyset$ , i per tant  $F(x_n) = \mu((-\infty, x_n]) \downarrow \mu(\emptyset) = 0$ .

En general, una funció  $F : \mathbb{R} \rightarrow [0, 1]$  que compleix les propietats (i) a (iv) s'anomena una funció de distribució.

Si  $F$  és la funció de distribució d'una probabilitat  $\mu$ , es compleix  $\mu((a, b]) = F(b) - F(a)$ , si  $a < b$ ; i  $\mu((a, +\infty)) = 1 - F(a)$ . Per tant, si dues probabilitats  $\mu_1$  i  $\mu_2$  tenen la mateixa funció de distribució  $F$ ,  $\mu_1$  i  $\mu_2$  coincideixen sobre la semialgebra  $\mathcal{S}$ , i, en conseqüència  $\mu_1 = \mu_2$  (veure els apartats 4.3 i 4.6). És a dir, la funció de distribució d'una probabilitat caracteritza completament aquesta probabilitat. D'altra banda, el següent teorema ens diu que, com a conseqüència dels resultats d'estensió de mesures, hi ha equivalència entre funcions de distribució i probabilitats sobre  $\mathbb{R}$ .

**Teorema 5.1.** Donada una funció de distribució  $F$ . Existeix una única probabilitat  $\mu$  sobre  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$  tal que  $F(x) = \mu((-\infty, x])$  per a tot  $x \in \mathbb{R}$ .

*DemOSTRació:* En primer lloc definirem  $\mu$  sobre la semialgebra  $\mathcal{S}$  posant

$$\begin{aligned}\mu((a, b]) &= F(b) - F(a), \text{ si } a < b, \\ \mu((a, +\infty)) &= 1 - F(a), \quad \mu((-\infty, a]) = F(a), \quad \mu(\emptyset) = 0, \quad \mu(\mathbb{R}) = 1.\end{aligned}$$

És fàcil veure que  $\mu$  és additiva en  $\mathcal{S}$ . En efecte, suposem per exemple que  $a < c < b$ , aleshores

$$\mu((a, c]) + \mu((c, b]) = F(c) - F(a) + F(b) - F(c) = F(b) - F(a) = \mu((a, b]).$$

Els altres casos es tracten de manera anàloga.

Veiem ara que  $\mu$  és  $\sigma$ -additiva en  $\mathcal{S}$ . Així podem estendre  $\mu$  a una probabilitat en  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$  (Teorema 4.13) que tindrà  $F$  per funció de distribució i haurem acabat.

Sigui  $I = \bigcup_{n=1}^{\infty} I_n$ , on  $I, I_n \in \mathcal{S}$  i els  $I_n$  són disjunts dos a dos. Recordem que  $\mu$  admet una extensió additiva a l'àlgebra generada per  $\mathcal{S}$ .

Tenim

$$\sum_{n=1}^m \mu(I_n) = \mu\left(\bigcup_{n=1}^m I_n\right) \leq \mu(I),$$

per a tot  $m \geq 1$ . i per tant

$$\sum_{n=1}^{\infty} \mu(I_n) \leq \mu(I).$$

Per a provar la desigualtat contrària és necessari utilitzar la següent propietat:

Donat  $\varepsilon > 0$  i  $J \in \mathcal{S}$ , existeixen  $J_1, J_2 \in \mathcal{S}$  tals que  $J_1$  és compacte,  $J_1 \subset J \subset J_2$ ,  $\mu(J_2) - \mu(J_1) < \varepsilon$  i  $\mu(J_2) - \mu(J) < \varepsilon$ .

En efecte, considerem les diferents possibilitats:

- (i) Si  $J = (a, b]$  prenem  $J_1 = (a', b]$ ,  $J_2 = (a, b']$  amb  $a < a'$ .  $F(a') - F(a) < \varepsilon$ ,  $b < b'$  i  $F(b') - F(b) < \varepsilon$ .

(ii) Si  $J = (a, +\infty)$  prenem  $J_1 = (a', c]$ ,  $J_2 = J$ . on  $a < a'$ .  $F(a') - F(a) < \frac{\varepsilon}{2}$ .  $a' < c$  i  $1 - F(c) < \frac{\varepsilon}{2}$ .

(iii) Si  $J = (-\infty, b]$  prenem  $J_1 = (a', b]$ ,  $J_2 = (-\infty, b']$ , on  $b > a'$ .  $F(a') < \varepsilon$ .  $b < b' i F(b') - F(b) < \varepsilon$ .

Fixem  $\varepsilon > 0$  i sigui  $I' \in \mathcal{S}$ .  $I'$  compacte, tal que  $\bar{I}' \subset I$  i  $\mu(I) - \mu(\bar{I}') < \frac{\varepsilon}{2}$ . També per cada  $n \geq 1$  prenem  $I'_n \in \mathcal{S}$  tals que  $I_n \subset I'_n$ , i  $\mu(I'_n) - \mu(I_n) < \frac{\varepsilon}{2^n}$ . Com que  $\{I'_n, n \geq 1\}$  és un recobriment obert del compacte  $\bar{I}'$  existirà un subrecobriment finit,  $I' \subset \bigcup_{n=1}^m I'_n$ . Llavors

$$\mu(I) < \mu(I') + \frac{\varepsilon}{2} \leq \sum_{n=1}^m \mu(I'_n) + \frac{\varepsilon}{2} \leq \sum_{n=1}^m \mu(I_n) + \sum_{n=1}^m \frac{\varepsilon}{2^{n+1}} + \frac{\varepsilon}{2} \leq \sum_{n=1}^{\infty} \mu(I_n) + \varepsilon,$$

i això acaba la demostració del teorema. ■

Suposem que  $\mu$  és una probabilitat amb funció de distribució  $F$ . Les següents fórmules són fàcils de comprovar i permeten calcular la probabilitat de qualsevol interval:

$$\mu([a, b]) = F(b) - F(a^-),$$

$$\mu([a, b)) = F(b^-) - F(a^-),$$

$$\mu([a]) = F(a) - F(a^-).$$

En particular veiem que  $F$  és contínua en un punt a si i només si  $\mu(\{a\}) = 0$ .

Considerem el cas d'una probabilitat *discreta*,  $\mu = \sum_{i \in I} \rho_i \delta_{a_i}$ , on  $\{a_i, i \in I\} \subset \mathbb{R}$  és un conjunt finit o numerable i els  $\rho_i \in [0, 1]$  són nombres tals que  $\sum_{i \in I} \rho_i = 1$ . L'espai de probabilitat  $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mu)$  correspon a una experiència aleatòria en la qual s'obtenen els resultats  $a_i$  amb probabilitats  $\rho_i$ .

En aquest cas, la funció de distribució de  $\mu$  és purament discontínua:

$$F(x) = \sum_{\{i : a_i \leq x\}} \rho_i.$$

Cal observar que els punts  $\{a_i, i \in I\}$  poden ser aïllats (per exemple,  $\mu = \frac{1}{2} \delta_1 + \frac{1}{2} \delta_{-1}$ ), o bé constituir un conjunt dens en  $\mathbb{R}$  (com  $\mathbb{Q}$ ). En aquest darrer cas la funció de distribució  $F$  és més difícil d'imaginat.

L'equivalència entre funcions de distribució i probabilitats es pot generalitzar immediatament al cas de mesures finites si es substitueix la propietat (iii) de les funcions de distribució per la condició

$$(iii') \quad \lim_{x \rightarrow +\infty} F(x) = \mu(\mathbb{R}).$$

Considerarem una funció de distribució  $F$  d'una probabilitat  $\mu$ . Designarem per  $S$  el conjunt finit o numerable dels seus punts de discontinuitat. Es defineix la part purament discontínua de  $F$  per

$$F_d(x) = \sum_{\{y : y \in S, y \leq x\}} [F(y) - F(y^-)].$$

$F_d$  és la funció de distribució de la mesura disreta  $\mu_d = \sum_{y \in S} [F(y) - F(y^-)] \delta_y$ . Observis que  $S = \{y : \mu(\{y\}) > 0\}$  i que  $\mu_d = \sum_{y \in S} \mu(\{y\}) \delta_y$ . D'aquí es dedueix que la probabilitat  $\mu$  és disreta si i només si  $\mu = \mu_d$ . La diferència  $\mu_c = \mu - \mu_d$  és una mesura (de massa total  $\leq 1$ ) contínua ( $\mu_c(\{x\}) = 0$  per a tot  $x$ ) que té per funció de distribució  $F_c = F - F_d$ . En conclusió, tota probabilitat  $\mu$  sobre  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$  és descomposa de forma única com a suma d'una mesura disreta i una contínua.

Es poden també construir probabilitats sobre la recta mitjançant funcions de densitat, utilitzant la integral de Lebesgue. Per definició una *densitat de probabilitat* és una funció  $f : \mathbb{R} \rightarrow [0, +\infty)$  mesurable i tal que

$$\int_{-\infty}^{+\infty} f(x) dx = 1.$$

Aleshores, la funció definida per  $F(x) = \int_{-\infty}^x f(y) dy$  és una funció de distribució. Si  $\mu$  és la probabilitat sobre  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$  associada a  $F$ , es compleix

$$\mu(A) = \int_A f(x) dx$$

per a tot  $A \in \mathcal{B}(\mathbb{R})$ . En efecte, degut a les propietats de la integral de Lebesgue, la funció  $A \mapsto \int_A f(x) dx$  és una probabilitat que coincideix amb  $\mu$  quan  $A$  és una semirecta del tipus  $(-\infty, x]$ .

Les funcions de distribució (o les probabilitats) d'aquest tipus s'anomenen *absolutament contínues*. Tota funció de distribució absolutament contínua és contínua, ja que per a tot  $x \in \mathbb{R}$

$$F(x) - F(x^-) = \mu(\{x\}) = \int_{\{x\}} f(y) dy = 0.$$

El recíproc no és cert i es poden construir funcions de distribució contínues que no son absolutament contínues mitjançant la utilització del conjunt de Cantor. Si la densitat  $f$  és contínua en un interval obert  $I$ , és fàcil veure que  $F$  és derivable en aquest interval i  $F'(x) = f(x)$  per a tot  $x \in I$ .

Exemples

És a dir, el temps de vida residual a l'instant  $s$  té la mateixa llei de probabilitat que el temps de vida des de l'instant inicial.

1.- *Lleis uniformes:* Si tenim un conjunt finit  $\Omega = \{\omega_1, \dots, \omega_n\}$ , el model probabilístic corresponent a l'elecció a l'atzar d'un element d'aquest conjunt s'obté prenent l'espai de probabilitat  $(\Omega, \mathcal{A}, P)$  on  $\mathcal{A} = \mathcal{P}(\Omega)$  i  $P$  és la probabilitat discreta

$$\sum_{i=1}^n \frac{1}{n} \delta_{\omega_i}.$$

Considerem ara l'experiència aleatòria següent: "Elecció a l'atzar d'un nombre de l'interval  $[a, b]$ ". Una forma aproximada de realitzar aquesta experiència consisteix en utilitzar una ruleta contínua de perímetre  $b - a$ . Observem que el resultat de l'experiència és un nombre real. Que l'elecció sigui a l'atzar vol dir que la probabilitat de que el nombre elegit estigui en un interval  $I \subset [a, b]$  és proporcional a la longitud d'aquest interval. Si  $\mu$  és la probabilitat sobre  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$  associada a aquesta experiència tindrem:

$$\mu(I) = \frac{d - c}{b - a} \quad \text{si} \quad I = [c, d] \subset [a, b],$$

$$\mu([a, b]^c) = 0.$$

$\mu$  és una probabilitat absolutament contínua amb densitat  $f(x) = \frac{1}{b-a} \mathbf{1}_{[a,b]}(x)$ . Aquesta probabilitat s'anomena *lleïa uniforme* en l'interval  $[a, b]$ , i la seva funció de distribució és

$$F(x) = \begin{cases} 0, & \text{si } x \leq a \\ \frac{x-a}{b-a}, & \text{si } x \in [a, b] \\ 1, & \text{si } x \geq b. \end{cases}$$

2.- *Lleis exponencials:* Considerarem una probabilitat  $\mu$  sobre la recta, concentrada en la semirrecta  $[0, +\infty)$ , que representi la llei de probabilitat del temps de vida d'un sistema (un individu, una bombeta elèctrica, etc.). Imposarem a  $\mu$  les següents condicions:

- (i) La funció  $\varphi(t) = \mu((t, +\infty))$ ,  $t \geq 0$  (igual a la probabilitat que el sistema tingui una duració superior a  $t$ ) és contínua i no nul·la.
- (ii) No hi ha envejament. En termes de probabilitats condicionades aquesta propietat s'expressa posant

$$\mu((t+s, +\infty)/(s, +\infty)) = \mu((t, +\infty)), \quad \text{per a tots } s, t \geq 0.$$

És a dir, el temps de vida residual a l'instant  $s$  té la mateixa llei de probabilitat que el temps de vida des de l'instant inicial.

La condició (ii) implica que la funció  $\varphi : [0, +\infty) \rightarrow [0, 1]$  compleix  $\varphi(t+s) = \varphi(t) \cdot \varphi(s)$ . Tenint en compte la condició (i) i que  $\varphi(0) = 1$ , obtenim  $\varphi(t) = e^{-\lambda t}$ , amb  $\lambda > 0$ , ja que  $\lim_{t \rightarrow \infty} \varphi(t) = 0$ .

En conseqüència, la funció de distribució de  $\mu$  és

$$F(x) = \begin{cases} 0, & \text{si } x \leq 0 \\ 1 - e^{-\lambda x}, & \text{si } x > 0. \end{cases}$$

La probabilitat  $\mu$  s'anomena *lleïa exponencial* de paràmetre  $\lambda$  i és absolutament contínua amb densitat

$$f(x) = \lambda e^{-\lambda x} \mathbf{1}_{[0, +\infty)}(x).$$

3.- *Lleïa normal:* És una distribució que juga un paper fonamental en el capital del Teorema central del límit i en l'Estadística. Per definició la lleïa normal  $N(0, 1)$  és una probabilitat sobre  $\mathbb{R}$  que té per densitat la funció

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \quad x \in \mathbb{R}.$$

És fàcil veure que  $f$  és efectivament una densitat de probabilitat, és a dir que  $\int_{\mathbb{R}} f(x) dx = 1$ . En efecte,

$$\begin{aligned} \left( \int_{\mathbb{R}} e^{-\frac{x^2}{2}} dx \right)^2 &= \int_{\mathbb{R}^2} e^{-\frac{(x^2+y^2)}{2}} dx dy \\ &= \int_0^{2\pi} d\theta \int_0^\infty e^{-\frac{r^2}{2}} r dr = 2\pi. \end{aligned}$$

Funcions de distribució n-dimensionals

Els resultats anteriors es poden estendre al cas de probabilitats en  $\mathbb{R}^n$ . Descriurem breument aquesta extensió sense precisar-ne tots els detalls.

Fixats dos punts  $a = (a_1, \dots, a_n)$ ,  $b = (b_1, \dots, b_n)$  tals que  $a_i < b_i$  per a tot  $i = 1, \dots, n$ , es defineix el rectangle n-dimensional  $(a, b)$  per

$$(a, b) = \prod_{i=1}^n (\sigma_i, b_i].$$

La família d'aquests rectangles (on les coordenades  $a_i$  poden ser  $-\infty$  i les  $b_i$  poden ser  $+\infty$ ) afegint-hi el conjunt buit, constitueix una semiàlgebra que genera la  $\sigma$ -àlgebra de Borel  $\mathcal{B}(\mathbb{R}^n)$  de  $\mathbb{R}^n$ . Recordem que la  $\sigma$ -àlgebra de Borel d'un espai topològic és la generada per la col·lecció dels oberts.

Sigui  $\mu$  una probabilitat en l'espai mesurable  $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ . La funció  $F : \mathbb{R}^n \rightarrow [0, 1]$  definida per

$$F(x) = \mu((-\infty, x]),$$

s'anomena la *funció de distribució* de  $\mu$  i té les següents propietats:

(i)  $\Delta_{ba} F \geq 0$ , per a tota paella de punts  $a, b$  tals que  $a_i \leq b_i$ ,  $i = 1, \dots, n$ , on per definició

$$\Delta_{ba} F = \sum_{\substack{\epsilon \in \{0, 1\}^n \\ \epsilon = (\epsilon_1, \dots, \epsilon_n)}} (-1)^{\epsilon_1 + \dots + \epsilon_n} F(b_1 + \epsilon_1(a_1 - b_1), \dots, b_n + \epsilon_n(a_n - b_n)).$$

$\Delta_{ba} F$  representa l'increment  $n$ -dimensional de la funció  $F$  en el rectangle  $(a, b]$ . La positivitat d'aquest increment es dedueix del fet que coincideix amb  $\mu((a, b])$ . Per exemple, si  $n = 2$  tindrem

$$\begin{aligned} \Delta_{ba} F &= F(b_1, b_2) - F(b_1, a_2) - F(a_1, b_2) + F(a_1, a_2) \\ &= \mu((-\infty, b_1] \times (a_1, b_2]) - \mu((-\infty, a_1] \times (a_2, b_2)) \\ &= \mu((a, b]). \end{aligned}$$

(ii)  $F$  és contínua per la dreta en totes les variables simultàniament:

$$\lim_{x_i \downarrow y_i, i=1, \dots, n} F(x) = F(y)$$

(iii) Es compleixen les següents propietats asymptòtiques

$$\lim_{x_i \rightarrow -\infty} F(x) = 0, \quad \text{per a tot } i = 1, \dots, n.$$

$$\lim_{x_i \rightarrow \infty} F(x) = 1.$$

Tota funció  $F : \mathbb{R}^n \rightarrow [0, 1]$  que compleix aquestes quatre propietats s'anomena funció de distribució, i existeix una única probabilitat  $\mu$  en  $\mathbb{R}^n$  tal que  $F(x) = \mu((-\infty, x])$ . La unicitat prové del fet que dues probabilitats  $\mu_1, \mu_2$  amb la mateixa funció de distribució  $F$  han de coincidir sobre la semiàlgebra  $\mathcal{S}$  ja que  $\mu_1((a, b]) = \mu_2((a, b]) = \mu_2((a, b]) = \Delta_{ba} F$ . D'altra

banda, fixada la funció de distribució  $F$ , es defineix  $\mu$  sobre la semiàlgebra  $\mathcal{S}$  posant  $\mu((a, b]) = \Delta_{ba} F$  i es demostra (com el cas  $n = 1$ ) que  $\mu$  és  $\sigma$ -additiva en  $\mathcal{S}$ . Això ens permet d'estendre  $\mu$  a la  $\sigma$ -àlgebra  $\mathcal{B}(\mathbb{R}^n)$ .

Sigui  $\mu$  una probabilitat en  $\mathbb{R}^n$ . El conjunt  $\mathcal{S} = \{x \in \mathbb{R}^n : \mu(\{x\}) > 0\}$  és finit o numerable. Podem definir aleshores la part discreta de  $\mu$  per  $\mu_d = \sum_{x \in \mathcal{S}} \mu(\{x\}) \delta_x$  i  $\mu$  es descompona de forma única en suma d'una mesura discreta  $\mu_d$  i una de contínua  $\mu_c = \mu - \mu_d$ .

Com en el cas unidimensional, tota funció  $f : \mathbb{R}^n \rightarrow [0, +\infty]$  mesurable i tal que  $\int_{\mathbb{R}^n} f(x) dx = 1$  es pot utilitzar com a *densitat de probabilitat*. És a dir, a partir de  $f$  definirem la funció de distribució

$$F(x) = \int_{-\infty}^{x_1} \dots \int_{-\infty}^{x_n} f(y_1, \dots, y_n) dy_1 \dots dy_n.$$

Es diu aleshores que la probabilitat  $\mu$  associada a  $F$  és *absolutament contínua* amb densitat  $f$ . Si  $f$  és una funció contínua, es compleix que  $f = \frac{\partial^n F}{\partial x_1 \dots \partial x_n}$ .

$X : (G, \mathcal{G}) \rightarrow (\Omega, \Sigma)$  és mesurable si i només si  $X_i \circ X : (G, \mathcal{G}) \rightarrow (F_i, \mathcal{F}_i)$  és mesurable per a tot  $i \in I$ .

En efecte, la implicació en un sentit ( $\Rightarrow$ ) és immediata i, d'altra banda, per a tot generador  $X_i^{-1}(\mathcal{A}_i)$  d' $\mathcal{E}$  es compleix  $X^{-1}[X_i^{-1}(\mathcal{A}_i)] = (X_i \circ X)^{-1}(\mathcal{A}) \in \mathcal{G}$ .

**Definició 6.2.** Considerem un espai de probabilitat  $(\Omega, \mathcal{A}, P)$ . Una aplicació mesurable  $X : \Omega \rightarrow \mathbb{R}$  que a cada resultat possible  $\omega$  de l'experiència li fa correspondre un nombre real  $X(\omega)$ . Per exemple, si l'experiència és una jugada en un joc d'atzar, el valor  $X(\omega)$  pot representar el guany del jugador quan es produeix el resultat  $\omega$ . L'esperança matemàtica representa el guany mitjà del jugador, i es defineix com la integral de la funció real  $X$  respecte la mesura  $P$ . Això ens porta a exigir que  $X$  sigui una funció mesurable i, d'altra banda ens interessarà considerar, per raons tècniques, variables aleatoris que prenguin valors en  $\mathbb{R} = [-\infty, +\infty]$ . L'objectiu d'aquest apartat és, per tant, estudiar l'esperança matemàtica d'una variable aleatoria com a cas particular de la integració de funcions mesurables reals.

## Propietats de les funcions mesurables reals

En tot el que segueix farem servir la notació  $X^{-1}(B) = \{X \in B\}$ .

- 1.-  $X : \Omega \rightarrow \mathbb{R}$  s'anomena una *variable aleatòria*. Analogament, una aplicació mesurable  $X : \Omega \rightarrow \mathbb{R}^n$  s'anomena un *vector aleatori*.
- 2.- Suposem que  $X_i : \Omega \rightarrow \mathbb{R}$ ,  $i = 1, \dots, n$  són mesurables. Aleshores aquesta condició és equivalent a que l'aplicació  $X : \Omega \rightarrow \mathbb{R}^n$ ,  $X = (X_1, \dots, X_n)$  sigui mesurable. En efecte, en un sentit és una conseqüència de la propietat

$$X^{-1}((-\infty, a]) = \{X_1 \leq a_1, \dots, X_n \leq a_n\} \in \mathcal{A},$$

on  $a = (a_1, \dots, a_n)$ .

D'altra banda  $X_i^{-1}((-\infty, a]) = X^{-1}(\mathbb{R} \times \dots \times (-\infty, a] \times \dots \times \mathbb{R})$ , per a tot  $a \in \mathbb{R}$ . La qual cosa prova l'altra implicació. Per tant les funcions del tipus  $h(X_1, \dots, X_n)$  on  $h : \mathbb{R}^n \rightarrow \mathbb{R}$  és mesurable, seran també mesurables. Per exemple  $X_1 + \dots + X_n$ ,  $\sup(X_1, \dots, X_n)$ ,  $\inf(X_1, \dots, X_n)$  són mesurables.

- 3.- Si  $\{X_n, n \geq 1\}$  és una successió de variables aleatòries, també seran mesurables les funcions  $\inf_n X_n$ ,  $\sup_n X_n$ ,  $\liminf_n X_n$  i  $\limsup_n X_n$ .

Aquesta propietat es dedueix de les següents relacions:

- $\{\sup_n X_n \leq a\} = \cap_n \{X_n \leq a\} \in \mathcal{A}$  per a tot  $a \in \mathbb{R}$ .
  - $\sup_n X_n = -\inf_n(-X_n)$ .
  - $\liminf_n X_n = \sup \inf_n X_m$ .
  - $\limsup_n X_n = \inf \sup_{m \geq n} X_m$ .
- Estructures iniciales:** Considerarem una família  $\{(F_i, \mathcal{F}_i), i \in I\}$  d'espais mesurables i una família d'aplicacions  $\{X_i : \Omega \rightarrow F_i, i \in I\}$ , on  $\Omega$  és un conjunt arbitrari. Aleshores, la mínima  $\sigma$ -àlgebra de parts d' $\Omega$  que fa totes les  $X_i$  mesurables és la  $\sigma$ -àlgebra
- $$\mathcal{E} = \sigma\{X_i^{-1}(\mathcal{A}_i); \mathcal{A}_i \in \mathcal{F}_i, i \in I\}.$$
- Aquesta  $\sigma$ -àlgebra té la propietat següent:

4.- Considerem un subconjunt  $A \subset \Omega$ . Aleshores la funció indicatriu  $\mathbf{1}_A : \Omega \rightarrow \mathbb{R}$  és measurable si i només si  $A \in \mathcal{A}$ .

**Definició 6.3.** Una funció measurable  $X : \Omega \rightarrow \mathbb{R}$  direm que és *elemental* si només pren un nombre finit de valors diferents. Això és equivalent a dir que  $X$  és una combinació lineal finita de funcions indicatrius measurable.  $X = \sum_{i=1}^n a_i \mathbf{1}_{A_i}$ , on  $a_i \neq a_j$  i  $A_i \cap A_j = \emptyset$  si  $i \neq j$ . Cal observar que una representació d'aquest tipus és única ja que els  $\{a_1, \dots, a_n\}$  són els valors que pren la variable  $X$  i d'altra banda  $A_i = \{X = a_i\}$ . El conjunt  $\mathcal{E}$  de les funcions elementals és un espai vectorial ja que en fer combinacions lineals d'elements d' $\mathcal{E}$  s'obteneu de nou funcions measurable que prenen només un conjunt finit de valors diferents.

**Proposició 6.4.** Tota funció measurable  $X : \Omega \rightarrow [0, +\infty]$  és límit creixent d'una successió de funcions elementals no negatives.

*Demostració:* Definim

$$X_n = \sum_{k=1}^{2^n-1} \frac{k}{2^n} \mathbf{1}_{\{\frac{k}{2^n} \leq X < \frac{k+1}{2^n}\}} + n \mathbf{1}_{\{n \leq X\}}.$$

En primer lloc cal veure que la successió de funcions elementals  $\{X_n, n \geq 1\}$  és creixent. Fixem  $\omega \in \Omega$ ,  $n \geq 1$  i suposem que  $X(\omega) \in [k2^{-n}, (k+1)2^{-n}]$ . Aleshores  $X_n(\omega) = k2^{-n}$

$$X_{n+1}(\omega) = \begin{cases} k2^{-n} & \text{si } X(\omega) \in [k2^{-n}, (2k+1)2^{-(n+1)}] \\ (2k+1)2^{-(n+1)} & \text{si } X(\omega) \in [(2k+1)2^{-(n+1)}, (k+1)2^{-n}]. \end{cases}$$

Per tant veiem que  $X_n(\omega) \leq X_{n+1}(\omega)$ . Es faria un raonament analog en el cas  $X(\omega) \geq n$ .

Finalment hem de provar que  $\lim X_n(\omega) = X(\omega)$  per a tot  $\omega \in \Omega$ . En efecte, si  $X(\omega) = \infty$ ,  $X_n(\omega) = n$  per a tot  $n \geq 1$ , i si  $X(\omega) < \infty$  aleshores  $|X(\omega) - X_n(\omega)| \leq 2^{-n}$  per  $n > X(\omega)$ . ■

**Corollari 6.5.** Tota funció measurable  $X : \Omega \rightarrow \mathbb{R}$  és límit puntual de funcions elementals.

*Demostració:* Només cal fer la descomposició  $X = X^+ - X^-$ , on  $X^+ = \sup(X, 0)$ ,  $X^- = \sup(-X, 0)$  i aplicar la proposició anterior a les funcions  $X^+$  i  $X^-$ . ■

## 6.2. Integració respecte d'una mesura

Considerem un espai de mesura  $(\Omega, \mathcal{A}, \mu)$ . Sigui  $X = \sum_{i=1}^n a_i \mathbf{1}_{A_i}$  una funció elemental no negativa. Recorden que els  $a_i$  són diferents entre si i els  $A_i$  són disjunts dos a dos. Es defineix la integral de  $X$  respecte  $\mu$  per

$$\int_{\Omega} X d\mu = \sum_{i=1}^n a_i \mu(A_i). \quad (6.1)$$

Farem el conveni  $0 \cdot \infty = 0$  quan algun dels conjunts  $A_i$  tingui mesura infinita i  $a_i$  valgui zero. La integral pren valors en  $[0, +\infty]$ . Utilitzarem també les següents notacions per la integral:

$$\int_{\Omega} X(\omega) d\mu(\omega), \quad \text{ó} \quad \int_{\Omega} X(\omega) \mu(d\omega).$$

### 6.2.1. Propietats de la integral de funcions elementals no negatives

- (i) Si  $a$  és un nombre real no negatiu es compleix  $\int_{\Omega} a X d\mu = a \int_{\Omega} X d\mu$ .
- (ii)  $\int_{\Omega} (X + Y) d\mu = \int_{\Omega} X d\mu + \int_{\Omega} Y d\mu$ .

*Demostració:* Considerem dues funcions elementals no negatives  $X = \sum_{i=1}^n a_i \mathbf{1}_{A_i}$ ,  $Y = \sum_{j=1}^m b_j \mathbf{1}_{B_j}$ . Suposem que  $X + Y$  pren els valors diferents  $c_1, \dots, c_p$  en els conjunts  $C_1, \dots, C_p$ . Aleshores,

$$\begin{aligned} \int_{\Omega} (X + Y) d\mu &= \sum_{k=1}^p c_k \mu(C_k) = \sum_{k=1}^p \sum_{i=1}^n \sum_{j=1}^m c_k \mu(C_k \cap A_i \cap B_j) \\ &= \sum_{k=1}^p \sum_{i=1}^n \sum_{j=1}^m (a_i + b_j) \mu(C_k \cap A_i \cap B_j) = \sum_{i=1}^n \sum_{j=1}^m (a_i + b_j) \mu(A_i \cap B_j) \\ &= \sum_{i=1}^n a_i \mu(A_i) + \sum_{j=1}^m b_j \mu(B_j) = \int_{\Omega} X d\mu + \int_{\Omega} Y d\mu. \end{aligned}$$

(iii) Si  $X \leq Y$ , aleshores  $\int_{\Omega} X d\mu \leq \int_{\Omega} Y d\mu$ .

*Demostració:* Suposem  $X = \sum_{i=1}^n a_i \mathbf{1}_{A_i}$  i  $Y = \sum_{j=1}^m b_j \mathbf{1}_{B_j}$ . Aleshores,

$$\begin{aligned} \int_{\Omega} X d\mu &= \sum_{i=1}^n a_i \mu(A_i) = \sum_{i=1}^n \sum_{j=1}^m a_i \mu(A_i \cap B_j) \leq \sum_{i=1}^n \sum_{j=1}^m b_j \mu(A_i \cap B_j) \\ &= \sum_{j=1}^m b_j \mu(B_j) = \int_{\Omega} Y d\mu. \end{aligned}$$

ja que  $a_i \leq b_j$  si  $A_i \cap B_j \neq \emptyset$ .

(iv) Si  $X_n$  és una successió creixent de variables aleatòries simples i positives i  $X$  és la variable aleatòria  $X = \lim_{n \rightarrow \infty} X_n$ , aleshores  $\int_{\Omega} X_n d\mu \uparrow \int_{\Omega} X d\mu$ .

**Demostració:** Suposem que  $X = \sum_{i=1}^p a_i 1_{A_i}$  i  $X_n = \sum_{j=1}^{p_n} a_j^n 1_{A_j}$ . Per la propietat (iii) sabem que  $\lim_n \int_{\Omega} X_n d\mu \leq \int_{\Omega} X d\mu$ . Hem de demostrar la desigualtat contrària. Com que

$$\int_{\Omega} X d\mu = \sum_{i=1}^p a_i \mu(A_i)$$

$$\int_{\Omega} X_n d\mu = \sum_{j=1}^{p_n} a_j^n \mu(A_j^n) = \sum_{i=1}^p \sum_{j=1}^{p_n} a_j^n \mu(A_i \cap A_j^n).$$

És doncs suficient veure que

$$a_i \mu(A_i) \leq \lim_n \sum_{j=1}^{p_n} a_j^n \mu(A_i \cap A_j^n), \quad (6.2)$$

per a tot  $i = 1, \dots, p$ . Si  $a_i = 0$  aquesta desigualtat és trivial. Suposem  $a_i > 0$  i fixem un nombre  $0 < \lambda < 1$ . Tindrem

$$\begin{aligned} \sum_{j=1}^{p_n} a_j^n \mu(A_i \cap A_j^n) &= \sum_{j: a_j^n > \lambda a_i} a_j^n \mu(A_i \cap A_j^n) + \sum_{j: a_j^n \leq \lambda a_i} a_j^n \mu(A_i \cap A_j^n) \\ &\geq \sum_{j: a_j^n > \lambda a_i} \lambda a_i \mu(A_i \cap A_j^n) = \lambda a_i \mu([X_n > \lambda a_i] \cap A_i]. \end{aligned}$$

Com que  $X_n \uparrow a_i$  sobre el conjunt  $A_i$ , la successió de conjunts  $\{X_n > \lambda a_i\} \cap A_i$  creix cap a  $A_i$ . En conseqüència, fent  $n \rightarrow \infty$  obtenim

$$\lim_n \sum_{j=1}^{p_n} a_j^n \mu(A_i \cap A_j^n) \geq \lambda a_i \mu(A_i),$$

i com que  $\lambda$  és arbitrari, fent tendir  $\lambda$  a 1 s'obté la desigualtat (6.2); la propietat (iv) queda demostrada.

La segona etapa en la construcció de la integral consisteix en definir la integral d'una funció measurable  $X : \Omega \rightarrow [0, +\infty]$ . Posarem, per definició,

$$\int_{\Omega} X d\mu = \lim_n \int_{\Omega} X_n d\mu.$$

on  $X_n$  és una successió de funcions elementals no negatives que creix cap a  $X$ . Dugent a la Proposició 6.4 sempre existeix una successió d'aquest tipus. Hem de veure que aquesta definició és consistent, és a dir, el valor de  $\int_{\Omega} X d\mu$  no depèn de la successió  $X_n$  escollida. Això es dedueix del següent resultat.

**Lema 6.6.** Sigui  $\{X_n, n \geq 1\} \subset \{Y_n, n \geq 1\}$  dues successions creixents de funcions elementals no negatives. Aleshores si  $\lim_n X_n \leq \lim_n Y_n$  es compleix  $\int_{\Omega} X_n d\mu \leq \lim_n \int_{\Omega} Y_n d\mu$ .

**DemOSTRACIÓ:** Fixat  $m$  tenim que  $X_m = \lim_n \uparrow \inf(X_m, Y_n)$ . Per tant, utilitzant la propietat (iv) s'obté

$$\int_{\Omega} X_m d\mu = \lim_n \int_{\Omega} \inf(X_m, Y_n) d\mu \leq \lim_n \int_{\Omega} Y_n d\mu$$

i fent  $m \rightarrow \infty$  el lema queda demostrat. ■

### 6.2.2. Propietats de la integral de funcions mesurables no negatives

Les propietats (i), (ii), (iii) de linealitat i monotonía es dedueixen per pas al límit de les mateixes propietats per les funcions elementals no negatives.

Demostrarem a continuació la propietat de convergència monòtona i altres propietats interessants.

(iv) Si  $\{X_n, n \geq 1\}$  és una successió creixent de funcions mesurables no negatives que convergeix puntualment cap a  $X$ , aleshores  $\int_{\Omega} X_n d\mu \uparrow \int_{\Omega} X d\mu$ .

**DemOSTRACIÓ:** Per cada  $n \geq 1$ , signi  $\{Y_{m,n}, m \geq 1\}$  una successió creixent de funcions elementals no negatives tal que  $\lim_m Y_{m,n} = Y_n$ . Possem  $Z_m = \sup_{n \leq m} Y_{m,n}$ , que és una funció elemental, tal que  $Z_m \leq Z_{m+1}$ ; i  $Y_{m,n} \leq Z_m \leq X_n$  si  $m \geq n$ . En conseqüència,

$$\int_{\Omega} Z_m d\mu \leq \int_{\Omega} Z_{m+1} d\mu,$$

$$\int_{\Omega} Y_{m,n} d\mu \leq \int_{\Omega} Z_m d\mu \leq \int_{\Omega} X_n d\mu,$$

Fent primer  $m \rightarrow \infty$  i després  $n \rightarrow \infty$  s'obté

$$X = \lim_n X_n = \lim_m Z_m.$$

$$\lim_n \int_{\Omega} X_n d\mu = \lim_m \int_{\Omega} Z_m d\mu = \int_{\Omega} X d\mu.$$

(v) *Lema de Fatou.* Per a tota successió  $\{X_n, n \geq 1\}$  de funcions mesurables positives es compleix

$$\int_{\Omega} (\liminf_n X_n) d\mu \leq \liminf_n \int_{\Omega} X_n d\mu.$$

*Demostració:* La successió  $\inf_{m \geq n} X_m$  creix cap a  $\liminf X_n$ . Aleshores, utilitzant la propietat (iv) de convergència monòtona s'obté

$$\begin{aligned} \int_{\Omega} (\liminf_n X_n) d\mu &= \int_{\Omega} \lim_n (\inf_{m \geq n} X_m) d\mu \\ &= \lim_n \int_{\Omega} (\inf_{m \geq n} X_m) d\mu \leq \liminf_n \int_{\Omega} X_m d\mu = \liminf_n \int_{\Omega} X_n d\mu. \end{aligned}$$

(vi) Si  $\int_{\Omega} X d\mu < \infty$ , aleshores  $\mu\{X = \infty\} = 0$ . És a dir, si una funció té integral finita, la funció és finita a menys d'un conjunt de mesura zero.

*Demostració:* Posem  $m = \int_{\Omega} X d\mu$ . Tindrem  $m \geq \int_{\Omega} X 1_{\{X \geq n\}} d\mu \geq n\mu\{X \geq n\}$ , i en conseqüència,  $\mu\{X = \infty\} = \lim_n \mu\{X \geq n\} = 0$ .

(vii) Si  $\int_{\Omega} X d\mu = 0$ , aleshores  $\mu\{X > 0\} = 0$ , i reciprocament. És a dir, la integral d'una funció és zero si i només si la funció positiva és zero a menys d'un conjunt de mesura nula.

*Demostració:* Suposem primer que  $\int_{\Omega} X d\mu = 0$ . Tindrem  $\mu\{X > 0\} = \lim_n \mu\{X \geq \frac{1}{n}\} \leq \lim_n \int_{\Omega} n X d\mu = 0$ , ja que  $1_{\{X \geq 1/n\}} \leq n X$ .

Recíprocament,

$$\int_{\Omega} X d\mu = \int_{\Omega} X 1_{\{X > 0\}} d\mu \leq \lim_n \int_{\Omega} n 1_{\{X > 0\}} d\mu = 0.$$

Considerarem dues funcions mesurables  $X, Y : \Omega \rightarrow \bar{\mathbb{R}}$ . Direm que  $X$  i  $Y$  coincideixen  $\mu$ -quasi per tot si  $\mu\{X \neq Y\} = 0$ . En aquest cas escriurem  $X = Y$ ,  $\mu$ -q.p.t. Si  $\mu$  és una probabilitat i  $\mu\{X \neq Y\} = 0$  es diu aleshores que  $X$  i  $Y$  coinciden quasi segurament i s'escriu  $X = Y$ ,  $\mu$ -q.s.

En la teoria de la probabilitat, dues variables aleatòries que coincideixen quasi segurament es poden identificar. De moment, a partir de la propietat (vii) tenim que si  $X, Y$  són

funcions mesurables no negatives.

$$X = Y \mu - q.p.t. \Rightarrow \int_{\Omega} X d\mu = \int_{\Omega} Y d\mu.$$

En efecte, si posem  $A = \{X \neq Y\}$ , per hipòtesi  $\mu(A) = 0$ , per tant

$$\int_{\Omega} X d\mu = \int_{\Omega} X 1_A d\mu = \int_{\Omega} Y 1_A d\mu = \int_{\Omega} Y d\mu.$$

ja que les funcions  $X 1_A$  i  $Y 1_A$  són zero a menys d'un conjunt de mesura nula.

En general, direm que una determinada propietat és certa  $\mu$ -quasi per tot si la propietat es satisfà per tots els elements fora d'un conjunt de mesura zero.

La darrera etapa en la construcció de la integral ve donada per la següent definició.

**Definició 6.7.** Direm que una funció mesurable  $X : \Omega \rightarrow \bar{\mathbb{R}}$  és integrable respecte  $\mu$  si  $\int_{\Omega} |X| d\mu < \infty$ . En aquest cas, definirem la integral de  $X$  per

$$\int_{\Omega} X d\mu = \int_{\Omega} X^+ d\mu - \int_{\Omega} X^- d\mu. \quad (6.3)$$

Recordem que  $X = X^+ - X^-$  i  $|X| = X^+ + X^-$ . Observi's que la condició  $\int_{\Omega} |X| d\mu < \infty$  implica que la funció  $X$  és finita a menys d'un conjunt de mesura zero.

### 6.2.3. Propietats de la integral

(i) Si  $X$  és integrable, tenim que per a tot  $a$  de  $\mathbb{R}$  la variable  $aX$  és integrable i

$$\int_{\Omega} aX d\mu = a \int_{\Omega} X d\mu.$$

(ii) Si  $X$  i  $Y$  són integrables,  $X + Y$  també ho és i es compleix

$$\int_{\Omega} (X + Y) d\mu = \int_{\Omega} X d\mu + \int_{\Omega} Y d\mu.$$

*Demostació:* La variable aleatòria  $X + Y$  és integrable ja que

$$(X + Y)^+ \leq X^+ + Y^+ \quad i \quad (X + Y)^- \leq X^- + Y^-.$$

D'altra banda, la relació  $(X + Y)^+ - (X + Y)^- = X^+ + Y^+ - (X^- + Y^-)$  implica

$$\begin{aligned} \int_{\Omega} (X + Y) d\mu &= \int_{\Omega} (X^+ + Y^+) d\mu - \int_{\Omega} (X^- + Y^-) d\mu \\ &= \int_{\Omega} [X^+ + Y^+ - (X^- + Y^-)] d\mu = \int_{\Omega} X d\mu - \int_{\Omega} Y d\mu. \end{aligned}$$

(iii) Si  $X$  i  $Y$  són integrables i  $X \leq Y$ , aleshores  $\int_{\Omega} X d\mu \leq \int_{\Omega} Y d\mu$ .

*Demostració:* Fent servir la linealitat de la integral, tindrem

$$X \leq Y \Rightarrow Y - X \geq 0 \Rightarrow \int_{\Omega} Y d\mu - \int_{\Omega} X d\mu \geq 0.$$

(iv) Si  $X$  és integrable, aleshores

$$\left| \int_{\Omega} X d\mu \right| \leq \int_{\Omega} |X| d\mu.$$

(v) *Teorema de convergència monotona.* Sigui  $\{X_n, n \geq 1\}$  una successió creixent de variables aleatories integrables amb límit  $X$ . Suposem que  $\sup_n \int_{\Omega} X_n d\mu < \infty$ . Aleshores  $X$  és integrable i

$$\int_{\Omega} X_n d\mu \uparrow \int_{\Omega} X d\mu.$$

Igualment, si  $\{X_n, n \geq 1\}$  és una successió decreixent de variables aleatories integrables amb límit  $X$ , i si es compleix la condició  $\inf_n \int_{\Omega} X_n d\mu > -\infty$ , es té que  $X$  és integrable respectivament i aplicar les propietats de monotonía de la integral de funcions mesurables no negatives. El cas decreixent es redueix al cas creixent mitjançant un canvi de signe.

$$\int_{\Omega} X_n d\mu \downarrow \int_{\Omega} X d\mu.$$

*Demostració:* En el cas creixent només cal substituir  $X_n$  i  $X$  per  $X_n - X_1$  i  $X - X_1$  respectivament i aplicar les propietats de monotonía de la integral de funcions mesurables no negatives. El cas decreixent es redueix al cas creixent mitjançant un canvi de signe.

(vi) *Lema de Fatou.*

(a) Sigui  $\{X_n, n \geq 1\}$  una successió de variables aleatories que compleix les següents propietats:

- (1)  $X_n \geq X$  per a tot  $n \geq 1$ , on  $X$  és una variable aleatorià.
- (2)  $X, \liminf X_n$  són integrables.

Aleshores

$$\int_{\Omega} (\liminf X_n) d\mu \leq \liminf \int_{\Omega} X_n d\mu.$$

(b) Anàlogament, si  $\{X_n, n \geq 1\}$  és una successió de variables aleatories que compleix

- (1')  $X_n \leq X$  per a tot  $n \geq 1$ , on  $X$  és una variable aleatorià.
- (2')  $X, \limsup X_n$  són integrables.

Aleshores

$$\int_{\Omega} (\limsup X_n) d\mu \geq \limsup \int_{\Omega} X_n d\mu.$$

*Demostració:* Com abans, la primera propietat es dedueix immediatament del Lema de Fatou per funcions no negatiuves i per la segona només cal fer un canvi de signe.

(vii) *Teorema de convergència dominada de Lebesgue.* Sigui  $\{X_n, n \geq 1\}$  una successió de variables aleatories que compleixen les següents propietats:

(1)  $\lim_n X_n = X$ .

(2)  $|X_n| \leq Y$ , per a tot  $n \geq 1$ , on  $Y$  és integrable.

Aleshores  $X$  és integrable i

$$\int_{\Omega} X d\mu = \lim \int_{\Omega} X_n d\mu.$$

Bàsicament aquest resultat ens diu que si una successió de funcions està dominada per una funció integrable, es pot commutar el límit amb la integral. Observis que aquesta comutació és també possible si la convergència és monòtona (proprietat (v)).

*Demostració:* La relació  $|X_n| \leq Y$  implica que  $X$  és integrable i aplicant el Lema de Fatou tindrem

$$\begin{aligned} \int_{\Omega} X d\mu &= \int_{\Omega} (\liminf X_n) d\mu \leq \liminf \int_{\Omega} X_n d\mu \\ &\leq \limsup \int_{\Omega} X_n d\mu \leq \int_{\Omega} (\limsup X_n) d\mu = \int_{\Omega} Y d\mu. \end{aligned}$$

Si una funció integrable la modifiquem en un conjunt de mesura nula (de manera que segueixi essent mesurable), la nova funció és integrable i el valor de la integral no ha canviat. Aquesta propietat que ja havíem establert per les funcions no negatives s'estén clarament a les funcions integrables.

En particular, totes les propietats de les funcions integrables són també vàlides si les relacions de desigualtat o convergència són certes  $\mu$ -quasi per tot.

Si  $X$  és una funció integrable i  $A \in \mathcal{A}$ , la funció  $X 1_A$  és integrable i escriurem per definició:

$$\int_A X d\mu = \int_{\Omega} (X 1_A) d\mu.$$

En el cas d'una funció mesurable no negativa  $X$ , es pot comprovar fàcilment que la funció  $\nu(A) = \int_A X d\mu$  és una mesura. Direm que aquesta mesura  $\nu$  és *absolutament contínua* respecte  $\mu$  amb densitat  $X$  i escriurem

$$X = \frac{d\nu}{d\mu}.$$

Si la mesura  $\nu$  és finita (o  $\sigma$ -finita), la densitat, si existeix, és única a mènys de conjunts de mesura  $\mu$  zero, com es dedueix del següent resultat.

**Proposició 6.8.** Signi  $X$  una funció integrable en un espai de mesura  $(\Omega, \mathcal{A}, \mu)$  tal que  $\int_A X d\mu = 0$  per a tot  $A \in \mathcal{A}$ . Aleshores  $X = 0$ ,  $\mu$ -q.p.t.

*Demostració:* tenim, en particular,

$$\int_{\{X > 0\}} X d\mu = 0 \quad i \quad \int_{\{X < 0\}} (-X) d\mu = 0.$$

Mitjançant les propietats de la integral de les funcions no negatives s'obté  $\mu(\{X > 0\}) = \mu(\{X < 0\}) = 0$ . És a dir,  $\mu(\{X \neq 0\}) = 0$ . ■

### 6.3. Càcul d'integrals

Considerem una funció integrable  $X : \Omega \rightarrow \mathbb{R}$  definida en un espai de mesura  $(\Omega, \mathcal{A}, \mu)$ . En aquest apartat veurem alguns criteris pel càlcul de la integral  $\int_\Omega X d\mu$ .

#### 6.3.1. Cas d'una mesura discreta

Signi  $\mu = \sum_{i \in I} \alpha_i \delta_{\omega_i}$  on  $I$  és un conjunt finit o numerable,  $\alpha_i \in [0, +\infty]$  i els  $\omega_i$  són punts d' $\Omega$  diferents entre si. En aquest cas tenim:

(i) Si  $X : \Omega \rightarrow [0, +\infty]$  és mesurable,

$$\int_\Omega X d\mu = \sum_{i \in I} \alpha_i X(\omega_i).$$

En efecte, si  $X$  és igual a  $1_A$ ,  $A \in \mathcal{A}$  el resultat és immediat, i en el cas general només cal utilitzar les propietats de linealitat i convergència monòtona de la integral.

(ii) Una funció mesurable  $X : \Omega \rightarrow \mathbb{R}$  és integrable si i només si  $\sum_{i \in I} |X(\omega_i)| \alpha_i < \infty$ , i en aquest cas,

$$\int_\Omega X d\mu = \sum_{i \in I} \alpha_i X(\omega_i).$$

#### 6.3.2. Mesures donades per densitats

Signi  $(\Omega, \mathcal{A}, \mu)$  un espai de mesura i  $X : \Omega \rightarrow [0, +\infty]$  mesurable. Considerem la mesura  $\nu$  que té densitat  $X$  respecte  $\mu$  i que ve donada per  $\nu(A) = \int_A X d\mu$ ,  $A \in \mathcal{A}$ . Aleshores

(i) Per a tota funció measurable  $Y : \Omega \rightarrow [0, +\infty]$  es compleix

$$\int_\Omega Y d\nu = \int_\Omega Y X d\mu.$$

Com abans, aquest resultat és immediat si  $Y = 1_A$ ,  $A \in \mathcal{A}$  i en el cas general s'obté per linealitat i convergència monòtona.

(ii) Una funció measurable  $Y : \Omega \rightarrow \mathbb{R}$  és integrable respecte  $\nu$  si i només si  $\int_\Omega |Y| X d\mu < \infty$ , i en aquest cas tenim també que

$$\int_\Omega Y d\nu = \int_\Omega Y X d\mu.$$

#### 6.3.3. Mesura imatge

Considerem un espai de mesura  $(\Omega, \mathcal{A}, \mu)$ , un espai measurable  $(E, \mathcal{E})$  i una aplicació measurable  $X : \Omega \rightarrow E$ . L'aplicació  $X$  induceix una mesura en l'espai  $(E, \mathcal{E})$  donada per

$$(\mu \circ X^{-1})(B) = \mu(X^{-1}(B)), \quad B \in \mathcal{E}.$$

La mesura  $\mu \circ X^{-1}$  s'anomena la mesura imatge de  $\mu$  per  $X$ . Es compleix aleshores el següent resultat.

**Proposició 6.9.** Signi  $f : E \rightarrow \mathbb{R}$  una funció measurable. Llavors  $f$  és integrable en l'espai de mesura  $(E, \mathcal{E}, \mu \circ X^{-1})$  si i només si  $f \circ X$  és integrable en  $(\Omega, \mathcal{A}, \mu)$  i en aquest cas

$$\int_\Omega (f \circ X) d\mu = \int_E f d(\mu \circ X^{-1}). \quad (6.4)$$

*Demostració:* Com en d'altres demostracions de teoria de la mesura seguirem les següents etapes:

(i) Suposem primer que la funció  $f$  és un indicador,  $f = 1_B$ ,  $B \in \mathcal{E}$ . Llavors  $f \circ X = 1_{X^{-1}(B)}$  i en conseqüència,

$$\int_\Omega (1_B \circ X) d\mu = \mu(X^{-1}(B)) = (\mu \circ X^{-1})(B) = \int_E 1_B d(\mu \circ X^{-1}).$$

(iii) Si  $f$  és una funció simple,  $f = \sum_{i=1}^n a_i \mathbf{1}_{B_i}$ ,  $B_i \in \mathcal{E}$ ,  $a_i \in [0, +\infty]$  tindrem que  $f \circ X = \sum_{i=1}^n a_i \mathbf{1}_{X^{-1}(B_i)}$ , i per linearitat es comprova també la igualtat (6.4).

(iv) La relació (6.4) és certa per a tota funció mesurable  $f : E \rightarrow [0, +\infty]$ . En efecte, si  $f_n$  és una successió de funcions simples tal que  $0 \leq f_n \uparrow f$ , tindrem

$$\int_E (f \circ X) d\mu = \lim_n \int_\Omega (f_n \circ X) d\mu = \lim_n \int_E f_n d(\mu \circ X^{-1}) = \int_E f d(\mu \circ X^{-1}).$$

(iv) Si  $f : E \rightarrow \mathbb{R}$  és una funció mesurable, la igualtat  $\int_E |f| \circ X d\mu = \int_\Omega |f| d(\mu \circ X^{-1})$  ens diu que  $f$  és integrable respecte  $\mu \circ X^{-1}$  si i només si  $f \circ X$  ho és respecte  $\mu$ . Finalment, en aquest cas tindrem

$$\begin{aligned} \int_\Omega (f \circ X) d\mu &= \int_\Omega (f^+ \circ X) d\mu - \int_\Omega (f^- \circ X) d\mu = \\ &= \int_E f^+ d(\mu \circ X^{-1}) - \int_E f^- d(\mu \circ X^{-1}) = \int_E f d(\mu \circ X^{-1}). \end{aligned}$$

#### 6.4. Llei d'una variable aleatòria. Esperança matemàtica

Considerem un espai de probabilitat  $(\Omega, \mathcal{A}, P)$ . Recordem que una variable aleatòria definida en aquest espai és una funció measurable  $X : \Omega \rightarrow \mathbb{R}$ .

Tota variable aleatòria  $X$  induceix una probabilitat sobre la recta,  $P_X = P \circ X^{-1}$ , que s'anomena la *llei* o la *distribució de probabilitat* de la variable  $X$ . La probabilitat  $P_X$  és la mesura imatge de  $P$  per  $X$ , és a dir,

$$P_X(B) = P(X^{-1}(B)), \quad B \in \mathcal{B}(\mathbb{R}).$$

Anomenarem funció de distribució de la variable  $X$  a la funció de distribució de la seva llei. O sigui

$$F_X(x) = P_X((-\infty, x]) = P\{X \leq x\}, \quad x \in \mathbb{R}.$$

Si la variable aleatòria  $X$  és integrable respecte  $P$  (o bé  $X$  pren valors en  $[0, +\infty]$ ) es defineix l'*esperança matemàtica* d' $X$  com la integral d' $X$  respecte la mesura  $P$ . És a dir,

$$E(X) = \int_\Omega X dP.$$

El teorema de la mesura imatge (Proposició 6.9) ens permet aleshores, calcular l'esperança d'una variable aleatòria mitjançant una integral en la recta real:

$$E(X) = \int_\Omega X dP = \int_{\mathbb{R}} x(P \circ X^{-1})(dx).$$

Més generalment, si  $f : \mathbb{R} \rightarrow \mathbb{R}$  és una funció measurable tal que  $f \circ X$  és integrable respecte  $P$  (o bé  $f$  pren valors en  $[0, +\infty]$ ), tindrem

$$E[f(X)] = \int_\Omega f(X) dP = \int_{\mathbb{R}} f(x)(P \circ X^{-1})(dx). \quad (6.5)$$

Si la llei de la variable  $X$  és discreta,  $P_X = \sum_{i \in I} \alpha_i \delta_{x_i}$ ,  $x_i \in \mathbb{R}$ ,  $\alpha_i \in [0, 1]$ ,  $\sum_{i \in I} \alpha_i = 1$ ,  $I$  finit o numerable, aleshores la integral que apareix en el darrer membre de l'expressió (6.5) es pot calcular fàcilment (vegi's l'apartat 6.3.1):

$$E[f(X)] = \sum_{i \in I} f(x_i) \alpha_i.$$

Observi's que en aquest cas la variable  $X$  pren els valors  $x_i$  amb probabilitats  $\alpha_i$ .

Si la llei de la variable  $X$  és absolutament contínua respecte la mesura de Lebesgue, amb densitat  $\varphi(x)$  tindrem (vegi's 6.3.2)

$$E[f(X)] = \int_{\mathbb{R}} f(x) \varphi(x) dx.$$

Per a cada natural  $n \geq 1$ , l'esperança  $E(X^n)$  (si existeix) s'anomena el *moment d'ordre n* de la variable aleatòria  $X$ .

Si  $E(X^2) < \infty$ , es defineix la *variància* de la variable  $X$  com

$$\sigma^2(X) = E[(X - E(X))^2] = E(X^2) - (E(X))^2.$$

Cal fer notar que  $E(X^2) < \infty$  implica  $E(|X|) < \infty$ , ja que  $E(|X|) \leq \sqrt{E(X^2)}$ , com a conseqüència de la desigualtat de Schwarz que venrem més endavant.

Els conceptes d'esperança, moment d'ordre  $n$  i variància que hem introduït en relació a una variable aleatòria, només depenen de la seva llei. És a dir que podem parlar també de moment d'ordre  $n$  i variància d'una probabilitat  $\mu$  sobre  $\mathbb{R}$ , que es defineixen com

$$m_n = \int_{\mathbb{R}} x^n \mu(dx), \quad \sigma^2 = \int_{\mathbb{R}} (x - m_1)^2 \mu(dx).$$

El moment d'ordre 1,  $m_1 = \int_{\mathbb{R}} x \mu(dx)$  s'anomena també *valor mig*. Representa el "centre de masses" de  $\mu$ , si interpretem la probabilitat  $\mu$  com una distribució de masses en la recta. L'esperança d'una variable aleatòria serà aleshores el *valor mig* de la seva llei.

La variància  $\sigma^2$  ens mesura el grau de dispersió de la mesura  $\mu$  respecte al valor mig  $m_1$ . Per exemple,  $\sigma^2 = 0$  si i només si  $\mu = \delta_{m_1}$ . En el cas d'una variable aleatòria  $X$  ( $\mu = P \circ X^{-1}$ ) que la variància de  $X$  sigui 0, equival a que  $X$  sigui igual a  $E(X)$  amb probabilitat 1.

Sigui  $X, Y$  variables aleatòries en un espai de probabilitat  $(\Omega, \mathcal{A}, P)$ . Recordarem tot seguit les següents desigualtats, que es demostrarien con en el cas de la integral de Lebesgue:

(i) *Desigualtat de Minkowski:* Per a tot  $p \geq 1$

$$(E(|X+Y|^p))^{1/p} \leq (E(|X|^p))^{1/p} + (E(|Y|^p))^{1/p}.$$

(ii) *Desigualtat de Hölder:* Per a qualsevol  $p, q > 1$  amb  $\frac{1}{p} + \frac{1}{q} = 1$ ,

$$E(|XY|) \leq (E(|X|^p))^{1/p} (E(|Y|^q))^{1/q}.$$

Com a cas particular de (ii) quan  $p = q = 2$  s'obté la desigualtat de Schwarz:

$$E(|XY|) \leq \{E(|X|^2) E(|Y|^2)\}^{1/2}.$$

Suposem  $1 \leq r < s$ . Aleshores  $E(|X|^s) < \infty \Rightarrow E(|X|^r) < \infty$ . Aquest resultat és una conseqüència de la desigualtat de Hölder.

En efecte, si prenem  $p = \frac{s}{r}$ ,  $q = \frac{r}{s-r}$  i  $Y = 1$  tindrem

$$E(|X|^r) \leq (E(|X|^p))^{\frac{1}{p}} = (E(|X|^s))^{\frac{r}{s}}.$$

Designarem per  $\mathcal{L} \subset \mathcal{L}(\Omega, \mathcal{A}, P)$  l'espai vectorial de totes les variables aleatòries definides en l'espai de probabilitat  $(\Omega, \mathcal{A}, P)$ .

Per a cada  $p \geq 1$ ,  $\mathcal{L}^p = \{X \in \mathcal{L} : E(|X|^p) < \infty\}$  és un subespai vectorial de  $\mathcal{L}$ , degut a la desigualtat de Minkowski i aquests subespais són cada vegada més petits si augmentem el valor de  $p$ , com acabem de veure.

La relació  $X \sim Y \iff P\{X \neq Y\} = 0$  és una relació d'equivalència en el conjunt  $\mathcal{L}$ , compatible amb l'estructura d'espai vectorial. Podem introduir aleshores l'espai quotient  $\mathcal{L}^0 = \mathcal{L}/\sim$ . Sovint identificarem una variable aleatòria amb la seva classe d'equivalència. Analogament s'introdueixen els espais  $L^p(\Omega, \mathcal{A}, P) = \mathcal{L}^p(\Omega, \mathcal{A}, P)/\sim$ , per a cada nombre real  $p \geq 1$ .

Els casos  $p = 1$  i  $p = 2$  són especialment interessants.

$L^1(\Omega, \mathcal{A}, P)$  és l'espai vectorial dotat d'un producte escalar definit per la forma bilineal  $\langle X, Y \rangle \mapsto E(XY)$ .

#### 6.4.1. Imatges de lleis amb densitat

Sigui  $X$  una variable aleatòria en un espai de probabilitat  $(\Omega, \mathcal{A}, P)$ , que pren valors en un interval obert  $I$  finit o no. Sigui  $f : I \rightarrow J$  una transformació bijectiva i de classe  $C^1$  de l'interval  $I$  en un altre interval obert  $J$ .

Si la llei de la variable  $X$  té una densitat  $\varphi_X(x)$  respecte la mesura de Lebesgue, ens podem preguntar quina és la densitat de la variable  $Y = f(X)$ . La resposta és la següent:

$$\varphi_Y(y) = \varphi_X(f^{-1}(y)) \frac{1}{|f'(f^{-1}(y))|}, \quad y \in J. \quad (6.6)$$

En efecte, si suposem per exemple que  $f$  és creixent i que  $I = (a, b)$ ,  $J = (c, d)$ ,  $-\infty \leq a < b \leq +\infty$ ,  $-\infty \leq c < d \leq +\infty$ , tindrem

$$\begin{aligned} F_Y(y) &= P\{Y \leq y\} = P\{X \leq f^{-1}(y)\} = \int_a^{f^{-1}(y)} \varphi_X(x) dx \\ &= \int_c^y \varphi_X(f^{-1}(z)) \frac{dz}{f'(f^{-1}(z))}. \end{aligned}$$

Per  $f$  decreixent la demostració és anàloga.

Si el domini de  $f$  es pot escriure com una reunió finita d'intervalss  $I_1 \cup \dots \cup I_n$ , i la restricció de  $f$  a cada  $I_j$ ,  $j = 1, \dots, n$  és bijectiva i de classe  $C^1$ , la densitat de la variable  $Y = f(X)$  és

$$\begin{aligned} \varphi_Y(y) &= \sum_{i=1}^n \varphi_X(f_i^{-1}(y)) \frac{1}{|(f'_i(f_i^{-1}(y)))|}, \\ \text{on} \quad f_i(x) &= \begin{cases} f(x), & \text{si } x \in I_i, \\ 0, & \text{si } x \notin I_i. \end{cases} \end{aligned}$$

**Exemple:** Sigui  $X$  una variable aleatòria amb distribució uniforme en un interval  $[a, b]$ , on  $a \geq 0$ . La variable  $X$  pot representar l'elecció a l'atzar d'un nombre de l'interval  $[a, b]$ . Suposem que aquest nombre l'elevem al quadrat i volem determinar la llei de probabilitat del nombre així obtingut. Degut a la fórmula (6.6), la variable  $X^2$  tindrà una densitat en l'interval  $[a^2, b^2]$  donada per

$$\varphi_{X^2}(y) = \frac{1}{b-a} \cdot \frac{1}{2\sqrt{y}}, \quad y \in [a^2, b^2].$$

Així, per exemple,  $E(X^2)$  pot ésser calculada indistintament a partir d'una de les expressions següents

$$\int_a^b x^2 \frac{1}{b-a} dx, \quad \text{ó} \quad \int_a^{b^2} y \frac{1}{b-a} \frac{1}{2\sqrt{y}} dy.$$

L'expressió (6.6) es pot estendre a vectors aleatoris. Considerem un vector aleatori  $X : \Omega \rightarrow \mathbb{R}^n$  en un espai de probabilitat  $(\Omega, \mathcal{A}, P)$ . Per definició la llei del vector  $X$  és la probabilitat en  $\mathbb{R}^n$  donada per  $P \circ X^{-1}$ .

Suposem que  $X$  pren valors en un obert  $U$  de  $\mathbb{R}^n$  i que la llei d' $X$  té una densitat  $\varphi_X(x)$  (que serà nulla si  $x \in U^c$ ) respecte la mesura de Lebesgue. Signi  $f$  una transformació bijectiva i de classe  $C^1$  de l'obert  $U$  en un altre obert  $V$ . Designarem per  $J_f(x) = \det \begin{bmatrix} \frac{\partial f_i}{\partial x_j} \end{bmatrix}$  el Jacobiat de l'aplicació  $f$ . Aleshores, el vector aleatori  $Y = f(X)$  té una densitat igual a

$$\varphi_Y(y) = \varphi_X(f^{-1}(y)) \frac{1}{|J_f(f^{-1}(y))|}, \quad y \in V. \quad (6.7)$$

En efecte, tenint en compte la fórmula del canvi de variable, per a tot obert  $A \subset V$  tindrem

$$P\{Y \in A\} = P\{X \in f^{-1}(A)\} = \int_{f^{-1}(A)} \varphi_X(x) dx = \int_A \varphi_X(f^{-1}(z)) \frac{dz}{|J_f(f^{-1}(z))|}.$$

En el llenguatge de l'apartat 6.1,  $\mathcal{A}$  és la  $\sigma$ -àlgebra inicial corresponent a les funcions  $\pi_i, i \in I$ . Aleshores, tal com hem vist a l'apartat 6.1 es compleix el següent resultat:

Una aplicació  $X : (E, \mathcal{E}) \rightarrow (\prod_{i \in I} \Omega_i, \otimes_{i \in I} \mathcal{A}_i)$  és measurable si i només si les funcions  $\pi_i \circ X : E \rightarrow \Omega_i, i \in I$ , són measurable.

En el cas d'un nombre finit d'espais  $(\Omega_i, \mathcal{A}_i), i = 1, 2, \dots, n$ , anomenem rectangle measurable a tot subconjunt d' $\Omega = \Omega_1 \times \dots \times \Omega_n$  de la forma  $A_1 \times \dots \times A_n$ , on  $A_i \in \mathcal{A}_i$  per a tot  $i = 1, \dots, n$ .

La família  $\mathcal{S}$  dels rectangles measurable és una semiàlgebra. En efecte,

- (i)  $\Omega_1 \times \dots \times \Omega_n \in \mathcal{S}$ .
- (ii) Si  $A_1 \times \dots \times A_n, B_1 \times \dots \times B_n$  son de  $\mathcal{S}$ , aleshores  $(A_1 \times \dots \times A_n) \cap (B_1 \times \dots \times B_n) = (A_1 \cap B_1) \times \dots \times (A_n \cap B_n) \in \mathcal{S}$ .
- (iii) Si  $A_1 \times \dots \times A_n$  és un conjunt de  $\mathcal{S}$ , el conjunt  $(A_1 \times \dots \times A_n)^c$  es pot escriure com una unió finita de conjunts de  $\mathcal{S}$  disjunts dos a dos.

La  $\sigma$ -àlgebra producte  $\mathcal{A}_1 \otimes \dots \otimes \mathcal{A}_n$  concideix amb la  $\sigma$ -àlgebra generada pels rectangles measurable, és a dir

$$\sigma\{\pi_i^{-1}(A_i); A_i \in \mathcal{A}_i, i = 1, \dots, n\} = \sigma\{A_1 \times \dots \times A_n; A_i \in \mathcal{A}_i, i = 1, \dots, n\}.$$

En efecte, tot cilindre és un rectangle measurable, car  $\pi_i^{-1}(A_i) = \Omega_1 \times \dots \times \Omega_{i-1} \times A_i \times \Omega_{i+1} \times \dots \times \Omega_n$ ; tot rectangle measurable és una intersecció finita de cilindres:  $A_1 \times \dots \times A_n = \bigcap_{i=1}^n \pi_i^{-1}(A_i)$ .

## 7. PRODUCTE D'ESPAIS DE PROBABILITAT. INDEPENDÈNCIA. DISTRIBUCIONS CONDICIONADES

### 7.1. Producte d'espais measurable

Considerem una família d'espais measurable  $\{(\Omega_i, \mathcal{A}_i), i \in I\}$ . Per definició, l'espai producte  $(\Omega, \mathcal{A})$  s'obté prenent  $\Omega = \prod_{i \in I} \Omega_i$  i com a  $\sigma$ -àlgebra producte  $\mathcal{A}$  (que designarem per  $\otimes_{i \in I} \mathcal{A}_i$ ) la mínima  $\sigma$ -àlgebra que fa measurable les aplicacions projecció  $\pi_i : \Omega \rightarrow \Omega_i, i \in I$ . És a dir,  $\mathcal{A}$  és la  $\sigma$ -àlgebra generada per la família de conjunts  $\{\pi_i^{-1}(A_i); A_i \in \mathcal{A}_i, i \in I\}$ . Un conjunt de la forma  $\pi_i^{-1}(A_i)$  s'anomena cilindre de base  $A_i$ .

En el llenguatge de l'apartat 6.1,  $\mathcal{A}$  és la  $\sigma$ -àlgebra inicial corresponent a les funcions  $\pi_i, i \in I$ . Aleshores, tal com hem vist a l'apartat 6.1 es compleix el següent resultat:

Una aplicació  $X : (E, \mathcal{E}) \rightarrow (\prod_{i \in I} \Omega_i, \otimes_{i \in I} \mathcal{A}_i)$  és measurable si i només si les funcions  $\pi_i \circ X : E \rightarrow \Omega_i, i \in I$ , són measurable.

En el cas d'un nombre finit d'espais  $(\Omega_i, \mathcal{A}_i), i = 1, 2, \dots, n$ , anomenem rectangle measurable a tot subconjunt d' $\Omega = \Omega_1 \times \dots \times \Omega_n$  de la forma  $A_1 \times \dots \times A_n$ , on  $A_i \in \mathcal{A}_i$  per a tot  $i = 1, \dots, n$ .

La família  $\mathcal{S}$  dels rectangles measurable és una semiàlgebra. En efecte,

- (i)  $\Omega_1 \times \dots \times \Omega_n \in \mathcal{S}$ .
- (ii) Si  $A_1 \times \dots \times A_n, B_1 \times \dots \times B_n$  son de  $\mathcal{S}$ , aleshores  $(A_1 \times \dots \times A_n) \cap (B_1 \times \dots \times B_n) = (A_1 \cap B_1) \times \dots \times (A_n \cap B_n) \in \mathcal{S}$ .
- (iii) Si  $A_1 \times \dots \times A_n$  és un conjunt de  $\mathcal{S}$ , el conjunt  $(A_1 \times \dots \times A_n)^c$  es pot escriure com una unió finita de conjunts de  $\mathcal{S}$  disjunts dos a dos.

La  $\sigma$ -àlgebra producte  $\mathcal{A}_1 \otimes \dots \otimes \mathcal{A}_n$  concideix amb la  $\sigma$ -àlgebra generada pels rectangles measurable, és a dir

$$\sigma\{\pi_i^{-1}(A_i); A_i \in \mathcal{A}_i, i = 1, \dots, n\} = \sigma\{A_1 \times \dots \times A_n; A_i \in \mathcal{A}_i, i = 1, \dots, n\}.$$

Considerem dos espais mesurables  $(\Omega_1, \mathcal{A}_1)$ ,  $(\Omega_2, \mathcal{A}_2)$  i el seu producte  $(\Omega_1 \times \Omega_2, \mathcal{A}_1 \otimes \mathcal{A}_2)$ . Si fixem un element  $\omega_1 \in \Omega_1$ , l'aplicació  $I_{\omega_1} : \Omega_2 \rightarrow \Omega_1 \times \Omega_2$  definida per  $I_{\omega_1}(\omega_2) = (\omega_1, \omega_2)$  és mesurable ja que per a tot rectangle measurable  $A_1 \times A_2$  es compleix

$$I_{\omega_1}^{-1}(A_1 \times A_2) = \begin{cases} A_2 \in \mathcal{A}_2, & \text{si } \omega_1 \in A_1 \\ \emptyset \in \mathcal{A}_2, & \text{si } \omega_1 \notin A_1. \end{cases}$$

Considerem un altre espai measurable  $(E, \mathcal{E})$  i una funció measurable  $X : \Omega_1 \times \Omega_2 \rightarrow E$ . Aleshores, per a cada  $\omega_1 \in \Omega_1$ , la secció  $X(\omega_1, \cdot) : \Omega_2 \rightarrow E$ , definida per  $X(\omega_1, \cdot)(\omega_2) = X(\omega_1, \omega_2)$ , és measurable car  $X(\omega_1, \cdot) = X(\omega_1, \cdot) = X(\omega_1, \omega_2)$ . Analogament, les aplicacions secció del tipus  $X(\cdot, \omega_2)$  són mesurables.

## 7.2. Construcció de probabilitats en un producte mitjançant probabilitats de transició

En primer lloc donarem la següent definició.

**Definició 7.1.** Una probabilitat de transició d'un espai measurable  $(\Omega_1, \mathcal{A}_1)$  en un altre  $(\Omega_2, \mathcal{A}_2)$  és una aplicació  $p : \Omega_1 \times \mathcal{A}_2 \rightarrow [0, 1]$  tal que

(i) Per a tot  $\omega_1 \in \Omega_1$ ,  $p(\omega_1, \cdot)$  és una probabilitat en l'espai  $(\Omega_2, \mathcal{A}_2)$ .

(ii) Per a tot  $A_2 \in \mathcal{A}_2$ , la funció  $p(\cdot, A_2) : \Omega_1 \rightarrow \mathbb{R}$  és measurable.

Si  $\Omega_1, \Omega_2$  representen els conjunts de resultats associats a dues experiències aleatòries, la funció  $p$  representa la probabilitat que es produexi l'esdeveniment  $A_2$  en la segona experiència si hem observat el resultat  $\omega_1$  en la primera.

Quan  $p(\omega_1, A_2) = p(A_2)$ , o sigui no depèn de  $\omega_1$ ,  $p$  és simplement una probabilitat en l'espai measurable  $(\Omega_2, \mathcal{A}_2)$ . Aquest cas corresindrà al cas de dues experiències independents.

**Teorema 7.2.** Considerem un espai de probabilitat  $(\Omega_1, \mathcal{A}_1, P_1)$ , un espai measurable  $(\Omega_2, \mathcal{A}_2)$  i una probabilitat de transició  $p$  de  $(\Omega_1, \mathcal{A}_1)$  a  $(\Omega_2, \mathcal{A}_2)$ . Existeix una única probabilitat  $Q$  en l'espai producte  $(\Omega, \mathcal{A}) = (\Omega_1 \times \Omega_2, \mathcal{A}_1 \otimes \mathcal{A}_2)$  tal que

$$Q(A_1 \times A_2) = \int_{A_1} p(\omega_1, A_2) P_1(d\omega_1), \quad \text{per a tots } A_1 \in \mathcal{A}_1, A_2 \in \mathcal{A}_2. \quad (7.1)$$

Si  $X : \Omega \rightarrow [0, +\infty]$  és una funció measurable, es compleix

$$\int_{\Omega} X dQ = \int_{\Omega_1} \left[ \int_{\Omega_2} X(\omega_1, \omega_2) p(\omega_1, d\omega_2) \right] P_1(d\omega_1). \quad (7.2)$$

Finalment, si  $X : \Omega \rightarrow \mathbb{R}$  és integrable respecte a  $Q$  val també la fórmula (7.2) en el sentit de que la integral  $\int_{\Omega_1} X(\omega_1, \omega_2) p(\omega_1, d\omega_2)$  existeix per a tot  $\omega_1$  a més d'un conjunt de probabilitat  $P_1$  zero i defineix una funció de  $\omega_1$  integrable respecte a  $P_1$ .

**Demàstració:** Degut als resultats sobre extensió de mesures (Teorema 4.13), l'existència i la unitat de  $Q$  quedaran demostrades si veiem que  $Q$  (definida per (7.1)) és  $\sigma$ -additiva en la semiàlgebra  $\mathcal{S}$  dels rectangles mesurables.

Suposem que  $A_1 \times A_2 = \bigcup_{i=1}^{\infty} [A_1^i \times A_2^i]$ , on els  $A_1^i \times A_2^i$  són disjunts dos a dos. Llavors podem escriure

$$\begin{aligned} \sum_{i=1}^{\infty} Q(A_1^i \times A_2^i) &= \sum_{i=1}^{\infty} \int_{A_1^i} p(\omega_1, A_2^i) P_1(d\omega_1) = \\ &= \int_{\Omega_1} \left[ \sum_{i=1}^{\infty} 1_{A_1^i}(\omega_1) p(\omega_1, A_2^i) \right] P_1(d\omega_1) = \\ &= \int_{\Omega_1} \left( \sum_{i=1}^{\infty} \int_{\Omega_2} 1_{A_1^i}(\omega_1) 1_{A_2^i}(\omega_2) p(\omega_1, d\omega_2) \right) P_1(d\omega_1) = \\ &= \int_{\Omega_1} \left( \int_{\Omega_2} \left( \sum_{i=1}^{\infty} 1_{A_1^i}(\omega_1) 1_{A_2^i}(\omega_2) \right) p(\omega_1, d\omega_2) \right) P_1(d\omega_1) = \\ &= \int_{\Omega_1} \left( \int_{\Omega_2} 1_{A_1 \times A_2}(\omega_1, \omega_2) p(\omega_1, d\omega_2) \right) P_1(d\omega_1) = \\ &= \int_{A_1} p(\omega_1, A_2) P_1(d\omega_1) = Q(A_1 \times A_2). \end{aligned}$$

Considererem la família  $\mathcal{C}$  formada pels conjunts  $A \in \mathcal{A}_1 \otimes \mathcal{A}_2$  tals que l'aplicació  $\omega_1 \mapsto \int_{\Omega_2} 1_A(\omega_1, \omega_2) p(\omega_1, d\omega_2)$  és measurable i  $\int_{\Omega_1} P_1(d\omega_1) \left[ \int_{\Omega_2} 1_A(\omega_1, \omega_2) p(\omega_1, d\omega_2) \right] = Q(A)$ . Volem veure que  $\mathcal{C} = \mathcal{A}_1 \otimes \mathcal{A}_2$ . Això demostrarà la segona part del teorema quan  $X$  és un indicador. La igualtat  $C = \mathcal{A}_1 \otimes \mathcal{A}_2$  surt de les consideracions següents:

(i)  $\mathcal{C}$  conté els rectangles measurable. En efecte

$$\int_{\Omega_1} 1_{A_1 \times A_2}(\omega_1, \omega_2) p(\omega_1, d\omega_2) = 1_{A_1}(\omega_1) p(\omega_1, d\omega_2),$$

$$Q(A_1 \times A_2) = \int_{\Omega_1} P_1(d\omega_1) \left[ \int_{\Omega_2} 1_{A_1}(\omega_1) p(\omega_1, d\omega_2) \right].$$

(ii)  $\mathcal{C}$  és estable per unions disjunes, i per límits creixents i decreixents.

La propietat (i) ens diu que  $\mathcal{C} \supseteq \mathcal{S}$  i la propietat (ii) implica que  $\mathcal{C}$  conté l'àlgebra generada per  $\mathcal{S}$  i és classe monòtona. Pel teorema de les classes monòtones (Teorema 4.4) obtenim que  $\mathcal{C} \supseteq \sigma(\mathcal{S}) = \mathcal{A}_1 \otimes \mathcal{A}_2$ .

Per tant, sabem que la igualtat (7.2) és certa quan la variable  $X$  és de la forma  $1_A$ ,  $A \in \mathcal{A}_1 \otimes \mathcal{A}_2$ . Llavors, per linearitat la igualtat serà certa quan  $X$  és una variable simple no negativa, i per convergència monòtona s'obté el resultat per a tota funció mesurable  $X : \Omega \rightarrow [0, +\infty]$ .

Suposem finalment que  $X : \Omega \rightarrow \mathbb{R}$  és integrable respecte  $Q$ .

Sabem que

$$\int_{\Omega_1} \left[ \int_{\Omega_2} |X(\omega_1, \omega_2)| p(\omega_1, d\omega_2) \right] P_1(d\omega_1) = \int_{\Omega_2} |X| dQ < \infty.$$

En conseqüència, existeix  $N_1 \in \mathcal{A}_1$  amb  $P_1(N_1) = 0$  tal que

$$\int_{\Omega_2} |X(\omega_1, \omega_2)| p(\omega_1, d\omega_2) < \infty, \quad \text{si } \omega_1 \notin N_1.$$

Definim

$$Y(\omega_1) = \begin{cases} 0 & \text{si } \omega_1 \in N_1, \\ \int_{\Omega_2} X(\omega_1, \omega_2) p(\omega_1, d\omega_2) & \text{si } \omega_1 \notin N_1. \end{cases}$$

$Y$  és una funció mesurable i integrable respecte  $P_1$ , ja que

$$\int_{\Omega_1} |Y(\omega_1)| P_1(d\omega_1) \leq \int_{\Omega_1} \left[ \int_{\Omega_2} |X(\omega_1, \omega_2)| p(\omega_1, d\omega_2) \right] P_1(d\omega_1) < \infty.$$

Finalment,

$$\begin{aligned} \int_{\Omega_1} Y(\omega_1) P_1(d\omega_1) &= \int_{\Omega_1} \left[ \int_{\Omega_2} X^+(\omega_1, \omega_2) p(\omega_1, d\omega_2) \right] P_1(d\omega_1) \\ &\quad - \int_{\Omega_1} \left[ \int_{\Omega_2} X^-(\omega_1, \omega_2) p(\omega_1, d\omega_2) \right] P_1(d\omega_1) \\ &= \int_{\Omega_1} X^+ dQ - \int_{\Omega_1} X^- dQ \\ &= \int_{\Omega_1} X dQ. \quad \blacksquare \end{aligned}$$

Observacions

- Suposem que  $\Omega_1$  i  $\Omega_2$  representen el conjunt de resultats de dues experiències aleatories. Si conceixem la llei de probabilitat  $P_1$  de la primera experiència i la llei

de probabilitat  $p(\omega_1, A_2)$  de la segona condicionada per la primera, el Teorema 7.2 ens permetrà de calcular la llei de probabilitat conjunta de les dues experiències. Utilitzarem aquesta idea en l'estudi de les distribucions de probabilitat condicionades.

- Considerem dos espais de probabilitat  $(\Omega_1, \mathcal{A}_1, P_1)$ ,  $(\Omega_2, \mathcal{A}_2, P_2)$ . Podem considerar  $P_2(A_2)$  com una probabilitat de transició que no depèn de  $\omega_1$ . En aquest cas, la probabilitat  $Q$  construïda mitjançant el Teorema 7.2 s'anomena el *producte* de  $P_1$  per  $P_2$ , es designa per  $P_1 \times P_2$  i compleix  $Q(A_1 \times A_2) = P_1(A_1)P_2(A_2)$ .  $A_1 \in \mathcal{A}_1$ ,  $A_2 \in \mathcal{A}_2$ .

La formula (7.2) ens diu que les integrals respecte la probabilitat producte  $P_1 \times P_2$  es poden calcular fent primer la integració respecte una de les probabilitats i després respecte l'altra. Es tracta d'una versió del teorema de Fubini per probabilitats. De forma semblant es pot construir el producte d'un nombre finit d'espais de probabilitat  $(\Omega_i, \mathcal{A}_i, P_i)$ ,  $i = 1, \dots, n$ , que representarem per  $(\Omega_1 \times \dots \times \Omega_n, \mathcal{A}_1 \otimes \dots \otimes \mathcal{A}_n, P_1 \times \dots \times P_n)$ .

### 7.3. Independència

Signifiqui  $(\Omega, \mathcal{A}, P)$  un espai de probabilitat. Recordem que en el Capítol 3 hem definit la independència de dos esdeveniments  $A, B$  per la condició  $P(A \cap B) = P(A)P(B)$ . Més generalment, la Definició 3.4 ens diu que una col·lecció d'esdeveniments  $\{A_i, i \in I\}$  és independent si per a tot subconjunt finit  $J \subset I$  es compleix  $P(\bigcap_{j \in J} A_j) = \prod_{j \in J} P(A_j)$ . En aquest apartat estendrem la noció d'independència a col·leccions de  $\sigma$ -àlgebres i de variables aleatories, i estableirem diverses caracteritzacions de la independència.

Definició 7.3. Signifiqui  $\{\mathcal{C}_i, i \in I\}$ ,  $\mathcal{C}_i \subset \mathcal{A}$ , una col·lecció de conjunts d'esdeveniments. Direm que aquesta col·lecció és independent si tota família de conjunts  $\{A_i, i \in I\}$ , amb  $A_i \in \mathcal{C}_i$  per a tot  $i \in I$ , és independent.

Definició 7.4. Signifiqui  $\{X_i, i \in I\}$  una família de variables aleatories. Direm que la família ( $\mathbf{x}$  les variables) és independent si la col·lecció de  $\sigma$ -àlgebres  $\{X_i^{-1}(\mathcal{B}(\mathbb{R}))\}_{i \in I}$  és independent.

Tenint en compte les definicions anteriors, la independència de la família  $\{X_i, i \in I\}$  és equivalent al fet que per a tot subconjunt finit  $\{i_1, \dots, i_n\} \subset I$ , i per a tots  $B_1, \dots, B_n \in \mathcal{B}(\mathbb{R})$  es compleix

$$P\{X_{i_1} \in B_1, \dots, X_{i_n} \in B_n\} = P\{X_{i_1} \in B_1\} \dots P\{X_{i_n} \in B_n\}.$$

En general, direm que una família  $\{X_i, i \in I\}$ ,  $X_i : (\Omega, \mathcal{A}) \rightarrow (E, \mathcal{E}_i)$  d'aplicacions mesurables és independent si ho és la col·lecció de  $\sigma$ -àlgebres  $\{X_i^{-1}(\mathcal{E}_i), i \in I\}$ .

**Propietats de les variables aleatòries independents**

- (i) Si  $\{X_i, i \in I\}$  són variables aleatòries independents i  $g_i : \mathbb{R} \rightarrow \mathbb{R}$  són funcions mesurables, aleshores les variables  $\{g_i(X_i), i \in I\}$  són també independents.

*Demostració:* Per a cada  $i \in I$  tenim

$$(g_i(X_i))^{-1}(\mathcal{B}(\mathbb{R})) = X_i^{-1}[g_i^{-1}(\mathcal{B}(\mathbb{R}))] \subset (X_i)^{-1}(\mathcal{B}(\mathbb{R})).$$

- (ii)  $X_1, \dots, X_n$  són v.a. independents si i només si la llei del vector aleatori  $X = (X_1, \dots, X_n)$  és el producte de les lleis marginals, és a dir,

$$P \circ (X_1, \dots, X_n)^{-1} = (P \circ X_1^{-1}) \times \dots \times (P \circ X_n^{-1}). \quad (7.3)$$

*Demostració:* Suposem primer que les variables són independents.

Per a tot  $B_1, \dots, B_n \in \mathcal{B}(\mathbb{R})$  tindrem

$$\begin{aligned} (P \circ X^{-1})(B_1 \times \dots \times B_n) &= P\{X_1 \in B_1, \dots, X_n \in B_n\} \\ &= P\{X_1 \in B_1\} \dots P\{X_n \in B_n\} = ((P \circ X_n^{-1}) \times \dots \times (P \circ X_1^{-1}))(B_1 \times \dots \times B_n), \end{aligned}$$

la qual cosa implica (7.3) ja que dues probabilitats en  $\mathbb{R}^n$  quedan determinades pels seus valors en els rectangles mesurables.

Recíprocament, si (7.3) és cert, per a tot conjunt finit  $\{i_1, \dots, i_m\} \subset \{1, \dots, n\}$  i per a tots  $B_1, \dots, B_m \in \mathcal{B}(\mathbb{R})$  tindrem

$$\begin{aligned} P\{X_{i_1} \in B_1, \dots, X_{i_m} \in B_m\} &= P\{X_{i_1} \in B_1, \dots, X_{i_m} \in B_m, X_j \in \Omega, \\ j \in [1, 2, \dots, n] - \{i_1, \dots, i_m\}\} &= P\{X_{i_1} \in B_1\} \dots P\{X_{i_m} \in B_m\}. \end{aligned}$$

Considerem  $n$  variables aleatòries  $X_1, \dots, X_n$ . Per definició, la seva funció de distribució conjunta és la funció de distribució  $n$ -dimensional de la llei de vector aleatori  $X = (X_1, \dots, X_n)$ , és a dir,

$$F_X(x) = P(X^{-1}((-\infty, x])) = P\{X_1 \leq x_1, \dots, X_n \leq x_n\},$$

si  $x = (x_1, \dots, x_n) \in \mathbb{R}^n$ .

Tenim aleshores el següent resultat.

- (iii) Les variables aleatòries  $X_1, \dots, X_n$  són independents si i només si la funció de distribució conjunta és el producte de les funcions de distribució marginals, o sigui,

$$F_X(x_1, \dots, x_n) = F_{X_1}(x_1) \dots F_{X_n}(x_n). \quad (7.4)$$

*Demostració:* Sabem que la independència de les variables  $X_1, \dots, X_n$  és equivalent a la igualtat (7.3). Llavors només cal tenir en compte que els dos membres de (7.3) són probabilitats en  $\mathbb{R}^n$  que estan determinades per les seves respectives funcions de distribució. La funció de distribució de  $P \circ X^{-1}$  és, per definició,  $F_X$  i la del segon membre val

$$[(P \circ X_1^{-1}) \times \dots \times (P \circ X_n^{-1})][((-\infty, x_1] \times \dots \times (-\infty, x_n))] = F_{X_1}(x_1) \dots F_{X_n}(x_n).$$

Cal fer notar que, en general, les funcions de distribució marginals es poden obtenir a partir de la funció de distribució conjunta mitjançant un pas al límit:

$$F_{X_i}(X_i) = \lim_{\substack{x_i \rightarrow +\infty \\ j \neq i}} F_X(x_1, \dots, x_n).$$

Ara bé, les funcions de distribució marginals  $F_{X_1}, \dots, F_{X_n}$  no determinen la distribució conjunta excepte en situacions particulars, com és el cas de la independència de les variables  $X_1, \dots, X_n$ .

Considerem un vector aleatori  $X = (X_1, \dots, X_n)$  que té llei  $P \circ X^{-1}$  absolutament continua respecte la mesura de Lebesgue en  $\mathbb{R}^n$  amb densitat  $f_X(x_1, \dots, x_n)$ . Aleshores, cada variable  $X_i$  té una llei absolutament contínua amb densitat donada per

$$f_{X_i}(x_i) = \int_{\mathbb{R}^{n-i}} f_X(x_1, \dots, x_n) dx_1 \dots dx_{i-1} dx_{i+1} \dots dx_n.$$

En efecte, per a tot borelià  $B$  de  $\mathbb{R}$ , si posem  $\tilde{B} = \mathbb{R}^{i-1} \times B \times \mathbb{R}^{n-i}$ , tindrem pel teorema de Fubini

$$P\{X_i \in B\} = P\{X \in \tilde{B}\} = \int_{\tilde{B}} f_X(x) dx = \int_B f_{X_i}(x_i) dx_i.$$

El recíproc d'aquest resultat no és cert, ja que si  $X_1$  és una variable aleatòria amb llei absolutament contínua en  $\mathbb{R}$ , el vector aleatori  $(X_1, X_1)$  no pot tenir densitat perquè la seva distribució està concentrada en la recta  $x_1 = x_2$  de  $\mathbb{R}^2$ .

En el cas absolutament contínuo tenim la següent caracterització de la independència.

(iv) Sigui  $X_1, \dots, X_n$  variables aleatòries amb lleis marginals absolutament contínues. Aleshores  $X_1, \dots, X_n$  són independents si i únicament si la llei conjunta és absolutament contínua, amb densitat

$$f_{X^n}(x_1, \dots, x_n) = f_{X_1}(x_1) \dots f_{X_n}(x_n). \quad (7.5)$$

*Demostració:* Degut a la propietat (ii) la independència de les variables  $X_1, \dots, X_n$  equival a que

$$P\{X_1 \in B_1, \dots, X_n \in B_n\} = \left( \int_{B_1} f_{X_1}(x_1) dx_1 \right) \dots \left( \int_{B_n} f_{X_n}(x_n) dx_n \right).$$

Llavors només cal observar que pel teorema de Fubini, el segon membre d'aquesta igualtat val

$$\int_{B_1 \times \dots \times B_n} f_{X_1}(x_1) \dots f_{X_n}(x_n) dx_1 \dots dx_n.$$

(v) Suposem que les variables aleatòries  $X_1, \dots, X_n$  prenen valors en un mateix conjunt finit o numerable  $S = \{a_i, i \in I\} \subset \mathbb{R}$ . En aquest cas tenim que  $X_1, \dots, X_n$  són independents si i únicament si per a tots  $a_{i_1}, \dots, a_{i_n} \in S$  es compleix que

$$P\{X_1 = a_{i_1}, \dots, X_n = a_{i_n}\} = P\{X_1 = a_{i_1}\} \dots P\{X_n = a_{i_n}\}. \quad (7.6)$$

(vi) Suposem que  $X_1, \dots, X_n$  són variables aleatòries independents i integrables. Aleshores, el producte  $X_1 \dots X_n$  és integrable i l'esperança del producte és igual al producte de les esperances dels factors, és a dir

$$E(X_1 \dots X_n) = E(X_1) \dots E(X_n). \quad (7.7)$$

*Demostració:* Suposarem primer que les variables són no negatives. Utilitzant la caracterització (7.3) de la independència i el teorema de Fubini, s'obté

$$\begin{aligned} E(X_1 \dots X_n) &= \int_{\mathbb{R}^n} x_1 \dots x_n [P \circ (X_1, \dots, X_n)^{-1}] (dx_1, \dots, dx_n) \\ &= \int_{\mathbb{R}^n} x_1 \dots x_n \left[ (P \circ X_1^{-1}) \times \dots \times (P \circ X_n)^{-1} \right] (dx_1, \dots, dx_n) \\ &= \left( \int_{\mathbb{R}} x_1 (P \circ X_1^{-1})(dx_1) \right) \dots \left( \int_{\mathbb{R}} x_n (P \circ X_n^{-1})(dx_n) \right) \\ &= E(X_1) \dots E(X_n). \end{aligned}$$

En el cas general, la integrabilitat del producte  $X_1 \dots X_n$  surt de la relació

$$E(|X_1 \dots X_n|) = E(|X_1|) \dots E(|X_n|) < \infty,$$

i de nou, el teorema de Fubini ens permet d'establir la igualtat (7.7).

**Definició 7.5.** Sigui  $X$  i  $Y$  dues variables aleatòries de quadrat integrable. Es defineix la seva covariància per

$$\text{cov}(X, Y) = E[(X - E(X))(Y - E(Y))] = E(XY) - E(X)E(Y).$$

Aleshores, de la propietat (vi) es dedueix que si  $X$  i  $Y$  són independents i de quadrat integrable, la seva covariància és nula. Si dues variables tenen covariància nula, es diu que estan *incorellacionades*. La propietat d'independència és més forta que el fet d'estar incorrelacionades. Per exemple, si  $\theta$  és una variable aleatòria amb distribució uniforme en l'interval  $[0, 2\pi]$ , les variables aleatòries  $X = \cos \theta$  i  $Y = \sin \theta$  estan incorrelacionades però no són independents ja que  $X^2 + Y^2 = 1$ .

Per a variables aleatòries de  $L^2(\Omega, \mathcal{A}, P)$ , la covariància  $\text{cov}(X, X)$  és precisament la variància d' $\Omega$ , que com sabem és igual a  $\sigma^2(X) = E(X^2) - (E(X))^2$ . La rel quadra de la variància,  $\sigma(X) = \sqrt{\sigma^2(X)}$  s'anomena la *desviació típica* d' $X$ .

**Proposició 7.6.** Sigui  $X_1, \dots, X_n$  variables aleatòries independents i de quadrat integrable. Aleshores

$$\sigma^2(X_1 + \dots + X_n) = \sigma^2(X_1) + \dots + \sigma^2(X_n).$$

*Demonstració:* Tenint en compte que la independència implica la incorrelació, podem escriure

$$\begin{aligned} \sigma^2(X_1 + \dots + X_n) &= E \left[ \left( \sum_{i=1}^n (X_i - E(X_i)) \right)^2 \right] = \\ &= E \sum_{i=1}^n (X_i - E(X_i))^2 + E \sum_{i \neq j} (X_i - E(X_i))(X_j - E(X_j)) = \\ &= \sum_{i=1}^n \sigma^2(X_i). \quad \blacksquare \end{aligned}$$

**Definició 7.7.** Considerem dues variables aleatòries  $X, Y$  de quadrat integrable i amb variancia no nula (o signifiquem, no són constants q.s.). Es defineix el *coeficient de correlació* entre  $X$  i  $Y$  com

$$\rho_{XY} = \frac{\text{cov}(X, Y)}{\sigma(X)\sigma(Y)}.$$

Si  $X$  i  $Y$  són variables centrades de  $L^2(\Omega, \mathcal{A}, P)$ . La covariància és precisament el producte escalar de  $X$  i  $Y$  en l'espai de Hilbert  $L^2(\Omega, \mathcal{A}, P)$  i el coeficient de correlació és el cosinus de l'angle que formen  $X$  i  $Y$ :

Com a conseqüència de la desigualtat de Schwarz tenim que  $|\rho_{XY}| \leq 1$ . D'altra banda sabem que si  $X$  i  $Y$  són independents aleshores  $\rho_{XY} = 0$ . També, com veurem tot seguit, la igualtat  $|\rho_{XY}| = 1$  equival al fet que hi hagi una relació *lineal* entre les variables,  $Y = aX + b$ . Per tant, el coeficient de correlació mesura el grau de dependència lineal entre les variables.

Signi  $X = (X_1, \dots, X_n)$  un vector aleatori tal que les seves components són de quadrat integrable. La matríg de variàncies-covariàncies de  $X$  es defineix com  $\Lambda = (\text{cov}(X_i, X_j))_{1 \leq i, j \leq n}$ . També es pot escriure com

$$\Lambda = E[(X - E(X))(X - E(X))^*],$$

amb el conveni de que els vectors s'escriuen com una matríg d'una columna.

La matríg  $\Lambda$  és simètrica i definida no negativa. En efecte, per a tot  $t \in \mathbb{R}^n$  hem té

$$t^* \Lambda t = E[t^*(X - E(X))(X - E(X))^* t] = E[(t^*(X - E(X)))^2] \geq 0.$$

Si les variables  $X_1, \dots, X_n$  són independents, llur matríg de variàncies-covariàncies és diagonal.

#### 7.4. Regressió lineal

Donades dues variables no constants  $X, Y$  de  $L^2(\Omega, \mathcal{A}, P)$ , volem trobar coeficients  $a$  i  $b$  tals que la quantitat  $E[(Y - aX - b)^2]$  sigui mínima, és a dir volem trobar una funció lineal de la variable  $X$  que approximi a la variable  $Y$  en el sentit de la norma de l'espai  $L^2(\Omega, \mathcal{A}, P)$ .

Signi  $\tilde{X} = X - E(X)$  i  $\tilde{Y} = Y - E(Y)$ , aleshores

$$E[(Y - aX - b)^2] = E[(\tilde{Y} - a\tilde{X})^2] + E[b^2 - 2bE(\tilde{Y}) + a^2 E(\tilde{X})^2],$$

Per tant haurem de prendre  $b = E(\tilde{Y}) - aE(\tilde{X})$ .

Finalment, l'expressió

$$E[(\tilde{Y} - a\tilde{X})^2] = \sigma^2(Y) + a^2 \sigma^2(X) - 2a\rho_{XY}\sigma(X)\sigma(Y)$$

és mínima quan  $a = \rho_{XY} \frac{\sigma(Y)}{\sigma(X)}$ .

El valor mínim de  $E[(Y - aX - b)^2]$  és igual a  $\sigma^2(Y)(1 - \rho_{XY}^2)$ , per tant, val zero si i només si  $|\rho_{XY}| = 1$ .

La recta  $y = ax + b$  s'anomena *recta de regressió* de la variable  $Y$  sobre la variable  $X$  i el procediment que hem seguit per a calcular-la és el *mètode dels mètodes quadrats*.

#### 7.5. Llei del 0-1 de Kolmogorov

Introduirem primer el següent lemma tècnic.

**Lema 7.8.** En un espai de probabilitat  $(\Omega, \mathcal{A}, P)$  considerem una semiàlgebra  $\mathcal{S} \subset \mathcal{A}$  i una col·lecció de conjunts qualsevol  $C \subset \mathcal{A}$ . Si  $S \in \mathcal{S}$  i  $C$  independents, aleshores  $\sigma(S) \cap C$  són també independents.

*DemOSTRACIÓ:* Signi  $B = \sigma(S)$  l'àlgebra generada per  $S$ . Veurem em primer lloc que  $B \cap C$  són independents. Tot conjunt  $A \in B$  és de la forma  $A = A_1 \cup \dots \cup A_n$  on els  $A_i$  són conjunts de la semiàlgebra  $S$  disjunts dos a dos. Aleshores

$$P[A \cap B] = P[(A_1 \cup \dots \cup A_n) \cap B] = \sum_{i=1}^n P(A_i \cap B) = \sum_{i=1}^n P(A_i)P(B) = P(A)P(B),$$

per a tot  $B \in C$ .  
En segon lloc cal provar que la  $\sigma$ -àlgebra  $\sigma(B) \cap C$  són independents. Fixat un conjunt  $B \in C$ , definim

$$\mathcal{D} = \left\{ A \in \mathcal{A} : P(A \cap B) = P(A)P(B) \right\}.$$

$\mathcal{D}$  és una classe monòtona ja que si  $A_n \in \mathcal{D}$ ,  $A_n \uparrow A$ , és té  $P(A \cap B) = \lim_n P(A_n \cap B) = \lim_n P(A_n)P(B) = P(A)P(B)$  i anàlogament per successions decreixents. D'altra banda,  $\mathcal{D}$  conté l'àlgebra  $B$  perquè  $B \in C$  són independents. Llavors, pel teorema de les classes monòtones,  $\mathcal{D} \supseteq \sigma(B)$  i, en conseqüència,  $\sigma(B) \cap C$  són independents. ■

D'aquest lemma es dedueix que si  $S_1$  i  $S_2$  són dues semiàlgabres independents, les àlgabres  $\sigma(S_1) \cap \sigma(S_2)$  són independents. Com a aplicació d'aquest lema es pot demostrar la següent propietat associativa de la independència.

**PropOSICIÓ 7.9.** Signi  $\{X_i, i \in I\}$  una família de variables aleatories independents i considerem una partició  $I = J \cup K$ . Aleshores les  $\sigma$ -àlgabres  $\sigma\{X_i, i \in J\}$ ,  $\sigma\{X_i, i \in K\}$  són independents.

*Demostració:* Les  $\sigma$ -àlgebres  $\sigma\{X_i, i \in J\}$ ,  $\sigma\{X_i, i \in K\}$  estan generades respectivament per les semiàlgebres

$$\mathcal{S}_J = \left\{ X_j^{-1}(B_1) \cap \dots \cap X_{j_n}^{-1}(B_n); \quad j_1, \dots, j_n \in J, \quad n \geq 1, \quad B_1, \dots, B_n \in \mathcal{B}(\mathbb{R}) \right\}$$

$$\mathcal{S}_K = \left\{ X_{k_1}^{-1}(C_1) \cap \dots \cap X_{k_m}^{-1}(C_m); \quad k_1, \dots, k_m \in K, \quad m \geq 1, \quad C_1, \dots, C_m \in \mathcal{B}(\mathbb{R}) \right\}.$$

Per tant, només cal veure que aquestes simiàlgebres són independents i això és una conseqüència immediata de la independència de les variables  $X_i, i \in I$ . En efecte, si  $B_1, \dots, B_n, C_1, \dots, C_m$  són boreians de  $\mathbb{R}$ ,

$$\begin{aligned} & P[X_{j_1}^{-1}(B_1) \cap \dots \cap X_{j_n}^{-1}(B_n) \cap X_{k_1}^{-1}(C_1) \cap \dots \cap X_{k_m}^{-1}(C_m)] \\ &= P(X_{j_1}^{-1}(B_1)) \dots P(X_{j_n}^{-1}(B_n)) P(X_{k_1}^{-1}(C_1)) \dots P(X_{k_m}^{-1}(C_m)) \\ &= P\left[X_{j_1}^{-1}(B_1) \cap \dots \cap X_{j_n}^{-1}(B_n)\right] \cdot P\left[X_{k_1}^{-1}(C_1) \cap \dots \cap X_{k_m}^{-1}(C_m)\right]. \blacksquare \end{aligned}$$

Considerem una successió de variables aleatories  $\{X_n, n \geq 1\}$  i definim les següents  $\sigma$ -àlgebres:

$$\mathcal{F}_n = \sigma\{X_k, k \geq n+1\}, \quad \mathcal{F} = \bigcap_{n=0}^{\infty} \mathcal{F}_n.$$

Els conjunts de la  $\sigma$ -àlgebra  $\mathcal{F}$  s'anomenen conjunts *asíntotics*.

**Proposició 7.10.** (Llei del 0-1 de Kolmogorov). Si les variables aleatories  $X_n$  són independents, aleshores tot esdeveniment asymptòtic té probabilitat zero ó u.

*Demostració:* Per a tot  $n \geq 1$  posem  $\mathcal{A}_n = \sigma\{X_1, \dots, X_n\}$  i  $\mathcal{A}_0 = \cup_{n=1}^{\infty} \mathcal{A}_n \cdot \mathcal{A}_0$  és una àlgebra, ja que és una unió creixent de  $\sigma$ -àlgebres. i genera la  $\sigma$ -àlgebra  $\sigma\{X_n, n \geq 1\}$ . Per la proposició 7.9, les  $\sigma$ -àlgebres  $\mathcal{A}_n \cdot \mathcal{F}_n$  són independents, per a tot  $n \geq 1$ .

Això implica que  $\mathcal{F}$  és independent de  $\mathcal{A}_n, \forall n \geq 1$ , és a dir,  $\mathcal{F}$  és independent de  $\mathcal{A}_0$ , i pel Lema 7.8 tenim que  $\mathcal{F}$  és independent de  $\sigma\{X_n, n \geq 1\}$ . Per tant la  $\sigma$ -àlgebra  $\mathcal{F}$  és independent d'ella mateixa, i en conseqüència, per a tot  $F \in \mathcal{F}$  tindrem  $P(F) = 0$  ó  $P(F) = 1$ . ■

Aplicant la llei de 0-1 resulta que tota variable aleòria  $Y$  que sigui  $\mathcal{F}$ -mesurable haurà de ser constant q.s. En efecte, la funció de distribució de la variable  $Y$  només pren els valors 0 i 1, això ens diu que la llei de  $Y$  és una massa puntual en algun valor  $c \in \mathbb{R}$  i aleshores  $Y = c$  q.s.

En particular, si  $\{X_n, n \geq 1\}$  és una successió de variables aleatories independents, les variables

$$\liminf X_n, \quad \limsup X_n, \quad \liminf \frac{X_1 + \dots + X_n}{n}, \quad \limsup \frac{X_1 + \dots + X_n}{n}$$

seran constants quasi segurament.

### 7.6. Distribucions de probabilitat condicionades

Ja hem introduït abans la probabilitat condicionada d'un esdeveniment  $A$  per un altre esdeveniment  $B$  quan  $P(B) > 0$ . Aquesta probabilitat condicionada es defineix per la fórmula  $P(A/B) = \frac{P(A \cap B)}{P(B)}$ .

Ara bé, en alguns casos ens interessarà parlar de la probabilitat condicionada per un esdeveniment de probabilitat nulla. Suposem, per exemple, que  $X$  és una variable aleatoria a valors en  $[0,1]$  amb una distribució contínua, i que si la variable  $X$  pren un valor  $x \in \mathbb{R}$ , aleshores tirem una moneda amb probabilitat de sortir cara igual a  $g(x)$  on  $g : \mathbb{R} \rightarrow [0,1]$  és una funció coneguda. En aquesta experiència aleatòria, ens donen la probabilitat de treure cara (igual a  $g(x)$ ) condicionada per l'esdeveniment  $\{X = x\}$ , que té probabilitat zero.

Això ens motiva, per a donar una definició de la distribució de probabilitat d'una variable aleatoria  $Y$  (en l'exemple anterior  $Y$  podrà ésser el resultat del llançament de la moneda) condicionada per la realització d'una variable  $X$ .

**Definició 7.11.** Donades dues variables  $X, Y$  s'anomena distribució de probabilitat d' $Y$  condicionada per  $X$  a tota probabilitat de transició  $p(x, C)$  de  $\mathbb{R}$  en  $\mathbb{R}$  tal que

$$P\{X \in B, Y \in C\} = \int_B p(x, C) P_X(dx), \quad (7.8)$$

per a tot  $B, C \in \mathcal{B}(\mathbb{R})$ .

**Observacions**

- 1.- Pel teorema de la mesura imatge, el segon membre de la igualtat (7.8) es pot escriure igualment com
- 2.- A vegades es fa servir també la següent notació

$$\int_{\{X \in B\}} p(X(\omega), C) P(d\omega).$$

$$p(x, C) = P\{Y \in C / X = x\}. \quad (7.9)$$

Cal observar que en el cas que  $P\{X = x\} > 0$  la notació anterior és consistent, és a dir,  $p(x, C)$  coincideix amb la probabilitat condicionada de l'esdeveniment  $\{Y \in C\}$  per l'esdeveniment  $\{X = x\}$ . En efecte, si prenem  $B = \{x\}$ , la relació (7.8) ens dóna

$$P\{X = x, Y \in C\} = \int_{\{x\}} p(y, C) P_X(dy) = p(x, C) P\{X = x\}.$$

3.- Com que el primer membre de (7.8) val  $[P \circ (X, Y)^{-1}](B \times C)$ , la definició anterior equival a dir que la llei  $P \circ (X, Y)^{-1} = P_{(X,Y)}^{-1}$  del vector aleatori  $(X, Y)$  s'obté a partir de la llei d' $X$  i de la probabilitat de transició  $p(x, C)$ .

Aleshores, tenint en compte el teorema 7.2, tindrem que per a tota funció mesurable  $\varphi : \mathbb{R}^2 \rightarrow \mathbb{R}$  tal que  $E(|\varphi(X, Y)|) < \infty$  (és a dir,  $\varphi$  és integrable respecte la llei del vector aleatori  $(X, Y)$ ) es compleix

$$E(\varphi(X, Y)) = \int_{\mathbb{R}^2} \varphi P_{(X,Y)} \left[ \int_{\mathbb{R}} \varphi(x, y) P(Y \in dy / X = x) \right]. \quad (7.10)$$

4.- La definició i els resultats precedents s'estenen clarament al cas de dos vectors aleatoris  $X = (X_1, \dots, X_n)$ ,  $Y = (Y_1, \dots, Y_n)$ .

Resumint, si coneixem la llei d'una variable  $X$  i la distribució d'una altra variable  $Y$  condicionada per  $X$ , podem calcular la llei del vector aleatori  $(X, Y)$ , i, en particular, la llei de la variable  $Y$ . En l'exemple que hem donat al començament, tindrem

$$p(x, C) = g(x)\delta_1(C) + (1 - g(x))\delta_0(C).$$

Llavors, si  $X$  té una densitat  $f(x)$  i  $Y$  representa el nombre de cares, obtindrem

$$P\{Y = 1\} = \int_{\mathbb{R}} p(x, \{1\}) P_X(dx) = \int_{\mathbb{R}} g(x)f(x)dx.$$

Es pot demostrar que existeixen sempre distribucions de probabilitat condicionada però aquest tipus de resultats són d'un nivell superior al d'aquest curs. Ton seguit demostrem la unicitat de les distribucions condicionades, llevat de conjunts de probabilitat zero.

**Proposició 7.12.** Siguin  $p$  i  $q$  dues distribucions de probabilitat condicionada de  $Y$  per  $X$ . Aleshores existeix un conjunt  $N \in \mathcal{B}(\mathbb{R})$  amb  $P_X(N) = 0$  tal que  $p(x, C) = q(x, C)$  per a tot  $x \notin N$  i per a tot  $C \in \mathcal{B}(\mathbb{R})$ .

**Demostració:** Fixem un conjunt  $C \in \mathcal{B}(\mathbb{R})$ . De la definició de distribució condicionada es dedueix que

$$P\{X \in B, Y \in C\} = \int_B p(x, C) P_X(dx) = \int_B q(x, C) P_X(dx).$$

per a tot  $B \in \mathcal{B}(\mathbb{R})$ .

Per tant, existeix  $N_C \in \mathcal{B}(\mathbb{R})$  tal que  $P_X(N_C) = 0$  i  $p(x, C) = q(x, C)$  per a tot  $x \notin N_C$ . Signi  $N = \bigcup_{r \in \mathbb{Q}} N_{(-\infty, r]}$ . Clarament  $P_X(N) = 0$ . A més, si  $x \notin N$  tindrem  $p(x, (-\infty, r]) = q(x, (-\infty, r])$  per a tot nombre racional  $r$ . Com que les funcions de distribució són contínues per la dreta, això ens diu que les probabilitats  $p(x, \cdot)$  i  $q(x, \cdot)$  tenen la mateixa funció de distribució i, en conseqüència, són iguals. ■

Existència i càlcul de distribucions condicionades en determinats casos particulars

1) Si les variables  $X$  i  $Y$  són independents, sabem que la llei del vector aleatori  $(X, Y)$  coincideix amb la probabilitat producte  $P_X \times P_Y$ . Per tant, podem prendre  $p(x, C) = P_Y(C)$  com a distribució condicionada de  $Y$  per  $X$ .

2) Suposem que la variable  $X$  pren un conjunt finit o numerable de valors  $S = \{a_i, i \in I\} \subset \mathbb{R}$ . Aleshores la distribució de  $Y$  condicionada per  $X$  ve donada per

$$p(x, C) = \begin{cases} \frac{P\{X=x, Y \in C\}}{P\{X=x\}}, & \text{si } x \in S \\ P_0(C), & \text{si } x \notin S. \end{cases}$$

on  $P_0$  és una probabilitat arbitrària en  $\mathbb{R}$ .

En efecte,  $p(x, C)$  és una probabilitat de transició que compleix

$$\begin{aligned} P\{X \in B, Y \in C\} &= \sum_{\{x: x \in B \cap S\}} P\{X = x, Y \in C\} = \sum_{\{x: x \in B \cap S\}} p(x, C) P\{X = x\} \\ &= \int_B p(x, C) P_X(dx). \end{aligned}$$

3) Suposem que el vector aleatori  $(X, Y)$  té una llei absolutament contínua amb densitat  $f(x, y)$ . Ja sabem que en aquest cas, les variables  $X$  i  $Y$  tenen lleis absolutament contínues amb densitats marginals  $f_X(x) = \int_{\mathbb{R}} f(x, y)dy$ ,  $f_Y(y) = \int_{\mathbb{R}} f(x, y)dx$ . Llavors, la distribució de  $Y$  condicionada per  $X$  ve donada per

$$p(x, C) = \begin{cases} \frac{\int_C f(x, y) dy}{f_X(x)}, & \text{si } f_X(x) \neq 0 \\ P_0(C), & \text{si } f_X(x) = 0. \end{cases}$$

En efecte,  $p(x, C)$  és una probabilitat de transició tal que

$$\begin{aligned} P\{X \in B, Y \in C\} &= \int_{B \times C} f(x, y) dx dy = \int_B f_X(x) \left( \int_C \frac{f(x, y)}{f_X(x)} dy \right) \mathbf{1}_{\{f_X(x) \neq 0\}} dx \\ &= \int_B f_X(x) p(x, C) dx, \end{aligned}$$

ja que  $f_X(x) = 0$  implica  $f(x, y) = 0$  quasi per a tot  $y$  respecte la mesura de Lebesgue. Cal observar que per a cada  $x$  amb  $f_X(x) \neq 0$ , la probabilitat  $p(x, \cdot)$  és absolutament contínua amb densitat  $\frac{f(x, y)}{f_X(x)}$ . Aquesta densitat, que es representa per  $f(y/x)$  s'anomena *densitat condicionada* de la variable  $Y$  per la variable  $X$ , i es calcula fent el quotient de la densitat conjunta per la densitat marginal de la variable  $X$ .

Sigui  $p(x, C)$  la distribució de probabilitat condicionada d'una variable  $Y$  per la variable  $X$ . Si  $f : \mathbb{R} \rightarrow \mathbb{R}$  és una funció measurable no negativa o integrable respecte  $p(x, \cdot)$  (fixat un valor  $x \in \mathbb{R}$ ), la integral

$$\int_{\mathbb{R}} f(y)p(x, dy)$$

s'anomena *esperança condicionada* de la variable  $f(Y)$  per la variable  $X$ , i es representa per  $E(f(Y)|X = x)$ . Si  $f$  és no negativa, o si  $E(|f(Y)|) < \infty$ , sabem que

$$\begin{aligned} E(f(Y)) &= \int_{\mathbb{R}} P_X(dx) \left[ \int_{\mathbb{R}} f(y)p(x, dy) \right] \\ &= \int_{\mathbb{R}} P_X(dx) E(f(Y)/X = x). \end{aligned}$$

Això ens diu que per a tot  $x \in \mathbb{R}$  ( $P_X$ -quasi segurament), l'esperança condicionada  $E(f(Y)/X = x)$  existeix, i la seva integral respecte la llei de la variable  $X$  ens dóna  $E(f(Y))$ .

### 7.7. Producte numerable d'espais de probabilitat

Considerem un espai measurable  $(\Omega, \mathcal{A})$ . Designarem per  $\Omega^{\mathbb{N}}$  el conjunt de totes les successions d'elements d' $\Omega$ . Per a cada natural  $n \geq 1$ ,  $\pi_n : \Omega^{\mathbb{N}} \rightarrow \Omega$  representarà la projecció en la coordenada  $n$ -èsima. Com en el cas d'un producte arbitrari d'espais measurable, la  $\sigma$ -àlgebra producte  $\mathcal{A}^{\mathbb{N}}$  serà la mínima  $\sigma$ -àlgebra que fa measurable les aplicacions  $\pi_n$ ,  $n \geq 1$ . És a dir,  $\mathcal{A}^{\mathbb{N}} = \sigma\{\pi_n^{-1}(A); A \in \mathcal{A}, n \geq 1\}$

Tenim aleshores el següent resultat.

**Proposició 7.13.** Considerem una successió de probabilitats  $\{P_n, n \geq 1\}$  en l'espai  $(\Omega, \mathcal{A})$ . Existeix una única probabilitat  $P$  en  $(\Omega^{\mathbb{N}}, \mathcal{A}^{\mathbb{N}})$  tal que

$$P \left[ \pi_1^{-1}(A_1) \cap \dots \cap \pi_n^{-1}(A_n) \right] = P_1(A_1) \dots P_n(A_n), \quad (7.11)$$

per a tot  $n \geq 1$ ,  $A_1, \dots, A_n \in \mathcal{A}$ .

**Demostració:** Per a cada natural  $n \geq 1$  podem considerar la projecció  $\pi^n : \Omega^{\mathbb{N}} \rightarrow \Omega^n$  definida per  $\pi^n(\omega) = (\omega_1, \dots, \omega_n)$  si  $\omega = (\omega_n)_{n=1}^{\infty}$ . Les aplicacions  $\pi^n$  són measurable ja que

$$(\pi^n)^{-1}(A_1 \times \dots \times A_n) = (\pi_1^{-1}(A_1) \cap \dots \cap \pi_n^{-1}(A_n)) \in \mathcal{A}^{\mathbb{N}}, \quad \text{si } A_1, \dots, A_n \in \mathcal{A}.$$

Designarem per  $\mathcal{A}^n$  la  $\sigma$ -àlgebra producte en  $\Omega^n$ , i escriurem  $C_n = (\pi^n)^{-1}(\mathcal{A}^n)$ . L'aplicació  $(\pi^n)^{-1} : \mathcal{A}^n \rightarrow \mathcal{A}^{\mathbb{N}}$  dóna lloc a un isomorfisme entre les  $\sigma$ -àlgebres  $\mathcal{A}^n$  i  $\mathcal{C}_n \cdot ((\pi^n)^{-1}$  és injectiva perquè  $\pi^n$  és exhaustiva).

Es compleix  $C_n \subset C_m$  si  $n \leq m$ . En efecte, sigui  $G = (\pi^n)^{-1}(A)$ ,  $A \in \mathcal{A}^n$ . Podem considerar la projecció  $\pi^{mn} : \Omega^m \rightarrow \Omega^n$  definida per  $\pi^{mn}(\omega_1, \dots, \omega_m) = (\omega_1, \dots, \omega_n)$  que és measurable i compleix  $\pi^n = \pi^{mn} \circ \pi^m$ ;

Llavors,  $G = (\pi^n)^{-1}(A) = (\pi^m)^{-1}[(\pi^{mn})^{-1}(A)] \in C_m$ , ja que  $(\pi^{mn})^{-1}(A) \in \mathcal{A}^m$ . La unió creixent  $\mathcal{C} = \cup_{n=1}^{\infty} C_n$  és una àlgebra que genera  $\mathcal{A}^{\mathbb{N}}$ .

En primer lloc definirem  $P$  sobre l'àlgebra  $\mathcal{C}$  de manera que signi *additiva* i compleixi la propietat (7.11). Si  $G \in C_n$ ,  $G = (\pi^n)^{-1}(A)$ ,  $A \in \mathcal{A}^n$ , definim

$$P(G) = P^n(A), \quad (7.12)$$

on  $P^n = P_1 \times \dots \times P_n$ .

Com que  $(\pi^n)^{-1}$  és un isomorfisme entre les  $\sigma$ -àlgebres  $C_n$  i  $\mathcal{A}^n$ , aquesta definició dóna lloc a una probabilitat en la  $\sigma$ -àlgebra  $C_n$ .

Per tal de que (7.12) defineixi  $P$  en l'àlgebra  $\mathcal{C}$ , hem de veure que la definició *no depèn de n*. Suposem  $n \leq m$ ,  $G \in C_m$ ,  $G = (\pi^m)^{-1}(B)$ ,  $B \in \mathcal{A}^m$ . Aleshores cal proveir que  $P^n(A) = P^m(B)$ . Tindrem

$$(\pi^m)^{-1}(B) = (\pi^n)^{-1}(A) = (\pi^m)^{-1}[(\pi^{mn})^{-1}(A)],$$

per tant,  $B = (\pi^{mn})^{-1}(A)$  i  $P^m(B) = [P^m \circ (\pi^{mn})^{-1}](A) = P^n(A)$ .

En efecte, les probabilitats  $P^n$  i  $P^m \circ (\pi^{mn})^{-1}$  definides en l'espai measurable  $(\Omega^n, \mathcal{A}^n)$  són iguals ja que coincideixen sobre els rectangles mesurables:

$$P^n(A_1 \times \dots \times A_n) = P_1(A_1) \dots P_n(A_n).$$

$$\begin{aligned} & [P^m \circ (\pi^{mn})^{-1}] (A_1 \times \dots \times A_n) = P^m(A_1 \times \dots \times A_n \times \Omega \times \dots \times \Omega) \\ & = P_1(A_1) \dots P_m(A_n). \end{aligned}$$

Per tant  $P$  està ben definida en  $C$ .

Vegem que és additiva. Sigui  $G_1, G_2$  conjunts de  $C$  tals que  $G_1 \cap G_2 = \emptyset$ . Alleshores  $G_1, G_2 \in C_m$  per algun  $m \geq 1$ . En conseqüència  $P(G_1 \cup G_2) = P(G_1) + P(G_2)$  perquè  $P$  és una probabilitat sobre  $C_m$ . Clarament aquest argument no es pot utilitzar per a demostrar que  $P$  és  $\sigma$ -additiva.

D'altra banda, per construcció  $P$  compleix la propietat (7.11):

$$\begin{aligned} P[\pi_1^{-1}(A_1) \cap \dots \cap \pi_n^{-1}(A_n)] &= P[(\pi^n)^{-1}(A_1 \times \dots \times A_n)] \\ &= P^n(A_1 \times \dots \times A_n) = P_1(A_1) \dots P_n(A_n). \end{aligned}$$

Tenint en compte els resultats sobre l'estensió de mesures, per tal d'estendre la funció  $P$  a la  $\sigma$ -àlgebra  $A^{\mathbb{N}}$  només cal veure que  $P$  és  $\sigma$ -additiva en l'àlgebra  $C$ . Per això n'hi ha prou en provar que si  $\{B_n, n \geq 1\}$  és successió de conjunts que decreix cap el  $\emptyset$ , i  $B_n \in C$  per a tot  $n \geq 1$ , és té  $P(B_n) \downarrow 0$ .

Demostarem aquesta propietat per reducció a l'absurd. Suposem que  $\lim_n P(B_n) > 0$ . Per a cada  $n \geq 1$ , el conjunt  $B_n$  pertany a una certa  $\sigma$ -àlgebra  $C_{k_n}$ . Podem suposar que  $k_n = n$ . En efecte, si  $k_n < n$  no hi ha cap problema i si  $k_n > n$  repetim tantes vegades com calgui el conjunt  $B_{n-1}$  i possem  $B_n$  en el lloc  $k_n$ . Suposem que  $B_n = (\pi^n)^{-1}(A_n)$ . Fixats dos índexos  $n \geq m$  escriurem

$$P(B_n) = P^n(A_n) = \int_{\Omega^m} P_1(d\omega_1) \dots P_m(d\omega_m) h_{n,m}(\omega_1, \dots, \omega_m)$$

on

$$h_{n,m}(\omega_1, \dots, \omega_m) = \int_{\Omega^{n-m}} 1_{A_n}(\omega_1, \dots, \omega_n) F_{m+1}(d\omega_{m+1}) \dots F_n(d\omega_n), \quad \text{si } n > m$$

$$h_{n,n} = 1_{A_n}.$$

Les funcions  $h_{n,m}$ , definides per  $n \geq m$ , són mesurables i decreixen quan  $n \rightarrow \infty$ . En efecte,  $A_{n+1} \subset A_n \times \Omega$ , ja que  $B_{n+1} \subset B_n$ , i per tant,  $h_{n,m}(\omega_1, \dots, \omega_m) \geq h_{n+1,m}(\omega_1, \dots, \omega_m)$ .

Sigui  $h_m(\omega_1, \dots, \omega_m) = \lim_m \downarrow h_{n,m}(\omega_1, \dots, \omega_m)$ . Per convergència monòtona tindrem

$$\int_{\Omega^m} h_m(\omega_1, \dots, \omega_m) P_1(d\omega_1) \dots P_m(d\omega_m) = \lim_n P(B_n) > 0.$$

D'altra banda, si  $n > m$

$$\begin{aligned} \int_{\Omega^m} h_{n,m}(\omega_1, \dots, \omega_m) P_m(d\omega_m) &= h_{n,m-1}(\omega_1, \dots, \omega_{m-1}), \\ &= P_1(A_1) \dots P_{m-1}(A_{m-1}). \end{aligned}$$

i fent  $n \rightarrow \infty$  obtenim

$$\int_{\Omega} h_m(\omega_1, \dots, \omega_m) P_m(d\omega_m) = h_{m-1}(\omega_1, \dots, \omega_{m-1}).$$

Com que  $\int_{\Omega} h_1(\omega_1) P_1(d\omega_1) > 0$ , existeix un element  $\omega_1$  tal que  $h_1(\omega_1) > 0$ .

Com que  $\int_{\Omega} h_2(\omega_1, \omega_2) P_2(d\omega_2) = h_1(\omega_1) > 0$ , existeix un altre element  $\omega_2$  tal que  $h_2(\omega_1, \omega_2) > 0$ . Aleshores, per recurrència podem trobar una successió  $\{\omega_n, n \geq 1\}$  tal que  $h_m(\omega_1, \dots, \omega_m) > 0$  per a tot  $m \geq 1$ . Això implica que  $h_{n,m}(\omega_1, \dots, \omega_m) = 1_{A_m}(\omega_1, \dots, \omega_m) > 0$ . Es a dir,  $(\omega_1, \dots, \omega_m) \in A_m$  per a tot  $m \geq 1$  i, en conseqüència,  $\{\omega_n, n \geq 1\} \in \cap_{n=1}^{\infty} B_n$ . En conclusió  $\cap_{n=1}^{\infty} B_n \neq \emptyset$  i la  $\sigma$ -additivitat de  $P$  queda demonstrada.

Finalment, la unicitat d'una probabilitat  $P$  en l'espai  $(\Omega^{\mathbb{N}}, \mathcal{A}^{\mathbb{N}})$  que compleixi (7.11) surt del fet que la família de conjunts  $\{\pi_n^{-1}(A_1) \cap \dots \cap \pi_n^{-1}(A_n); n \geq 1, A_1, \dots, A_n \in \mathcal{A}\}$  és una semialgebra que genera la  $\sigma$ -àlgebra  $\mathcal{A}^{\mathbb{N}}$ .

La probabilitat  $P$  es representa per  $\otimes_{n=1}^{\infty} P_n$ , i si totes les probabilitats  $P_n$  són iguals a una probabilitat  $Q$ , escriurem  $P = Q^{\mathbb{N}}$ .

Considererem una certa probabilitat  $\mu$  en  $\mathbb{R}$ . El Teorema anterior ens permet de construir l'espai producte  $(\mathbb{R}^{\mathbb{N}}, \mathcal{B}(\mathbb{R})^{\mathbb{N}}, \mu^{\mathbb{N}})$ . Les projeccions cànoniques  $\{\pi_n, n \geq 1\}$  definides en aquest espai de probabilitat constitueixen una successió de variables aleatories independents i idènticament distribuïdes (i.i.d.), amb la mateixa distribució  $\mu$ .

## 8. LA DISTRIBUCIÓ BINOMIAL. PASSEJADA ALEATÒRIA SOBRE ELS ENTERS

En aquest capítol introduirem alguns exemples de lleis de probabilitat unidimensionals i estudiarem problemes relacionats amb elles.

### 8.1. Lleis de Bernoulli i Binomial

Considerem un nombre  $p$  de l'interval  $[0,1]$  i posem  $q = 1 - p$ . S'anomena llei de Bernoulli de paràmetre  $p$  a la probabilitat  $p\delta_1 + q\delta_0$ .

Aquesta llei de probabilitat es pot associar a tota experiència aleatòria en la qual hi hagin dos resultats possibles, que sempre podem representar per 1 i 0, respectivament.

Per exemple, si en un espai de probabilitat  $(\Omega, \mathcal{A}, P)$  considerem un esdeveniment  $A \in \mathcal{A}$  de probabilitat  $p$ , és immediat comprovar que la variable aleatòria  $X = 1_A$  té una llei de Bernoulli de paràmetre  $p$ . És a dir,  $X$  és una variable que val 1 ó 0 segons que l'esdeveniment  $A$  s'hagi realitzat o no.

Considerem ara  $n$  variables aleatories independents  $X_1, \dots, X_n$  amb lleis de Bernoulli de paràmetre  $p$ . Aleshores, la llei de la variable  $S_n = X_1 + \dots + X_n$  s'anomena *lleí binomial de paràmetres  $n, p$*  i es representa per  $B(n, p)$ . La variable  $S_n$  representa el nombre de vegades que s'ha produït un cert esdeveniment  $A$  de probabilitat  $p$  quan repetim  $n$  vegades i de manera independent una experiència aleatòria.

La variable  $S_n$  pot prendre els valors  $0, 1, \dots, n$  amb les següents probabilitats

$$\begin{aligned} P\{S_n = k\} &= P\{X_1 + \dots + X_n = k\} \\ &= \sum_{\substack{(i_1, \dots, i_n) \in \{0,1\}^n, \\ i_1 + \dots + i_n = k}} P\{X_1 = i_1, \dots, X_n = i_n\} \\ &= \binom{n}{k} p^k (1-p)^{n-k}, \end{aligned}$$

per  $k = 0, 1, \dots, n$ . Per tant, la lleí binomial serà

$$B(n, p) = \sum_{k=0}^n \binom{n}{k} p^k (1-p)^{n-k} \delta_k.$$

Tenint en compte les propietats de les variables aleatories independents, és fàcil calcular

la mitjana i la variància d'una lleí binomial:

$$\begin{aligned} E(S_n) &= np \\ \sigma^2(S_n) &= \sum_{i=1}^n \sigma_i^2(X_i) = np(1-p). \end{aligned}$$

Tractem tot seguit un problema d'aproximació de la lleí binomial  $B(n, p)$ . Considerem una successió de lleis binomials  $B(n, p_n)$  i suposem que els valors del paràmetre  $p_n$  decreixen de manera  $\lim_n np_n = \lambda > 0$ . Observi's que això implica  $\lim_n p_n = 0$ . Aleshores per a tot nombre natural  $k \geq 0$  hem té

$$\begin{aligned} \lim_n B(n, p_n)(\{k\}) &= \lim_n \binom{n}{k} p_n^k (1-p_n)^{n-k} \\ &= \lim_n \frac{n!}{k!(n-k)!} p_n^k (1-p_n)^{n-k} \\ &= \lim_n \frac{(np_n)^k}{k!} \frac{n!}{n^k(n-k)!} [(1-p_n)^{1/p_n}]^{np_n} (1-p_n)^{-k} = \frac{\lambda^k}{k!} e^{-\lambda}, \end{aligned}$$

ja que  $\lim_n (1-p_n)^{-k} = 1$  i  $\lim_n \frac{(n-k+1)\dots n}{n^k} = 1$ .

Anomenem lleí de Poisson de paràmetre  $\lambda > 0$  la lleí de probabilitat discreta, que té per suport els nombres naturals i ve donada per

$$\sum_{k=0}^{\infty} \frac{\lambda^k}{k!} e^{-\lambda} \delta_k.$$

Les aproximacions de la lleí binomial per la lleí de Poisson són bones quan  $p < 0.1$  i  $np < 5$ . La lleí de Poisson s'utilitza com a model probabilístic en molts problemes pràctics: Considerem una successió d'esdeveniments que es produeixen al llarg del temps com les trucades que rep un aparell de telèfon, les persones que arriben a una cua o les partícules que emet un material radioactiu. Si fixem un interval de temps  $[t_1, t_2]$  i comptem el nombre total d'esdeveniments que s'han produït en aquest període de temps, aquest nombre aleatori  $k \geq 0$  segueix una lleí de Poisson.

Representarem la lleí de Poisson de paràmetre  $\lambda > 0$  per  $\text{Pois}(\lambda)$ . Es pot comprovar que té mitjana  $\lambda$  i variància  $\lambda$ .

### 8.2. La passejada aleatòria de Bernoulli

Considerem un individu que "passeja" en el conjunt dels nombres enters de la forma següent. En l'instant inicial es troba a l'origen, i si en l'instant  $n$  es troba en un punt

$x \in \mathbb{Z}$ . en l'instant següent es desplaça al punt  $x+1$  amb probabilitat  $p$  ( $p \in [0, 1]$ ) o bé al punt  $x-1$  amb probabilitat  $q = 1-p$ . Suposarem que els desplaçaments successius són independents.

Aquesta passejada podria fer-se prenent en cada instant la decisió de desplaçar-se cap a la direta o cap a l'esquerra segons el resultat del llançament d'una moneda.

Si  $S_n$  representa la posició d'aquest individu a l'instant  $n$ , podem escriure  $S_n = X_1 + \dots + X_n$ ,  $S_0 = 0$ , on les  $\{X_n, n \geq 1\}$  són una successió de variables independents tals que

$$P\{X_n = 1\} = p, \quad P\{X_n = -1\} = q.$$

La successió  $\{S_n, n \geq 0\}$  s'anomena una *passejada aleatòria de Bernoulli* de paràmetre  $p$  que surt de l'origen.

Observem que si s'han produït  $k$  desplaçaments cap a la direta i  $n-k$  cap a l'esquerra, la variable  $S_n$  pren el valor  $2k-n$ , i

$$P\{S_n = 2k-n\} = \binom{n}{k} p^k q^{n-k},$$

$$k = 0, 1, \dots, n.$$

En altres paraules  $\frac{1}{2}(S_n + n)$  té una llei  $B(n, p)$  i  $S_n$  pren els valors  $-n, 2-n, 4-n, \dots, n$ .

Ens proposem ara donar alguns resultats interessants sobre la passejada aleatòria de Bernoulli.

- (a) Considerem una passejada aleatòria de Bernoulli  $\{S_n, n \geq 0\}$  que surt de l'origen, amb  $p = q = \frac{1}{2}$ . Sigui  $x = 2k-n$ ,  $k = 0, 1, \dots, n$ , un possible valor de  $S_n$ . Si  $x > 0$ , es compleix

$$P\{S_1 > 0, S_2 > 0, \dots, S_n > 0 / S_n = x\} = \frac{x}{n}. \quad (\text{S.1})$$

Aquesta propietat justifica el següent "Teorema de la votació", establert per W.A. Whitworth l'any 1878 i també per J. Bertrand l'any 1887, que diu la cosa següent:

"Considerem una votació entre dos candidats A i B. En el moment del recompte, i en treure la n-èssima papereta de l'urna, el candidat A té  $k$  vots i el candidat B  $j = n-k$ , amb  $k > j$ . Aleshores, la probabilitat que en tot moment del recompte de vots A hagi tingut més vots que B és igual a  $\frac{2k-n}{n}$ ".

La demostració de la igualtat (S.1) es basa en l'anomenat "príncipi de reflexió" que enunciem una mica més endavant. En primer lloc observem que els esdeveniments  $\{X_1 = \varepsilon_1, \dots, X_n = \varepsilon_n\}$ , on  $(\varepsilon_1, \dots, \varepsilon_n) \in \{-1, 1\}^n$  i  $S_n = X_1 + \dots + X_n$ ,

tenen tots la mateixa probabilitat igual a  $2^{-n}$ . Per a calcular  $P\{S_1 > 0, S_2 > 0, \dots, S_n > 0 / S_n = x\}$  haurim de determinar el nombre d'esdeveniments del tipus  $\{X_1 = \varepsilon_1, \dots, X_n = \varepsilon_n\}$  favorables a  $\{S_1 > 0, S_2 > 0, \dots, S_n > 0, S_n = x\}$ .

Definim  $\sigma_0 = \varepsilon_0 = 0$  i  $\sigma_k = \varepsilon_1 + \dots + \varepsilon_k$ . Donar una  $n$ -pla ordenada  $\varepsilon = (\varepsilon_1, \dots, \varepsilon_n)$  equival a donar la "trajectòria"  $\sigma = (\sigma_1, \dots, \sigma_n)$  on  $\varepsilon_k = \sigma_k - \sigma_{k-1}$ .

Cal doncs calcular el nombre de trajectòries que surten de 1, van a parar a  $x$  i no passen per 0. Mitjançant una reflexió respecte l'eix del temps es pot veure que

El nombre de trajectòries que van de 1 a  $x$  i toquen o passen pel zero coincideix amb el nombre de totes les trajectòries que van del -1 a  $x$ .

Sigui  $x = 2k-n$  i definim

$N =$  nombre de trajectòries  $\sigma$  que van del punt 1 al  $x$  sense passar pel zero,

$N(1, x) =$  nombre de trajectòries  $\sigma$  que van del punt 1 al  $x$ ,

$N(-1, x) =$  nombre de trajectòries  $\sigma$  que van del punt -1 al  $x$ .

Tenim aleshores que  $N = N(1, x) - N(-1, x)$ . Si  $\sigma = (\sigma_1, \dots, \sigma_n)$  és una trajectòria que va de 1 a  $x$  haurà de ser  $\varepsilon_1 = 1$  i  $\varepsilon_2 + \dots + \varepsilon_n = x$ . Això equival a dir que  $\varepsilon_1 = 1$  i que dels  $\varepsilon_2, \dots, \varepsilon_n$  n'hi ha  $k-1$  iguals a 1. En efecte,

$$\frac{1}{2}[(n-1) + \varepsilon_2 + \dots + \varepsilon_n] = \frac{1}{2}[n-1 + 2k-n-1] = k-1.$$

En conseqüència,

$$N(1, x) = \binom{n-1}{k-1}.$$

Anàlogament, si  $\sigma = (\sigma_1, \dots, \sigma_n)$  és una trajectòria que va de -1 a  $x$  haurà de ser  $\varepsilon_1 = -1$  i els  $\varepsilon_2, \dots, \varepsilon_n$  hauran de ser  $k$  vegades iguals a 1 ja que

$$\frac{1}{2}[(n-1) + \varepsilon_2 + \dots + \varepsilon_n] = \frac{1}{2}[n-1 + 2k-n+1] = k.$$

Per tant,

$$N(-1, x) = \binom{n-1}{k},$$

Així doncs,

$$\begin{aligned} P\{S_1 > 0, \dots, S_n > 0 / S_n = x\} &= \frac{P\{S_1 > 0, \dots, S_{n-1} > 0, S_n = x\}}{P\{S_n = x\}} = \frac{N 2^{-n}}{\binom{n}{k} 2^{-n}} \\ &= \frac{N}{\binom{n}{k}} = \frac{\binom{n-1}{k-1} - \binom{n-1}{k}}{\binom{n}{k}} = \frac{2k-n}{n}. \end{aligned}$$

(b) El problema de la *ruïna del jugador*.

Un jugador  $A$  té un capital  $x \in [0, b]$  i juga contra un altre jugador  $B$  que té un capital  $b - x$ . A cada jugada el jugador  $A$  guanya una unitat amb probabilitat  $p$  o bé perd una unitat amb probabilitat  $q = 1 - p$ . El procés continua fins que  $A$  ho perd tot o guanya un capital  $b$ . Es tracta de calcular la probabilitat que el jugador  $A$  s'arriuni.

Aquest problema es pot formular en termes de la passejada aleatorià. Signi  $S_n = x + \sum_{i=1}^n X_i$ , una passejada aleatorià de Bernoulli, de paràmetre  $p$  que surt del punt  $x$ . És a dir,  $S_0 = x$  i les  $\{X_i, i \geq 1\}$  són variables aleatories independents tals que  $P\{X_i = 1\} = p$ ,  $P\{X_i = -1\} = q$ . Considerem l'esdeveniment

$$\begin{aligned} G_x &= \cup_{k=0}^{\infty} \{S_k = 0, S_i < b \text{ per a tot } i = 0, 1, \dots, k\} \\ &= \{\text{la successió } S_n \text{ arriba alguna vegada al zero i abans no ha arribat a } b\}. \end{aligned}$$

$S_n$  representa la fortuna del jugador  $A$  a l'instant  $n$  i  $P(G_x)$  és la probabilitat que el jugador  $A$  s'arriuni.

Considerem una nova successió de sumes parcials definida per  $S'_k = X_2 + \dots + X_k$ ,  $k \geq 2$ .

Aleshores podem escriure, per  $1 \leq x \leq b - 1$

$$\begin{aligned} G_x \cap \{X_1 = 1\} &= \left[ \bigcup_{k=2}^{\infty} \{x + 1 + S'_k = 0, 0 < x + 1 + S'_i < b \mid i \geq k - 1\} \right] \\ &\cap \{X_1 = 1\} = G'_x \cap \{X_1 = 1\}, \end{aligned}$$

i anàlogament,

$$\begin{aligned} G_x \cap \{X_1 = -1\} &= \left[ \bigcup_{k=2}^{\infty} \{x - 1 + S'_k = 0, 0 < x - 1 + S'_i < b \mid i \leq k - 1\} \right] \\ &\cap \{X_1 = -1\} = G''_x \cap \{X_1 = -1\}. \end{aligned}$$

Les successions de variables aleatories  $\{X_n, n \geq 1\}$  i  $\{X_{n+1}, n \geq 1\}$  tenen la mateixa llei (és a dir, indueixen la mateixa probabilitat en l'espai producte  $\mathbb{R}^N$ ). En conseqüència,

$$P(G'_x) = P(G_{x+1}) \quad i \quad P(G''_x) = P(G_{x-1}),$$

si  $1 \leq x \leq b - 1$ .

Posem  $h(x) := P(G_x)$ . Aleshores, per  $1 \leq x \leq b - 1$ , tindrem

$$\begin{aligned} h(x) &= P(G_x) = P[G_x \cap \{X_1 = 1\}] + P[G_x \cap \{X_1 = -1\}] \\ &= P(G'_x)P\{X_1 = 1\} + P(G''_x)P\{X_1 = -1\} \\ &= ph(x+1) + qh(x-1). \end{aligned} \tag{8.2}$$

Per  $x = 1$ , obtenim

$$h(1) = P[G_1 \cap \{\xi_1 = 1\}] + P\{\xi_1 = -1\} = ph(2) + q,$$

i l'equació (8.2) també es compleix ja que  $h(0) = P(\Omega) = 1$ .

Per  $x = b - 1$ , obtenim

$$h(b-1) = P[G_{b-1} \cap \{\xi_1 = -1\}] = qh(b-2),$$

en conseqüència  $h(b) = 0$ . Per tant, hem de resoldre l'equació en diferències finites

$$h(x) = ph(x+1) + qh(x-1), \quad 1 \leq x \leq b-1,$$

amb les condicions a la frontera,  $h(0) = 1$ ,  $h(b) = 0$ .

Podem escriure aquesta equació de la forma següent

$$ph(x+2) - h(x+1) + qh(x) = 0, \quad 0 \leq x \leq b-2. \tag{8.3}$$

Considerem una solució de la forma  $h(x) = \lambda^x$ ,  $\lambda \in \mathbb{R}$ . Aleshores de

$$p\lambda^{x+2} - \lambda^{x+1} + q\lambda^x = 0,$$

se'n dedueix:

$$\lambda = \frac{1 \pm \sqrt{1 - 4pq}}{2p} = \frac{1}{2p}(1 \pm |p - q|),$$

ja que

$$(p - q)^2 = (p - 1 + p)^2 = 4p^2 - 4p + 1 = 1 - 4p(1 - p) = 1 - 4p^2.$$

Cal doncs considerar dos casos diferents:

(i) Si  $p \neq q$ , les dues raons de l'equació són diferents:

$\lambda_1 = 1, \lambda_2 = \frac{q}{p}$  i la solució general de l'equació (8.3) és  $h(x) = A\lambda_1^x + B\lambda_2^x = A + B\left(\frac{q}{p}\right)^x$ . Els coeficients  $A$  i  $B$  es poden determinar a partir de les condicions de contorn. La solució és

$$h(x) = \frac{\left(\frac{q}{p}\right)^x - \left(\frac{q}{p}\right)^b}{1 - \left(\frac{q}{p}\right)^b}.$$

(ii) Suposem  $p = q = \frac{1}{2}$ . En aquest cas,  $\lambda_1 = \lambda_2 = 1 = \lambda$ . Aleshores podem trobar dues solucions linealment independents prenent  $\lambda^x$  i  $x\lambda^x$ . La solució general és  $A\lambda^x + Bx\lambda^x = A + Bx$ ; i fent servir les condicions de contorn obtenim

$$h(x) = \frac{b-x}{b}.$$

No és difícil comprovar que l'equació (8.3) amb les condicions de contorn  $h(0) = 1$ ,  $h(b) = 0$  té una solució única. En efecte, les funcions  $f_1(x) = \lambda_1^x$  i  $f_2(x) = \lambda_2^x$  (o bé  $f_1(x) = \lambda^x$  i  $f_2(x) = x\lambda^x$  en el cas d'una rel doble) són linealment independents i donada qualsevol altre solució  $f(x)$  es poden trobar constants  $\alpha$  i  $\beta$  tals que

$$\begin{pmatrix} f(0) \\ f'(0) \end{pmatrix} = \alpha \begin{pmatrix} f_1(0) \\ f_1'(0) \end{pmatrix} + \beta \begin{pmatrix} f_2(0) \\ f_2'(0) \end{pmatrix}.$$

Però, aleshores,

$$\begin{aligned} f(2) &= \frac{1}{p}f(0) - \frac{q}{p}(1) \\ &= \frac{1}{p}(\alpha f_1(0) + \beta f_2(0)) - \frac{q}{p}(\alpha f_1(1) + \beta f_2(1)) \\ &= \alpha f_1(2) + \beta f_2(2). \end{aligned}$$

Recursivament s'obté  $f(x) = \alpha f_1(x) + \beta f_2(x)$  i com que les constants  $\alpha$  i  $\beta$  quedan determinades per les condicions de contorn conclouem que la solució és única.

La funció  $h(x)$  representa la probabilitat de que el jugador  $A$  s'arruïni. De forma anàloga podem introduir la probabilitat  $g(x)$  que el jugador  $A$  guanyi. Anem a veure que  $h(x) + g(x) = 1$ . Això no és immediat ja que el joc podrà no acabar-se mai.

Definim

$$F_x = \sum_{k=0}^{\infty} \{S_k = b, S_i > 0, \forall i = 0, 1, \dots, k\}.$$

Llavors  $g(x) = P(F_x)$  però pot acórrer que  $F_x \cup G_x \neq \Omega$ .

Mitjançant un argument similar a l'anterior podem obtenir una equació en diferències finites  $g(x) = pg(x+1) + qg(x-1)$ ,  $1 \leq x \leq b-1$  amb les condicions de contorn  $g(0) = 0$ ,  $g(b) = 1$ . Com que  $1 - h(x)$  compleix aquestes condicions, de la unitat de la solució es dedueix que  $g(x) = 1 - h(x)$ .

Suposem ara que fem  $b \uparrow \infty$ . Els conjunts  $G_x$  creixen cap al conjunt

$$H = \bigcup_{k=0}^{\infty} \{S_k = 0\}.$$

Aleshores,

$$P(H) = \lim_{b \rightarrow \infty} P(G_z) = \begin{cases} 1 & \text{si } q \geq p, \\ \left(\frac{q}{p}\right)^z & \text{si } q < p, \end{cases}$$

suposant  $z > 0$ .

$P(H)$  representa la probabilitat d'arruinar-se contra un jugador de capital infinit.  $1 - P(H)$  serà la probabilitat que en aquesta situació el joc no acabi mai, que és no nul·la si  $q < p$ . Escriurem  $f_p(x) = P(H)$ .

(c) Considerem finalment el següent problema.  $\{S_n, n \geq 0\}$  és una passejada aleatòria de Bernoulli que surt de l'origen ( $S_0 = 0$ ) i volem calcular la probabilitat de tornar a alguna vegada a l'origen. Sigui

$$A = \{S_n = 0, \text{ per algun } n \geq 1\}.$$

Tindrem

$$\begin{aligned} P(A) &= P(A \cap \{S_1 = 1\}) + P(A \cap \{S_1 = -1\}) \\ &= pP\{1 + X_2 + \dots + X_n = 0, \text{ per algun } n \geq 2\} \\ &\quad + qP\{-1 + X_2 + \dots + X_n = 0, \text{ per algun } n \geq 2\} \\ &= pf_p(1) + qP\{1 - X_2 - \dots - X_n = 0, \text{ per algun } n \geq 2\} \\ &= pf_p(1) + qf_q(1) \\ &= \begin{cases} 1, & \text{si } p = q \\ 1 - |p - q|, & \text{si } p \neq q. \end{cases} \end{aligned}$$

En efecte, si  $p < q$ , obtenim  $p + q\left(\frac{p}{q}\right) = 2p$ , i si  $p > q$ , obtenim  $p\left(\frac{q}{p}\right) + q = 2q$ .

Aquest resultat expressa la coneguda propietat de que la passejada aleatòria de Bernoulli és recurrent per  $p = q$  i transitòria per  $p \neq q$ .

### 8.3 Altres distribucions de probabilitat relacionades amb la binomial

**Llei hipergeomètrica.** Considerem una població de  $N$  individus dels quals  $D$  tenen una determinada característica i els  $N - D$  restants no la tenen. Un problema estadístic important és l'estimació de la proporcio  $p = \frac{D}{N}$  d'individus que tenen la propietat esmentada. Podem imaginar que els individus es representen per un conjunt de boles contingudes en una urna de les quals  $D$  són negres i  $N - D$  són blancs.

Suposarem que no es poden observar tots els individus de la població i aleshores ens caldrà fer una estimació de  $p$  a partir d'una mostra de mida  $n \leq N$ . Hi ha bàsicament dues maneres d'escollar aquesta mostra:

- (a) *Mostreig amb remplaçament.* Seleccionem a l'atzar i de forma independent  $n$  boles de la urna de manera que cada bola es torna a la urna després d'esser observada. És a dir, les boles es treuen de forma successiva i la composició de la urna és sempre la mateixa, amb  $D$  boles negres i  $N - D$  blanques. En aquest cas el nombre  $X$  de boles negres de la mostra segueix una llei binomial  $B(n, p)$  amb esperança  $np = \frac{nD}{N}$  i variància  $np(1 - p) = \frac{nD(N - D)}{N^2}$ .

- (b) *Mostreig sense remplaçament.* Seleccionem a l'atzar i de forma independent  $n$  boles sense tornar-les a la urna. En aquest cas la composició de la urna va variant després de cada selecció. Això equival a treure simultàniament  $n$  boles de la urna.

Si designem per  $X$  el nombre de boles negres en aquesta mostra de mida  $n$  tindrem que

$$P\{X = k\} = \frac{\binom{D}{k} \binom{N-D}{n-k}}{\binom{N}{n}}.$$

Això es dedueix d'un càlcul elemental de combinatòria:

Suposem que totes les possibles mostres de  $n$  elements tenen la mateixa probabilitat i aleshores  $\binom{N}{n}$  és el nombre total de mostres amb  $n$  boles, i  $\binom{D}{k} \binom{N-D}{n-k}$  és el nombre total de mostres amb exactament  $k$  boles negres.

Observem en primer lloc que el nombre  $k$  ha de complir les següents restriccions:

$$k \leq \min(D, n), \quad n - k \leq \min(N - D, n),$$

O sigui,

$$\max(n - N + D, 0) \leq k \leq \min(D, n).$$

La llei de probabilitat de  $X$  s'anomena *lleï hipergeomètrica*.

Comprovarem tot seguit que els nombres  $p_k = \frac{\binom{D}{k} \binom{N-D}{n-k}}{\binom{N}{n}}$  formen una llei de probabilitat. és a dir,

$$\sum_{k=(n-N+D)^+}^{n \wedge D} p_k = 1.$$

i

$t = (n - N + D)^+$

Uma manera senzilla de veure això és la següent:

Si  $t \in \mathbb{R}$ , tenim

$$\sum_{n=0}^N t^n \binom{N}{n} = (1+t)^N = (1+t)^D (1+t)^{N-D}$$

$$= \left[ \sum_{k=0}^D t^k \binom{D}{k} \right] \left[ \sum_{\ell=0}^{N-D} t^\ell \binom{N-D}{\ell} \right] = \sum_{k=0}^D \sum_{\ell=0}^{N-D} t^{k+\ell} \binom{D}{k} \binom{N-D}{\ell}.$$

En conseqüència, igualant coeficients s'obté

$$\binom{N}{n} = \sum_{\substack{(k, \ell) \\ 0 \leq k \leq n \\ 0 \leq \ell \leq N-D}} \binom{D}{k} \binom{N-D}{\ell} = \sum_{k=(n-N+D)^+}^{n \wedge D} \binom{D}{k} \binom{N-D}{n-k}.$$

Calculem l'esperança i la variància d'aquesta distribució:

$$\begin{aligned} E(X) &= \binom{N}{n}^{-1} \sum_{k=(n-N+D)^+}^{n \wedge D} k \binom{D}{k} \binom{N-D}{n-k} \\ &= \binom{N}{n}^{-1} D \sum_{k=(n-N+D)\vee 1}^{n \wedge D} D \binom{D-1}{k-1} \binom{N-D}{n-k} \\ &= \binom{N}{n}^{-1} D \sum_{j=(n-1-N+D)^+}^{(n-1)\wedge D} \binom{D-1}{j} \binom{N-D}{n-1-j} \\ &= \binom{N}{n}^{-1} D \binom{N-1}{n-1} = \frac{Dn}{N} = np. \end{aligned}$$

$$\begin{aligned} E[X(X-1)] &= \sum_{k=(n-N+D)^+}^{n \wedge D} k(k-1) \binom{N}{n}^{-1} \binom{D}{k} \binom{N-D}{n-k} = \\ &= \sum_{k=(n-N+D)\vee 2}^{n \wedge D} \binom{N}{n}^{-1} D(D-1) \binom{D-2}{k-2} \binom{N-D}{n-k} = \\ &= D(D-1) \binom{N}{n}^{-1} \sum_{j=(n-2-N+D)^+}^{(n-2)\wedge D} \binom{D-2}{j} \binom{N-D}{n-2-j} = \\ &= D(D-1) \binom{N}{n}^{-1} \binom{N-2}{n-2} = \frac{D(D-1) N(N-1)}{N(N-1)}. \end{aligned}$$

$$\begin{aligned}\sigma^2(X) &= E[X^2] - [E(X)]^2 = \frac{D(D-1)n(n-1)}{N(N-1)} + \frac{Dn}{N} - \left(\frac{Dn}{N}\right)^2 \\ &= \frac{D(N-D)n(N-n)}{N^2(N-1)} = n p(1-p) \frac{N-n}{N-1}.\end{aligned}$$

Observem que l'esperança matemàtica del nombre  $X$  de boles negres de la mostra és igual a  $np$  en els dos tipus de mostreig (amb reemplaçament i sense reemplaçament), mentre que la variància en el mostreig sense reemplaçament conté un factor igual a  $\frac{N-n}{N-1}$  que és proxim a 1 quan  $N$  és molt gran.

### Llei geomètrica

Considerem una successió  $\{X_n, n \geq 1\}$  de variables aleatòries de Bernoulli; independents, i amb paràmetre  $p \in (0, 1)$ . La variable  $X = \inf \{n \geq 1 : X_n = 1\}$  representa el primer instant on la successió pren el valor 1. Si les variables  $X_n$  ens indiquen si un determinat esdeveniment es produeix o no en una sèrie infinita de realitzacions independents de l'experiència aleatòria, aleshores  $X$  és l'instant on l'esdeveniment es produeix per primera vegada.

La distribució de probabilitat de la variable  $X$  s'anomena *lleï geomètrica* de paràmetre  $p$ . Es tracta d'una lleï discreta concentrada sobre els nombres naturals, donada per:

$$P\{X = k\} = (1-p)^{k-1}p, \quad k \geq 1.$$

Clarament,  $\sum_{k=1}^{\infty} (1-p)^{k-1} p = 1$ .

Es pot comprovar que l'esperança d' $X$  val  $\frac{1}{p}$  i la variància  $\frac{q}{p^2}$ .

### Llei binomial negativa

En la mateixa situació d'abans, sigui  $T_n$  l'instant en el qual la successió pren el valor 1 per  $n$ -èsima vegada i designem per  $Y$  el nombre total de zeros en la seqüència finita  $X_1, X_2, \dots, X_{T_n}$ .

Que la variable  $Y$  prengui un valor  $k \geq 0$  vol dir que  $T_n = n+k$  i que en la seqüència  $X_1, X_2, \dots, X_{T_n}$  hi ha  $k$  zeros i  $n$  uns. En conseqüència,

$$P\{Y = k\} = \binom{n+k-1}{n-1} p^n (1-p)^k.$$

Per tal de comprovar que  $\sum_{k=0}^{\infty} P\{Y = k\} = 1$ , cal fer servir la relació

$$\binom{-n}{k} = \frac{(-n)(-n-1)(-n-2)\dots(-n-k+1)}{k!} = (-1)^k \binom{n+k-1}{k}.$$

Aleshores, tindrem

$$\begin{aligned}&\sum_{k=0}^{\infty} \binom{n+k-1}{k} p^n (1-p)^k = \sum_{k=0}^{\infty} (-1)^k \binom{-n}{k} p^n (1-p)^k \\ &= p^n \sum_{k=0}^{\infty} \binom{-n}{k} (p-1)^k = p^n (1-(1-p))^{-n} = 1.\end{aligned}$$

La lleï de  $Y$  s'anomena *distribució binomial negativa*. La seva esperança val  $\frac{np}{p}$  i la variància  $\frac{np}{p^2}$ .

Aquesta darrera desigualtat ens proporciona una estimació de la probabilitat de la desviació de la variable  $X$  al voltant del seu valor mig  $E(X)$  en termes de la variància  $\sigma^2(X)$ . En particular, per un valor fix  $\alpha \in (0, 1)$ , podem assegurar que, amb probabilitat d'error inferior a  $\alpha$ , la variable  $X$  pertany a l'interval  $[E(X) - \frac{\sigma(X)}{\sqrt{\alpha}}, E(X) + \frac{\sigma(X)}{\sqrt{\alpha}}]$ . Observis que la longitud d'aquest interval és proporcional a la desviació típica  $\sigma(X)$ .

## 9. CONVERGÈNCIA DE VARIABLES ALEATÒRIES

En aquest capítol estudiarem diferents tipus de convergència de successions de variables aleatòries i les seves relacions.

### 9.1. Preliminars

Dediquem aquesta secció a presentar alguns resultats que seran útils en l'estudi de les convergències de successions de variables aleatòries.

#### La desigualtat de Txebixef

**Proposició 9.1.** Sigui  $X$  una variable aleatòria no negativa i  $f = \mathbb{R}_+ \rightarrow \mathbb{R}_+$  una funció creixent tal que  $E(f(X)) < \infty$ . Sigui  $a$  un nombre real tal que  $f(a) > 0$ . Aleshores es compleix

$$P\{X \geq a\} \leq \frac{1}{f(a)} E(f(X)). \quad (9.1)$$

**Demostració:** Considerem la desigualtat

$$f(a) 1_{\{X \geq a\}} \leq f(X).$$

Prentent esperances hom obté

$$f(a) P\{X \geq a\} \leq E(f(X)),$$

que és la desigualtat buscada. ■

En particular, si  $X \in L^p(\Omega, \mathcal{A}, P)$  amb  $p \geq 1$ , i  $a > 0$ , tindrem

$$P\{|X| \geq a\} \leq \frac{E(|X|^p)}{a^p}. \quad (9.2)$$

En efecte, (9.2) s'obté aplicant la desigualtat (9.1) a la funció  $f(x) = x^p$  i a la variable  $|X|$ .

Sigui  $X$  una variable aleatòria de quadrat integrable. La desigualtat (9.2) aplicada a la variable aleatòria  $|X - E(X)|$  i a  $p = 2$  ens dóna

$$P\{|X - E(X)| \geq a\} \leq \frac{\sigma^2(X)}{a^2}. \quad (9.3)$$

### Lemes de Borel-Cantelli

Considerem una successió  $\{A_n, n \geq 1\}$  de conjunts d'un espai mesurable  $(\Omega, \mathcal{A})$ . Definim

$$\limsup_n A_n = \bigcap_{n \geq 1} \bigcup_{k \geq n} A_k,$$

$$\liminf_n A_n = \bigcup_{n \geq 1} \bigcap_{k \geq n} A_k.$$

D'aquestes definicions se'n dedueix fàcilment que

- (i)  $\omega$  pertany a  $\limsup_n A_n$  si i únicament si  $\omega$  pertany a infinitis  $A_n$ .
- (ii)  $\omega$  pertany a  $\liminf_n A_n$  si i únicament si  $\omega$  pertany a tots els  $A_n$  per  $n \geq n_0(\omega)$ .

Són també immediates de comprovar les següents relacions

$$\begin{aligned} \liminf_n A_n &\subset \limsup_n A_n, \\ (\limsup_n A_n)^c &= \liminf_n A_n^c, \end{aligned}$$

i

$$(\liminf_n A_n)^c = \limsup_n A_n^c.$$

Signi  $P$  una probabilitat en l'espai  $(\Omega, \mathcal{A})$ . Tindrem aleshores el següent resultat

**Lema 9.2.** Donada una successió  $\{A_n, n \geq 1\}$  d'esdeveniments de  $\mathcal{A}$ , es compleix

$$P\{\limsup_n A_n\} \geq \limsup_n P(A_n),$$

$$P\{\liminf_n A_n\} \leq \liminf_n P(A_n).$$

**Demostració:** És suficient demostrar la primera desigualtat. La segona se'n dedueix per pas al complementari.

Podem escriure

$$\begin{aligned} P\{\limsup_n A_n\} &= \lim_n P\{\bigcup_{m \geq n} A_m\} \geq \limsup_n P(A_n) \\ &= \limsup_n P(A_n). \end{aligned}$$

El resultat queda així provat. ■

Presentem tot seguit els lemes de Borel-Cantelli.

**Lema 9.3. (Primer lema de Borel-Cantelli).** Sigui  $\{A_n, n \geq 1\}$  una successió de conjunts de  $\mathcal{A}$ . Si  $\sum_{n=1}^{\infty} P(A_n) < \infty$  es compleix

$$P\{\limsup_n A_n\} = 0.$$

**Demostació:** Hom té

$$\begin{aligned} P\{\limsup_n A_n\} &= P\left(\bigcap_{n=1}^{\infty} \bigcup_{m \geq n} A_m\right) = \lim_n P\left(\bigcup_{m \geq n} A_m\right) \\ &\leq \lim_n \sum_{m \geq n} P(A_m) = 0, \end{aligned}$$

i en conseqüència el resultat. ■

**Lema 9.4 (Segon lema de Borel-Cantelli).** Sigui  $\{A_n, n \geq 1\}$  una successió d'esdeveniments independents. Si  $\sum_{n=1}^{\infty} P(A_n) = \infty$ , aleshores

$$P\{\limsup_n A_n\} = 1.$$

**Demostació:** Veurem que  $P\{(\limsup_n A_n)^c\} = P\left(\bigcup_{n=1}^{\infty} \bigcap_{m \geq n} A_m^c\right) = 0$ . I per això demostrarem que  $P\left(\bigcap_{m \geq n} A_m^c\right) = 0$  per tot  $n \geq 1$ . Utilitzant la desigualtat  $1 - x \leq e^{-x}$ ,  $x \geq 0$ , tindrem

$$\begin{aligned} P\left(\bigcap_{m=n}^{n+j} A_m^c\right) &= \prod_{m=n}^{n+j} (1 - P(A_m)) \leq \prod_{m=n}^{n+j} e^{-P(A_m)} \\ &= \exp\left(-\sum_{m=n}^{n+j} P(A_m)\right), \end{aligned}$$

expressió que tendeix a zero quan  $j$  tendeix a infinit. ■

Véiem un exemple d'aplicació del segon lema de Borel-Cantelli. Sigui  $\{\varepsilon_n, n \geq 1\}$  una successió de variables aleatòries independents que prenen els valors 0 i 1 amb probabilitat  $\frac{1}{2}$ . Sigui  $\alpha = (\alpha_1, \dots, \alpha_j)$  una col·lecció ordenada de zeros i uns que considerem fixada. Tenim la propietat següent:

Amb probabilitat 1, la col·lecció ordenada  $\alpha$  apareix infinites vegades en la successió aleatòria  $\{\varepsilon_n, n \geq 1\}$ .

En efecte: Els esdeveniments

$$A_n = \{\varepsilon_{(n-1)j+1} = \alpha_1, \dots, \varepsilon_{nj} = \alpha_j\},$$

$n \geq 1$ , són independents i tots tenen probabilitat  $2^{-j}$ . Per tant la sèrie  $\sum_{n=1}^{\infty} P(A_n)$  convergeix i, pel segon lema de Borel-Cantelli  $P\{\limsup_n A_n\} = 1$ , és a dir, amb probabilitat 1 es produiran infinitis esdeveniments  $A_n$ .

## 9.2. Convergència quasi segura de variables aleatòries

**Definició 9.5.** Direm que una successió de variables aleatòries  $\{X_n, n \geq 1\}$  convergeix quasi segurament cap a una variable aleatòria  $X$  si existeix un conjunt  $N \in \mathcal{A}$  de probabilitat zero tal que

$$\lim_{n \rightarrow \infty} X_n(\omega) = X(\omega),$$

per tot  $\omega \notin N$ .

Si  $X_n$  convergeix quasi segurament cap a  $X$  escriurem  $X_n \xrightarrow{\text{q.s.}} X$ . La variable límit és única llevat en conjunts de probabilitat zero. És clar també que la successió  $\{X_n, n \geq 1\}$  convergeix quasi segurament si i només si és de Cauchy amb probabilitat 1.

**Exemple 1.** Sigui  $\Omega = [0, 1]$ ,  $\mathcal{A} = \mathcal{B}([0, 1])$  i  $P$  la mesura de Lebesgue. Definim

$$f_n(x) = \begin{cases} e^n, & \text{si } 0 \leq x \leq \frac{1}{n}, \\ 0, & \text{en cas contrari.} \end{cases}$$

És immediat comprovar que  $f_n \rightarrow 0$ , q.s., donat que l'únic punt en que  $\lim_{n \rightarrow \infty} f_n(x) \neq 0$  és  $x = 0$ .

El següent resultat ens dóna una condició equivalent per a la convergència quasi segura d'una successió de variables aleatories.

**Proposició 9.6.** La successió  $\{X_n, n \geq 1\}$  convergeix quasi segurament cap a la variable aleatòria  $X$  si i únicament si per a tot  $\varepsilon > 0$

$$\lim_{m \rightarrow \infty} P \left\{ \sup_{n \geq m} |X_n - X| \leq \varepsilon \right\} = 1.$$

*Demostració:* Suposem primer que  $\lim_{n \rightarrow \infty} X_n = X$ , q.s. i sigui  $N$  el subconjunt de  $\mathcal{A}$  de probabilitat zero on falla la convergència puntual. Designem per  $\Omega_0$  al conjunt  $\Omega - N$ . Es compleix

$$\left\{ \sup_{n \geq m} |X_n - X| \leq \varepsilon \right\} = \bigcap_{n \geq m} \{|X_n - X| \leq \varepsilon\} := A_m^\varepsilon.$$

Els conjunts  $A_m^\varepsilon$  formen una successió creixent en  $m$ . D'altra banda es té la inclusió

$$\Omega_0 \subset \bigcup_{m=1}^{\infty} A_m^\varepsilon,$$

resulta doncs

$$1 = P(\Omega_0) \leq P \left( \bigcup_{m=1}^{\infty} A_m^\varepsilon \right) = \lim_{m \rightarrow \infty} P(A_m^\varepsilon),$$

i, en conseqüència,

$$\lim_{m \rightarrow \infty} P(A_m^\varepsilon) = 1.$$

Recíprocament, fixem  $\varepsilon > 0$ . Podrem determinar un conjunt  $N_\varepsilon$  de probabilitat zero tal que per tot  $\omega \notin N_\varepsilon$ , existeix  $m_0(\varepsilon)$  i per tot  $m \geq m_0$ ,  $|X_m(\omega) - X(\omega)| \leq \varepsilon$ . Sigui  $N = \bigcup_{\varepsilon \in Q} N_\varepsilon$ . Observi's que  $P(N) = 0$ . Aleshores per tot  $\omega \notin N$  tenim  $\lim_{n \rightarrow \infty} X_n(\omega) = X(\omega)$ , és a dir, la convergència q.s. de la successió  $\{X_n, n \geq 1\}$  cap a  $X$ . Això acaba la demostració de la proposició. ■

Donem tot seguit un criteri de convergència quasi segura.

**Proposició 9.7.** Sigui  $\{\varepsilon_n, n \geq 1\}$  una successió de nombres reals positius tal que  $\sum_{n=1}^{\infty} \varepsilon_n < \infty$ . Sigui  $\{X_n, n \geq 1\}$  una successió de variables aleatories i suposem que

$$\sum_{n=1}^{\infty} P(|X_{n+1} - X_n| \geq \varepsilon_n) < \infty.$$

Aleshores  $\{X_n, n \geq 1\}$  convergeix quasi segurament.

*Demostració:* Sigui  $A_n = \{|X_{n+1} - X_n| > \varepsilon_n\}$ . Com que per hipòtesi  $\sum_{n=1}^{\infty} P(A_n) < \infty$ , aplicant el primer lema de Borel-Cantelli hom obté  $P(\limsup_n A_n) = 0$  o equivalentment,  $P(\liminf_n A_n^c) = 1$ . Per tot  $\omega \in \liminf_n A_n^c$  existeix un  $n_0(\omega)$  tal que  $\omega \in A_n^c$  per tot  $n \geq n_0(\omega)$ . Per tant, si  $n \geq n_0(\omega)$  tindrem que  $|X_{n+1}(\omega) - X_n(\omega)| \leq \varepsilon_n$ . És a dir. la sèrie  $\sum_{n=1}^{\infty} [X_{n+1}(\omega) - X_n(\omega)]$  convergeix absolutament. D'aquí se'n dedueix que la successió

$$X_n(\omega) = X_1(\omega) + \sum_{k=1}^{n-1} [X_{k+1}(\omega) - X_k(\omega)]$$

és convergent. ■

### 9.3. Convergència en probabilitat de variables aleatòries

**Definició 9.8.** Direm que una successió de variables aleatòries  $\{X_n, n \geq 1\}$  convergeix en probabilitat cap a una variable aleatòria  $X$  si per tot  $\varepsilon > 0$

$$\lim_{n \rightarrow \infty} P(|X_n - X| > \varepsilon) = 0.$$

Escriurem  $P - \lim X_n = X$ .

Intuitivament la definició anterior expressa que, en augmentar  $n$  és cada vegada menys probable que  $X_n$  i  $X$  difereixin en més de  $\varepsilon$ , per tot  $\varepsilon > 0$ .

La convergència en probabilitat és un cas particular de la convergència en mesura.

**Exemple 2.** Sigui  $\Omega = (0, 1]$ ,  $\mathcal{A} = \mathcal{B}((0, 1])$  i  $P$  la mesura de Lebesgue. Per tot  $n \geq 1$ ,  $1 \leq m \leq n$  definim

$$f_{n,m}(x) = \begin{cases} 1, & \text{si } \frac{m-1}{n} < x \leq \frac{m}{n}, \\ 0, & \text{en cas contrari.} \end{cases}$$

La successió de variables aleatòries  $\{f_{n,m}, n \geq 1, 1 \leq m \leq n\}$  convergeix en probabilitat cap a 0. En efecte:

$$P(|f_{n,m}| > \delta) = \begin{cases} 0, & \text{si } \delta > 1, \\ P\{x : f_{n,m}(x) = 1\}, & \text{si } \delta \leq 1. \end{cases}$$

La convergència cap a zero es dedueix del fet que  $P\{x : f_{n,m}(x) = 1\} = \frac{1}{n}$ .

La convergència en probabilitat és metrizable. Abans de demostrar aquesta propietat introdúrem alguna notació.

Recordem que  $L^0$  designa l'espai de classes d'equivalència de variables aleatòries finites, mòdul la relació

$$X \sim Y \iff X = Y \quad \text{q.s.}$$

Definim l'aplicació  $d : L^0 \times L^0 \rightarrow \mathbb{R}$  mitjançant  $d(X, Y) = E\left(\frac{|X - Y|}{1 + |X - Y|}\right)$ .

**Proposició 9.9.** L'aplicació  $d$  definida anteriorment és una distància en  $L^0$  que metritz la convergència en probabilitat.

*DemOSTRaciÓ:* Provarrem en primer lloc que  $d$  és una distància. És clar que  $d$  és simètrica i que  $d(X, Y) = 0$  si i únicament si  $X = Y$ . Per tant només cal demostrar la desigualtat triangular.

Donats nombres reals  $x, y, z$  és immediat comprovar que es compleix la següent desigualtat

$$\frac{|x + y|}{1 + |x + y|} \leq \frac{|x|}{1 + |x|} + \frac{|y|}{1 + |y|}.$$

Preneint  $x = X - Y$  i  $y = Y - Z$  (per tant  $x + y = X - Z$ ) i integrant respecte de  $P$  resulta

$$d(X, Z) \leq d(X, Y) + d(Y, Z),$$

com volíem demostrar.

Veiem ara que  $d$  metritz la convergència en probabilitat. Considerem una successió de variables aleatòries  $\{X_n, n \geq 1\}$  que convergeix en probabilitat cap a una variable aleatòria  $X$ . Sense pèrduda de generalitat suposem que  $X = 0$ . Donat  $\varepsilon > 0$ , sigui  $n_0 \in \mathbb{N}$  tal que per tot  $n \geq n_0$

$$P\{|X_n| > \varepsilon\} < \varepsilon.$$

Aleshores, si designem per  $A_n^\varepsilon$  al conjunt  $\left\{\frac{|X_n|}{1 + |X_n|} > \frac{\varepsilon}{1 + \varepsilon}\right\}$  tindrem

$$E\left(\frac{|X_n|}{1 + |X_n|}\right) = E\left(\frac{|X_n|}{1 + |X_n|} 1_{A_n^\varepsilon}\right) + E\left(\frac{|X_n|}{1 + |X_n|} 1_{A_n^\varepsilon c}\right) \leq 2\varepsilon.$$

Per tant

$$d(X_n, 0) \longrightarrow 0.$$

quan  $n$  tendeix a infinit.

Recíprocament, amb les notacions introduïdes anteriorment tenim la següent desigualtat

$$E\left(\frac{|X_n|}{1 + |X_n|}\right) \geq E\left(\frac{|X_n|}{1 + |X_n|} 1_{A_n^\varepsilon}\right) > \frac{\varepsilon}{1 + \varepsilon} P(A_n^\varepsilon).$$

D'on en resulta

$$P\{|X_n| > \varepsilon\} = P(A_n^\varepsilon) < \frac{1 + \varepsilon}{\varepsilon} E\left(\frac{|X_n|}{1 + |X_n|}\right).$$

Aquesta darrera expressió tendeix a zero quan  $n$  tendeix a infinit si suposem que  $\lim_{n \rightarrow \infty} d(X_n, 0) = 0$ . ■

De la proposició anterior se'n dedueixen els següents resultats:

- (1) El límit en probabilitat d'una successió de variables aleatòries, si existeix, és únic.
- (2) Si una successió de variables aleatòries  $\{X_n, n \geq 1\}$  convergeix en probabilitat, aleshores és una successió de Cauchy en probabilitat, és a dir  $P - \lim_{m,n \rightarrow \infty} (X_m - X_n) = 0$ .

L'espai  $L^0$  amb la convergència en probabilitat és un espai mètric *complet*, és a dir, tota successió de Cauchy en probabilitat té límit. Aquest resultat es podrà demostrar fàcilment una vegada s'hagi relacionat la convergència en probabilitat amb la convergència quasi segura.

**Proposició 9.10.** Sigui  $f : \mathbb{R}^2 \rightarrow \mathbb{R}$  una funció contínua, i sigui  $\{X_n, n \geq 1\}$ ,  $\{Y_n, n \geq 1\}$  seqüències successives de variables aleatòries que convergeixen en probabilitat cap a les variables aleatòries  $X$  i  $Y$  respectivament. Aleshores, la successió  $\{f(X_n, Y_n), n \geq 1\}$  convergeix en probabilitat cap a  $f(X, Y)$ .

*DemOSTRaciÓ:* Fixem  $\varepsilon > 0$ .  $\eta > 0$ . És fàcil comprovar que existeix  $k > 0$  tal que  $P\{|X| > \frac{k}{2}\} \leq \eta$  i  $P\{|Y| > \frac{k}{2}\} \leq \eta$ . Com que  $f$  és uniformement contínua en el compacte  $[-k, k]^2$ , existirà  $\delta > 0$  tal que si  $|x'|, |x|, |y'|, |y| \leq k$ ,  $|x - x'| \leq \delta$ ,  $|y - y'| \leq \delta$ ,

aleshores  $|f(x, y) - f(x', y')| \leq \varepsilon$ . Podrem doncs escriure

$$P\{|f(X_n, Y_n) - f(X, Y)| > \varepsilon\} \leq P\{|X| > \frac{k}{2}\}$$

$$\begin{aligned} &+ P\left\{|Y| > \frac{k}{2}\right\} + P\left\{|X - X_n| > \frac{k}{2}\right\} \\ &+ P\left\{|Y - Y_n| > \frac{k}{2}\right\} + P\left\{|X| \leq \frac{k}{2}\right\}, \\ |Y| \leq k, |X_n| \leq k, |Y_n| \leq k, &|f(X_n, Y_n) - f(X, Y)| > \varepsilon\} \\ \leq 2\eta + P\left\{|X - X_n| > \frac{k}{2}\right\} + P\left\{|Y - Y_n| > \frac{k}{2}\right\} \\ &+ P\left\{|X - X_n| > \delta\right\} + P\left\{|Y - Y_n| > \delta\right\}. \end{aligned}$$

Fent tendir  $n$  cap a infinit i tenint en compte que  $\eta > 0$  és arbitrari, obtindrem la convergència desitjada. ■

#### 9.4. Convergència en mitjana d'ordre $p$

**Definició 9.11.** Sigui  $\{X_n, n \geq 1\}$  una successió de variables aleatòries de l'espai  $L^p(\Omega, \mathcal{A}, P)$ , amb  $1 \leq p < \infty$ . Direm que aquesta successió convergeix en mitjana d'ordre  $p$  cap a una variable aleatòria  $X$  de  $L^p(\Omega, \mathcal{A}, P)$ , si  $X_n$  convergeix cap a  $X$  en la norma d'aquest espai, és a dir

$$\lim_{n \rightarrow \infty} E(|X_n - X|^p) = 0.$$

Escriurem  $L^p - \lim X_n = X$ .

**Exemple 3.** Considerem l'espai de probabilitat de l'exemple 1. Donat un nombre natural  $n \geq 1$  existeixen dos únics naturals  $p$  i  $q$  tals que  $n = 2^p + q$ ,  $p > 0$ ,  $0 \leq q < 2^p$ . Considerem la successió de variables aleatòries definida per  $X_n = 1_{[q2^{-p}, (q+1)2^{-p}]}$ . Per tot  $p \geq 1$ ,  $\{X_n, n \geq 1\}$  convergeix en mitjana d'ordre  $p$  cap a 0. En efecte,

$$E(|X_n|^p) = \frac{1}{2^p},$$

que tendeix a zero quan  $n$  tendeix a infinit.

#### 9.5. Relacions entre els diferents tipus de convergència

Torrem a l'exemple 2. És fàcil veure que la successió de variables aleatòries que allí proposavem no convergeix quasi segurament cap a zero. De fet, per tot  $x$  fix de  $[0, 1]$  la successió  $\{f_{nm}(x), n \geq 1, 1 \leq m \leq n\}$  conté infinitos zeros i infinitos uns. Resulta doncs que la convergència en probabilitat no implica la convergència quasi segura. D'altra banda, en l'exemple 3 hem vist que hi ha convergència en mitjana d'ordre  $p$ , per tot  $p \geq 1$ , però en canvi no hi ha convergència quasi segura. En efecte, fixat un element  $x \in [0, 1]$  existeixen infinitos intervals del tipus  $[q2^{-p}, (q+1)2^{-p}]$  que el contenen i també infinitos que no el contenen. És a dir  $\limsup_n X_n(x) = 1$  i  $\liminf_n X_n(x) = 0$ . Per tant la successió  $X_n$  no convergeix en cap punt.

Tampoc és cert que la convergència quasi segura impliqui la convergència en mitjana d'ordre  $p$  per algun  $p \geq 1$ . En efecte, considerant novament el mateix espai de probabilitat que en l'Exemple 3, la successió  $\{X_n, n \geq 1\}$  definida per  $X_n = 2^n 1_{[0, \frac{1}{n}]}$  convergeix quasi segurament cap a zero, en canvi

$$E(|X_n|^p) = \frac{2^{np}}{n},$$

que tendeix a infinit quan  $n$  tendeix a infinit, qualsevol que sigui  $p \geq 1$ .

Donem tot seguit resultats sobre les possibles relacions entre els tres tipus de convergències introduïdes en els apartats anteriors.

**Teorema 9.12.** Sigui  $\{X_n, n \geq 1\}$  una successió de variables aleatòries i  $X$  una variable aleatòria.

- (i) Si  $X_n \rightarrow X$  q.s. quan  $n$  tendeix a infinit, aleshores  $P - \lim_{n \rightarrow \infty} X_n = X$ .
- (ii) Si  $P - \lim_{n \rightarrow \infty} X_n = X$ , existeix una subsuccessió  $\{X_{n_i}, i \geq 1\}$  que convergeix quasi segurament cap a la variable  $X$  quan  $i$  tendeix a infinit.

*DemOSTRaciÓ:*

- (i) Fixem  $\varepsilon > 0$ . Sabem que amb probabilitat 1 es compleix la desigualtat  $|X_n(\omega) - X(\omega)| \leq \varepsilon$  per  $n \geq n_0(\omega)$ . És a dir  $P(\liminf_n \{|X_n - X| \leq \varepsilon\}) = 1$ . Aleshores aplicant el Lema 9.2 obtenim

$$\limsup_n P\{|X_n - X| > \varepsilon\} \leq P(\limsup_n \{|X_n - X| > \varepsilon\}) = 0.$$

i, en conseqüència  $P - \lim_{n \rightarrow \infty} X_n = X$ .

(ii) Recíprocament, suposem que  $P - \lim_{n \rightarrow \infty} X_n = X$ . Considerem dues sèries de termes positius  $\sum_{n=1}^{\infty} \varepsilon_n < \infty$  i  $\sum_{n=1}^{\infty} \delta_n < \infty$ . Recordem que es compleix la condició de Cauchy  $\lim_{m,n \rightarrow \infty} P\{|X_n - X_m| > \varepsilon\} = 0$ , per tot  $\varepsilon > 0$ .

Definirem una successió estrictament creixent de nombres naturals de la manera següent. Posem  $n_0 = 0$  i per tot  $k \geq 1$  sigui  $n_k$  el primer natural més gran que  $n_{k-1}$  que satisfa la condició

$$P\{|X_n - X_{n_k}| > \varepsilon_k\} < \delta_k,$$

per tot  $n, m \geq n_k$ .

Aleshores

$$\sum_{k=1}^{\infty} P\{|X_{n_{k+1}} - X_{n_k}| > \varepsilon_k\} < \sum_{k=1}^{\infty} \delta_k < \infty.$$

Utilitzant el criteri demostrat en la Proposició 9.7 resulta que la successió parcial  $\{X_{n_k}, k \geq 1\}$  convergeix quasi segurament. El límit haurà d'ésser necessàriament la variable  $X$ , donat que la convergència quasi segura implica la convergència en probabilitat (tal i com hem demostrat a l'apartat (i)) i sabem que  $P - \lim_{k \rightarrow \infty} X_{n_k} = X$ .

La Proposició queda totalment demostrada. ■

**Proposició 9.14.** L'espai  $L^0$  amb la convergència en probabilitat és un espai mètric complet.

*DemOSTRaciÓ:* Sigui  $\{X_n, n \geq 1\}$  una successió de Cauchy en probabilitat. Existeix una subsuccessió  $\{X_{n_k}, k \geq 1\}$  que tendeix a un cert límit  $X$  en la convergència quasi segura. Resulta doncs

$$\begin{aligned} P\{|X_n - X| > \varepsilon\} &\leq P\{|X_n - X_{n_k}| \geq \frac{\varepsilon}{2}\} \\ &\quad + P\{|X_{n_k} - X| \geq \frac{\varepsilon}{2}\}. \end{aligned}$$

Com que  $\{X_n, n \geq 1\}$  és una successió de Cauchy en probabilitat, existeix un  $n_0$  natural tal que si  $n, n_k \geq n_0$  es té  $P\{|X_n - X_{n_k}| \geq \frac{\varepsilon}{2}\} \leq \frac{\varepsilon}{2}$ . D'altra banda, com que  $\{X_{n_k}, k \geq 1\}$  convergeix en probabilitat cap a  $X$ , existeix un  $n_1$  natural tal que per tot  $n_k \geq n_1$  es compleix  $P\{|X_{n_k} - X| \geq \frac{\varepsilon}{2}\} \leq \frac{\varepsilon}{2}$ . Aleshores, prenent  $N_0 = \max(n_0, n_1)$  tindrem  $P\{|X_n - X| \geq \varepsilon\} \leq \varepsilon$ , per tot  $n \geq N_0$ . ■

Com a conseqüència immediata de la desigualtat de Txebixef (Proposició 9.1) es té que la convergència en mitjana d'ordre  $p$ , per a qualsevol  $p \geq 1$ , implica la convergència en probabilitat.

Sota condicions de dominació per una variable de  $L^p$ , la convergència quasi segura implica la convergència en mitjana d'ordre  $p$ . Tenim concretament el següent resultat

**Proposició 9.14.** Sigui  $p \geq 1$  i  $\{X_n, n \geq 1\}$  una successió de variables aleatòries que convergeix quasi segurament cap a una variable aleatòria  $X$ . Suposem que  $|X_n|^p \leq Y$ , amb  $Y \in L^1(\Omega, \mathcal{A}, P)$ . Aleshores  $L^p - \lim_{n \rightarrow \infty} X_n = X$ .

*DemOSTRaciÓ:* Demostrarem primer, utilitzant el lema de Fatou, que  $X \in L^p(\Omega, \mathcal{A}, P)$ . En efecte:

$$E(|X|^p) \leq \liminf_{n \rightarrow \infty} E(|X_n|^p) \leq E(Y) < \infty.$$

La successió de variables aleatòries  $|X_n - X|^p$  tendeix cap a zero quasi segurament i està dominada per una variable aleatòria integrable, car

$$|X_n - X|^p \leq (|X_n| + |X|)^p \leq 2^{p-1} (|X_n|^p + |X|^p) \leq 2^{p-1} (Y + |X|^p).$$

Aleshores, aplicant el teorema de la convergència dominada de Lebesgue se'n dedueix que  $\lim_{n \rightarrow \infty} E(|X_n - X|^p) = 0$ , que és el que volem demostrar. ■

on hem utilitzat la igualtat  $E\left(\frac{S_n}{n}\right) = m$  i el fet que la variància d'una suma de v.a. independents és igual a la suma de les seves variàncies. ■

Tot seguit, i com aplicació de la llei feble dels grans nombres que acabem de veure, demostrarem que tota funció contínua  $f : [0, 1] \rightarrow \mathbb{R}$  es pot aproximar uniformement per polinomis. Es a dir, donarem una demostració del teorema de Stone–Weierstrass que es basa en tècniques probabilístiques.

Considerem una funció contínua  $f : [0, 1] \rightarrow \mathbb{R}$ . Per a cada nombre  $x \in [0, 1]$  considerem una successió  $\{X_n, n \geq 1\}$  de variables aleatories i.i.d. amb llei de Bernoulli de paràmetre  $x$ , definides en un cert espai de probabilitat  $(\Omega, \mathcal{A}, P)$ . Posem  $S_n = X_1 + \dots + X_n$ . Sabem que  $S_n$  té una llei binomial  $B(n, x)$ , és a dir,

$$P\{S_n = k\} = \binom{n}{k} x^k (1-x)^{n-k}, \quad k = 0, 1, \dots, n.$$

d'una successió  $\{X_n, n \geq 1\}$  de variables aleatories independents. Sota certes condicions, la llei feble (forta) ens dirà que la successió  $\frac{S_n}{n}$  convergeix en probabilitat (quasi segurament) cap una constant. El fet de que el límit sigui constant era d'esperar degut a la llei del 0–1 de Kolmogorov (observis que  $\frac{S_n}{n}$  té el mateix límit que  $\frac{1}{n}(X_1 + \dots + X_n)$ ,  $n \geq k$ ).

Si suposem que les variables  $X_n$  són idènticament distribuïdes, aquesta constant serà precisament la seva esperança.

Considerem el següent exemple fonamental: Les variables  $X_n$  són independents i amb llei de Bernoulli de paràmetre  $p$ . Això vol dir que  $X_n$  val 1 ó 0 segons que un cert esdeveniment A s'hagi realitzat o no a l'instant  $n$ , en una successió de realitzacions independents d'una experiència aleatòria. Aleshores, el quotient  $\frac{S_n}{n}$  representa la freqüència relativa d'aquest esdeveniment en les  $n$  primeres realitzacions de l'experiència i les lleis dels grans nombres ens diuen que la successió de les freqüències relatives convergeix (quasi segurament i, per tant, també en probabilitat) cap a la probabilitat de l'esdeveniment que estem considerant (igual a la constant  $p$ ). Aquest és l'anomenat teorema de Bernoulli (en el cas de la convergència en probabilitat) i es troba a la seva obra "Ars Conjectandi" (1713).

Establirem a continuació, la llei feble dels grans nombres per a variables aleatories de quadrat integrable, idènticament distribuïdes. Recordi's que la convergència en  $L^2$  implica la convergència en probabilitat.

**Proposició 10.1.** Sigui  $\{X_n, n \geq 1\}$  una successió de variables aleatories independents i idènticament distribuïdes (i.i.d.). Suposem  $E(X_1) = m$  i  $E(X_1^2) < \infty$ . Aleshores,

$$\frac{S_n}{n} \xrightarrow{L_2} m.$$

*Demostració.* Tenim

$$E\left(\left|\frac{S_n}{n} - m\right|^2\right) = \sigma^2\left(\frac{S_n}{n}\right) = \frac{1}{n^2}\sigma^2(S_n) = \frac{1}{n}\sigma^2(X_1) \xrightarrow{n \rightarrow \infty} 0,$$

## 10. LLEIS DELS GRANS NOMBRES

En aquest apartat farem servir els diversos tipus de convergència analitzats en l'apartat 9 per tal d'estudiar les anomenades "Lleis dels grans nombres". Les lleis dels grans nombres es refereixen al comportament asymptòtic de la successió de sumes parcials

$$S_n = X_1 + \dots + X_n$$

d'una successió  $\{X_n, n \geq 1\}$  de variables aleatories independents. Sota certes condicions, la llei feble (forta) ens dirà que la successió  $\frac{S_n}{n}$  convergeix en probabilitat (quasi segurament) cap una constant. El fet de que el límit sigui constant era d'esperar degut a la llei del 0–1 de Kolmogorov (observis que  $\frac{S_n}{n}$  té el mateix límit que  $\frac{1}{n}(X_1 + \dots + X_n)$ ,  $n \geq k$ ). Si suposem que les variables  $X_n$  són idènticament distribuïdes, aquesta constant serà precisament la seva esperança.

Considerem el següent exemple fonamental: Les variables  $X_n$  són independents i amb llei de Bernoulli de paràmetre  $p$ . Això vol dir que  $X_n$  val 1 ó 0 segons que un cert esdeveniment A s'hagi realitzat o no a l'instant  $n$ , en una successió de realitzacions independents d'una experiència aleatòria. Aleshores, el quotient  $\frac{S_n}{n}$  representa la freqüència relativa d'aquest esdeveniment en les  $n$  primeres realitzacions de l'experiència i les lleis dels grans nombres ens diuen que la successió de les freqüències relatives convergeix (quasi segurament i, per tant, també en probabilitat) cap a la probabilitat de l'esdeveniment que estem considerant (igual a la constant  $p$ ). Aquest és l'anomenat teorema de Bernoulli (en el cas de la convergència en probabilitat) i es troba a la seva obra "Ars Conjectandi" (1713).

Establirem a continuació, la llei feble dels grans nombres per a variables aleatories de quadrat integrable, idènticament distribuïdes. Recordi's que la convergència en  $L^2$  implica la convergència en probabilitat.

$$\begin{aligned} E\left(\left|f(x) - f\left(\frac{S_n}{n}\right)\right|\right) &= \int_{\{|(\frac{S_n}{n}-x)|>\delta\}} |f\left(\frac{S_n}{n}\right) - f(x)| dP \\ &\quad + \int_{\{|(\frac{S_n}{n}-x)|\leq\delta\}} |f\left(\frac{S_n}{n}\right) - f(x)| dP \\ &\leq 2\|f\|_\infty P\left\{\left|\frac{S_n}{n} - x\right| > \delta\right\} + \varepsilon \\ &\leq 2\|f\|_\infty \frac{1}{\delta^2} \sigma^2\left(\frac{S_n}{n}\right) + \varepsilon \\ &= 2\|f\|_\infty \frac{1}{\delta^2} \frac{x(1-x)}{n} + \varepsilon \leq \frac{\|f\|_\infty}{2\delta^2} + \varepsilon. \end{aligned}$$

En conseqüència,

$$\limsup_{n \rightarrow \infty} \sup_{x \in [0,1]} |f(x) - p_n(x)| \leq \varepsilon,$$

i com que  $\varepsilon$  és arbitrari, obtenim el resultat que buscavem. ■

Per tal de preparar la demostració de la llei fortta dels grans nombres establimos alguns resultats previs. En primer lloc demostrarem la següent desigualtat.

**Proposició 10.2.** (Desigualtat de Kolmogorov). Sigui  $X_1, \dots, X_n$  variables aleatòries independents, de quadrat integrable i centrades. Posem  $S_k = X_1 + \dots + X_k$ ,  $1 \leq k \leq n$ . Aleshores, per a tot  $\varepsilon > 0$  tenim

$$P\left\{\max_{1 \leq k \leq n} |S_k| > \varepsilon\right\} \leq \frac{\sigma^2(S_n)}{\varepsilon^2}.$$

**Demostració:** Observi's en primer lloc que per  $n = 1$  el resultat es redueix a la desigualtat de Tchebychev.

Fixem  $\varepsilon > 0$  i posem  $A = \{\max_{1 \leq k \leq n} |S_k| > \varepsilon\}$ . Definim

$$\begin{aligned} A_1 &= \{|S_1| > \varepsilon\}, & i \\ A_k &= \left\{\max_{1 \leq j \leq k-1} |S_j| \leq \varepsilon, |S_k| > \varepsilon\right\} \quad \text{per } 2 \leq k \leq n. \end{aligned}$$

Els conjunts  $A_k$  són mesurables, disjunts dos a dos i  $A = A_1 \cup A_2 \cup \dots \cup A_n$ .

Podem interpretar aquesta descomposició que hem fet del conjunt  $A$  de la forma següent: En el conjunt  $A$ , algun dels nombres  $|S_1|, |S_2|, \dots, |S_n|$  és més gran que  $\varepsilon$ , aleshores  $A_k$  representa el conjunt on el primer d'aquests nombres que és més gran que  $\varepsilon$  és precisament  $|S_k|$ .

Tenim

$$\begin{aligned} \sigma^2(S_n) &= \int_A S_n^2 dP \geq \sum_{k=1}^n \int_{A_k} S_n^2 dP \\ &= \sum_{k=1}^n \int_{A_k} [S_k + (S_n - S_k)]^2 dP \geq \sum_{k=1}^n \int_{A_k} [S_k^2 + 2S_k(S_n - S_k)] dP. \end{aligned} \quad (10.2)$$

D'altra banda

$$\int_{A_k} S_k(S_n - S_k) dP = E[1_{A_k} S_k(S_n - S_k)] = E(1_{A_k} S_k)E(S_n - S_k) = 0.$$

En efecte, les variables  $1_{A_k} S_k : S_n - S_k$  són independents ja que  $S_n - S_k = X_{k+1} + \dots + X_n$  i  $1_{A_k} S_k$  és una funció mesurable del vector aleatori  $(X_1, \dots, X_n)$ .

Finalment, com que  $S_k^2$  és més gran que  $\varepsilon^2$  en el conjunt  $A_k$ , a partir de (10.2) obtenim

$$\sigma^2(S_n) \geq \sum_{k=1}^n \int_{A_k} S_k^2 dP \geq \varepsilon^2 \sum_{k=1}^n P(A_k) = \varepsilon^2 P(A). \blacksquare$$

Utilitzant la desigualtat de Kolmogorov podem establir el següent criteri de convergència quasi segura per a sèries aleatories:

**Proposició 10.3.** Sigui  $\{X_n, n \geq 1\}$  una successió de v.a. independents, de quadrat integrable i centrades. Aleshores, si  $\sum_{n=1}^{\infty} \sigma^2(X_n) < \infty$ , la sèrie  $\sum_{n=1}^{\infty} X_n$  convergeix quasi segurament.

**Demostració:** Aplicant la desigualtat de Kolmogorov a les variables  $X_{m+1}, \dots, X_{m+k}$  obtenim

$$\begin{aligned} P\left\{\sup_{j \geq 1} |S_{m+j} - S_m| > \varepsilon\right\} &= \lim_{k \rightarrow \infty} P\left\{\sup_{1 \leq j \leq k} |S_{m+j} - S_m| > \varepsilon\right\} \\ &\leq \lim_{k \rightarrow \infty} \frac{1}{\varepsilon^2} \sum_{j=1}^k \sigma^2(X_{m+j}) = \frac{1}{\varepsilon^2} \sum_{j=1}^{\infty} \sigma^2(X_{m+j}). \end{aligned}$$

Aquesta darrera expressió tendeix a zero quan  $m$  tendeix a infinit, i això implica la convergència quasi segura de la successió  $S_n$ . En efecte, la convergència en probabilitat

$$\sup_{j \geq 1} |S_{m+j} - S_m| \xrightarrow{n \rightarrow \infty} 0$$

implica l'existència d'una successió estrictament creixent de nombres naturals  $m_i$  tal que  $\sup_{j \geq 1} |S_{m_i+j} - S_{m_i}| \xrightarrow{j \rightarrow \infty} 0$ , q.s. i, per tant, la successió  $S_n$ ,  $n \geq 1$  és de Cauchy amb probabilitat 1. ■

**Exemple:** És ben sabut que la sèrie harmònica  $\sum_{n=1}^{\infty} \frac{1}{n}$  és divergent mentre que la sèrie alternada  $\sum_{n=1}^{\infty} \frac{(-1)^n}{n}$  convergeix. Considerem una successió  $\{\varepsilon_n, n \geq 1\}$  de v.a. i.i.d. tals que  $P\{\varepsilon_n = 1\} = P\{\varepsilon_n = -1\} = 1/2$ . Aleshores, la sèrie aleatòria  $\sum_{n=1}^{\infty} \frac{\varepsilon_n}{n^p}$  convergeix q.s. si  $p > 1/2$  ja que la sèrie de variàncies és igual a  $\sum_{n=1}^{\infty} \frac{1}{n^{2p}} < \infty$ . Observi's que per  $1/2 < p \leq 1$  aquesta sèrie aleatòria no convergeix absolutament. Necessitarem també el següent resultat tècnic.

**Lema 10.4.** (Lema de Kronecker). Sigui  $\{x_n, n \geq 1\}$  una successió de nombres reals i  $\{a_n, n \geq 1\}$  una altra successió de nombres reals tal que  $0 < a_n \uparrow \infty$ . Aleshores, si

ja que el primer sumand té límit zero quan  $n \rightarrow \infty$ . Com que  $\varepsilon$  és arbitrari obtenim el resultat desitjat. ■

El lema de Kronecker i el criteri de convergència q.s. per sèries aleatories ens permeten demostrar la següent versió de la llei forta dels grans nombres per a variables de quadrat integrable, no necessàriament idènticament distribuïdes.

*Demostració:* Per a cada  $n \geq 1$  posem  $b_n = \sum_{j=1}^n \frac{x_j}{a_j}$ . També escrivim, per conveni,

$$\begin{aligned} a_0 &= b_0 = 0. Llavors, x_n = a_n(b_n - b_{n-1}), \text{ per } n \geq 1, \text{ i mitjançant el mètode de sumació parcial d'Abel, obtenim} \\ &\frac{1}{a_n} \sum_{j=1}^n x_j = \frac{1}{a_n} \sum_{j=1}^n a_j(b_j - b_{j-1}) = \\ &= \frac{1}{a_n} (a_1 b_1 + a_2 b_2 - a_2 b_1 + a_3 b_3 - a_3 b_2 + \dots \\ &+ a_{n-1} b_{n-1} - a_{n-1} b_{n-2} + a_n b_n - a_n b_{n-1}) \\ &= b_n - \frac{1}{a_n} \sum_{j=0}^{n-1} b_j(a_{j+1} - a_j). \end{aligned}$$

Sigui  $b_\infty = \lim_n b_n$ . Aleshores, hem de veure que

$$\lim_n \frac{1}{a_n} \sum_{j=0}^{n-1} b_j(a_{j+1} - a_j) = b_\infty.$$

Observen que  $\frac{1}{a_n} \sum_{j=0}^{n-1} (a_{j+1} - a_j) = 1$ . En conseqüència,

$$\begin{aligned} \left| \frac{1}{a_n} \sum_{j=0}^{n-1} b_j(a_{j+1} - a_j) - b_\infty \right| &= \left| \frac{1}{a_n} \sum_{j=0}^{n-1} (b_j - b_\infty)(a_{j+1} - a_j) \right| \\ &\leq \frac{1}{a_n} \left| \sum_{j=0}^{m-1} (b_j - b_\infty)(a_{j+1} - a_j) \right| \\ &\quad + \frac{1}{a_n} \left| \sum_{j=m}^{n-1} (b_j - b_\infty)(a_{j+1} - a_j) \right|, \end{aligned}$$

per  $n > m$ .

Fixat un  $\varepsilon > 0$ , sigui  $m$  tal que  $|b_j - b_\infty| < \varepsilon$ , per a tot  $j \geq m$ . Tenint en compte que  $a_{j+1} - a_j \geq 0$ , el segon sumand de l'expressió anterior estarà afiat per  $\varepsilon$  per aquest valor de  $m$ . Per tant, si fixem  $\varepsilon \leq m$ , tindrem

$$\limsup_n \left| \frac{1}{a_n} \sum_{j=0}^{n-1} b_j(a_{j+1} - a_j) - b_\infty \right| \leq \varepsilon.$$

■

**Teorema 10.5.** Sigui  $\{X_n, n \geq 1\}$  una successió de v.a. independents, que quadrat integrable i centrades. Sigui  $\{a_n, n \geq 1\}$  una successió de nombres reals tal que  $0 < a_n \uparrow \infty$ .

Aleshores si  $\sum_{n=1}^{\infty} \frac{\sigma^2(X_n)}{a_n^2} < \infty$ , es compleix

$$\frac{S_n}{a_n} \rightarrow 0,$$

quasi segurament, quan  $n$  tendeix a infinit.

*Demostració:* Tenim

$$\sum_{n=1}^{\infty} \sigma^2\left(\frac{X_n}{a_n}\right) = \sum_{n=1}^{\infty} \frac{\sigma^2(X_n)}{a_n^2} < \infty.$$

La proposició 10.3 aplicada a la successió  $\frac{X_n}{a_n}$  ens diu que la sèrie  $\sum_{n=1}^{\infty} \frac{X_n}{a_n}$  convergeix q.s. i pel lema de Kronecker obtenim

$$\lim_n \frac{1}{a_n} \sum_{j=1}^n X_j = 0, \text{ q.s.}$$

Si apliquem el resultat anterior a la successió definida per  $a_n = n$ , tenim que la condició  $\sum_{n=1}^{\infty} \frac{\sigma^2(X_n)}{n^2} < \infty$  és suficient per a assegurar la convergència

$$\frac{S_n}{n} \rightarrow 0, \quad (10.3)$$

quan  $n$  tendeix a infinit.

En particular, si  $\{X_n, n \geq 1\}$  és una successió de variables aleatories i.i.d., de quadrat integrable, amb  $E(X_1) = m$  i  $\sigma^2(X_1) = \sigma^2$  obtenim

$$\lim_{n \rightarrow \infty} \frac{S_n}{n} = m, \quad (10.4)$$

quasi segurament.

En efecte, només cal aplicar (10.3) a les variables centrades  $X_n - m$  i observar que

$$\sum_{n=1}^{\infty} \frac{\sigma^2(X_n - m)}{n^2} = \sum_{n=1}^{\infty} \frac{\sigma^2}{n^2} < \infty.$$

Per tant, (10.4) ens dóna la llei forta dels grans nombres sota les mateixes hipòtesis que la Proposició 10.1 (variables i.i.d. i de quadrat integrable). A continuació veurem que aquest resultat es pot afegir. Més precisament, arribarem a la mateixa conclusió suposant únicament existència de moments de primer ordre.

Abans d'abordar aquest problema donarem un resultat previ.

**Lema 10.6.** Per a tota variable aleatòria  $Y$  es compleix

$$\sum_{n=1}^{\infty} P\{|Y| \geq n\} \leq E(|Y|) \leq 1 + \sum_{n=1}^{\infty} P\{|Y| \geq n\}.$$

Aquest lema ens diu que la variable  $Y$  és integrable si i només si la sèrie  $\sum_{n=1}^{\infty} P\{|Y| \geq n\}$  és convergent.

*Demostració:* Podem escriure

$$\begin{aligned} \sum_{n=1}^{\infty} P\{|Y| \geq n\} &= \sum_{n=1}^{\infty} \sum_{k=n}^{\infty} P\{k \leq |Y| < k+1\} \\ &= \sum_{k=1}^{\infty} k P\{k \leq |Y| < k+1\} = \sum_{k=0}^{\infty} k P\{k \leq |Y| < k+1\}. \end{aligned}$$

Aquesta darrera sèrie està majorada per  $\sum_{k=0}^{\infty} \int_{\{k \leq |Y| < k+1\}} |Y| dP = E(|Y|)$ , i obtenim així la primera desigualtat.

D'altra banda,

$$E(|Y|) = \sum_{k=0}^{\infty} \int_{\{k \leq |Y| < k+1\}} |Y| dP \leq \sum_{k=0}^{\infty} (k+1) P\{k \leq |Y| < k+1\} \leq \sum_{n=1}^{\infty} P\{|Y| \geq n\} + 1.$$

El següent resultat s'anomena llei forta de Kolmogorov i ens diu que l'hipòtesi de integritat de les variables és necessària i suficient per tal d'assegurar la convergència de la successió  $\left\{ \frac{S_n}{n}, n \geq 1 \right\}$ .

**Teorema 10.7.** Sigui  $\{X_n, n \geq 1\}$  una successió de variables aleatòries i.i.d. Aleshores,

- (i) Si  $E(|X_1|) < \infty$ , aleshores  $\lim_{n \rightarrow \infty} \frac{S_n}{n} = E(X_1)$ , q.s.
- (ii) Si  $E(|X_1|) = \infty$ , aleshores  $\limsup_n \frac{|S_n|}{n} = +\infty$ . q.s.

*Demostració:* La idea de la demostració consisteix en truncar les variables  $X_n$  posant el valor zero si estan fora de l'interval  $(-n, n)$  i aplicar després el Teorema 10.5.

Per a demostrar (i) definim  $Y_n = X_n \mathbf{1}_{\{|X_n| < n\}}$ . Fent servir el Lema 10.6, tenim

$$\sum_{n=1}^{\infty} P\{|X_n \neq Y_n\} \leq \sum_{n=1}^{\infty} P\{|X_n| \geq n\} = \sum_{n=1}^{\infty} P\{|X_i| \geq n\} \geq E(|X_1|) < \infty.$$

Pel lemma del Borel-Cantelli tindrem

$$P\left\{\limsup_n (X_n \neq Y_n)\right\} = 0,$$

és a dir, si  $G = \liminf_n \{X_n = Y_n\}$ , tindrem  $P(G) = 1$ .

Si  $\omega \in G$ , existeix un  $n_0(\omega)$  tal que  $X_n(\omega) = Y_n(\omega)$  per a tot  $n \geq n_0(\omega)$ . Per tant,

$$\frac{1}{n} \sum_{i=1}^n X_i(\omega) = \frac{1}{n} \sum_{i=1}^{n_0(\omega)} Y_i(\omega) + \frac{1}{n} \sum_{i=1}^{n-n_0(\omega)} (X_i - Y_i)(\omega),$$

per a tot  $n \geq n_0(\omega)$ ,  $\omega \in G$ .

Per tant és suficient demostrar que

$$\frac{1}{n} \sum_{i=1}^n Y_i \rightarrow E(X_1), \quad (10.5)$$

quan  $n$  tendeix a infinit, q.s.

Tenim

$$E(Y_n) = E(X_n \cdot \mathbf{1}_{\{|X_n| < n\}}) = E(X_1 \mathbf{1}_{\{|X_n| < n\}}) \rightarrow E(X_1),$$

quan  $n$  tendeix a infinit

Per tant

$$\frac{1}{n} \sum_{i=1}^n E(Y_i) \rightarrow E(X_1)$$

quan  $n$  tendeix a infinit.

En conseqüència, per provar (10.5) només ens caldrà demostrar que

$$\frac{1}{n} \sum_{i=1}^n (Y_i - E(Y_i)) \rightarrow 0, \quad (10.6)$$

q.s. quan  $n$  tendeix a infinit.

Tenint en compte el Teorema 10.5 (aplicat a la successió  $\{Y_n - E(Y_n), n \geq 1\}$  de variables independents, centrades i de quadrat integrable) (10.6) serà cert si  $\sum_{n=1}^{\infty} \frac{E(Y_n^2)}{n^2} < \infty$ .

Podem escriure

$$\begin{aligned} \sum_{n=1}^{\infty} \frac{\sigma^2(Y_n)}{n^2} &\leq \sum_{n=1}^{\infty} \frac{E(Y_n^2)}{n^2} = \sum_{n=1}^{\infty} \frac{1}{n^2} E(X_n^2 1_{\{|X_n|< n\}}) \\ &= \sum_{n=1}^{\infty} \frac{1}{n^2} E(X_1^2 1_{\{|X_1|< n\}}) = \sum_{n=1}^{\infty} \frac{1}{n^2} \sum_{k=1}^{\infty} E(X_1^2 1_{\{k-1 \leq |X_1| < k\}}) \\ &= \sum_{k=1}^{\infty} E(X_1^2 1_{\{k-1 \leq |X_1| < k\}}) \sum_{n=k}^{\infty} \frac{1}{n^2} \\ &\leq \sum_{k=1}^{\infty} k E(|X_1| 1_{\{k-1 \leq |X_1| < k\}}) \frac{2}{k} \\ &\leq 2E(|X_1|) < \infty. \end{aligned}$$

On hem fet servir l'affitació

$$\sum_{n=k}^{\infty} \frac{1}{n^2} \leq \frac{1}{k^2} + \int_k^{\infty} \frac{1}{x^2} dx = \frac{1}{k^2} + \frac{1}{k} \leq \frac{2}{k}.$$

Això acaba la demostració de l'apparatat (i).

Per a demostrar (ii) fent servir el Lema 10.6. Per a tota constant  $K > 0$ ,

$$\sum_{n=1}^{\infty} P\{|X_n| \geq Kn\} = \sum_{n=1}^{\infty} P\left\{\frac{|X_1|}{K} \geq n\right\} \leq E\left(\frac{|X_1|}{K}\right) - 1 = \infty.$$

El segon lema de Borel-Cantelli (cas independent) ens diu aleshores que  $P(G_K) = 1$ . on

$$G_K = \limsup \{|X_n| \geq Kn\}.$$

Posem  $G = \cap_{K=1}^{\infty} G_K$ . Clarament  $P(G) = 1$ .

Sobre el conjunt  $G$ , es compleix  $\frac{|X_n|}{n} \geq K$  per infinitis valors de  $n$ , per a tot natural  $K \geq 1$ .

Això implica que

$$\frac{|S_n| + \frac{|S_{n-1}|}{n}}{n-1} > \frac{|S_n| + |S_{n-1}|}{n} \geq \frac{|S_n - S_{n-1}|}{n} = \frac{|X_n|}{n} \geq K$$

per infinitis valors de  $n$ , és a dir  $\frac{|S_n|}{n} \geq \frac{K}{2}$  per infinitis valors de  $n$ , i en conseqüència,  $\limsup_n \frac{|S_n|}{n} = \infty$  en el conjunt  $G$ .

Amb això acaba la demostració del teorema. ■

**Exemple:** Els nombres normals de Borel.

Un cas interessant d'aplicació de la llei forta dels grans números apareix en l'estudi dels anomenats "nombres normals" de Borel (Borel, 1909).

Per a tot nombre  $x \in [0, 1]$  designarem per  $\xi_n(x)$  la  $n$ -èsima xifra decimal de  $x$ . Cal indicar que els nombres del conjunt  $M = [0, 1] \cap \{m10^{-n}, m, n \leq 0\}$  tenen dues expressions decimals diferents (per exemple,  $0,2 \text{ i } 0,19$ ) i cal escollir-ne una, per exemple, la que té zeros a partir d'un cert lloc.

Posem  $X_n^{(k)}(x) = 1_{\{X_n=k\}}(x)$  on  $k \in \{0, 1, 2, \dots, 9\}$  és una xifra que fixem.

Aleshores  $\frac{1}{n} \sum_{i=1}^n X_i^{(k)}(x)$  representa la freqüència relativa amb què la xifra  $k$  apareix en les  $n$  primeres xifres decimals de  $x$ .

Es diu aleshores que el nombre  $x$  és *normal* si

$$\frac{1}{n} \sum_{i=1}^n X_i^{(k)}(x) \rightarrow \frac{1}{10}, \quad (10.7)$$

quan  $n$  tendeix a infinit, per a tota xifra  $k$ .

No se sap si nombres iracionals ben coneguts com  $e - 2 \circ \pi - 3$  són normals, però en canvi el *teorema de Borel* ens diu que llevat per un subconjunt de  $[0, 1]$  de mesura de Lebesgue zero, tots els nombres són normals.

Donem tot seguit una demostració d'aquest resultat. Considerem l'espai de probabilitat  $([0, 1], \mathcal{B}([0, 1]), P)$  on  $P$  és la mesura de Lebesgue. En aquest espai,  $\{X_n, n \geq 1\}$  és una successió de variables aleatòries independents amb distribució uniforme sobre el conjunt  $\{0, 1, 2, \dots, 9\}$ . En efecte, fixem  $n$  xifres  $k_1, k_2, \dots, k_n$  i calculem  $P\{X_1 = k_1, \dots, X_n = k_n\}$ , que és la probabilitat del conjunts de  $x \in [0, 1]$  tals que les primeres  $n$  xifres decimals de  $x$  són  $k_1, k_2, \dots, k_n$ .

Obtenim

$$P\{X_1 = k_1, \dots, X_n = k_n\} = P\{[0, k_1 \dots k_n + 10^{-n})\} = 10^{-n}.$$

Les variables  $\{X_n^{(k)}, n \geq 1\}$  (fixada una xifra  $k$ ) també seran independents i amb lleis de Bernoulli de paràmetre  $P\{X_n^{(k)} = 1\} = P\{X_n = k\} = 10^{-1}$ . Per tant, la convergència (10.7) serà certa, per a tots els  $x \in [0, 1]$  excepte en un conjunt de mesura de Lebesgue zero, degut a la llei forta dels grans nombres. Queda així provat el resultat.

*Demostració:* Suposem primer que  $w = \lim_{n \rightarrow \infty} \mu_n = \mu$ . Fixem un punt  $x$  de continuïtat de la funció  $F$  i considerem les funcions contínues i afitades definides per

$$f_\varepsilon^+(y) = 1_{(-\infty, x)}(y) + \left(1 - \frac{y-x}{\varepsilon}\right) 1_{(x, x+\varepsilon)}(y),$$

$$f_\varepsilon^-(y) = 1_{(-\infty, x-\varepsilon)}(y) + \frac{x-y}{\varepsilon} 1_{(x-\varepsilon, x)}(y),$$

on  $\varepsilon$  és un nombre real estrictament positiu que fixem.

11.1. Convergència feble de probabilitats i convergència en llei d'una successió de variables aleatòries

El tipus de convergència que anem a introduir difereix de les convergències que hem estudiats fins ara (en probabilitat, quasi segura i en  $L^p$ ) en què es referix a una successió de probabilitats sobre  $\mathbb{R}$  en lloc d'una successió de variables aleatòries.

Considerem una successió  $\{\mu_n, n \geq 1\}$  de probabilitats sobre  $\mathbb{R}$ . Ens podem plantejar la qüestió següent: Quan direm que  $\mu_n$  convergeix cap a una probabilitat  $\mu$ ? La definició més natural seria dir que  $\lim_{n \rightarrow \infty} \mu_n(B) = \mu(B)$  per a tot  $B \in \mathcal{B}(\mathbb{R})$ . Ara bé aquesta definició és clarament inadequada pels nostres objectius. En efecte, ens interessarà en alguns casos (per exemple en l'aproximació de la llei binomial per la llei normal, o en l'aproximació de la llei geomètrica per una exponencial) poder dir que una successió de probabilitats discretes  $\mu_n$  convergeix cap a una llei contínua  $\mu$ . Això no és possible amb la definició anterior ja que si  $S$  és el conjunt numerable igual a la unió dels suports de les  $\mu_n$ , tindrem  $\mu_n(S) = 1$  per a tot  $n$  però  $\mu(S) = 0$ .

**Definició 11.1.1.** Sigui  $\{\mu_n, n \geq 1\}$  una successió de probabilitats en  $\mathbb{R}$ . Direm que aquesta successió *convergeix feblement* cap a una probabilitat  $\mu$  i escriurem  $w - \lim_{n \rightarrow \infty} \mu_n = \mu$ .

$$\lim_n \int_{\mathbb{R}} f d\mu_n = \int_{\mathbb{R}} f d\mu,$$

per a tota funció contínua i afitada  $f : \mathbb{R} \rightarrow \mathbb{R}$ .

Al llarg d'aquest capítol designarem per  $C_b(\mathbb{R})$  el conjunt de les funcions  $f : \mathbb{R} \rightarrow \mathbb{R}$  contínues i afitades.

El següent resultat caracteritza la convergència feble en termes de les funcions de distribució  $F_n$  i  $F$  de  $\mu_n$  i  $\mu$ , respectivament.

**Teorema 11.1.2.** Una successió  $\{\mu_n, n \geq 1\}$  convergeix feblement cap a una probabilitat  $\mu$  si i únicament si  $\lim_{n \rightarrow \infty} F_n(x) = F_n(a) + 1 - F_n(b) \longrightarrow F(a) + 1 - F(b) = \mu((a, b])$ .

## 11. CONVERGÈNCIA FEBLE DE PROBABILITATS

En aquest capítol analitzarem un tipus de convergència de caràcter diferent als estudiatss en el Capítol 9. Els resultats que demostrarrem seran utilitzats en la prova del teorema central del límit, que donem en el Capítol 13.

11.1.1. Convergència feble de probabilitats i convergència en llei d'una successió de variables aleatòries

El tipus de convergència que anem a introduir difereix de les convergències que hem estudiats fins ara (en probabilitat, quasi segura i en  $L^p$ ) en què es referix a una successió de probabilitats sobre  $\mathbb{R}$  en lloc d'una successió de variables aleatòries.

Considerem una successió  $\{\mu_n, n \geq 1\}$  de probabilitats sobre  $\mathbb{R}$ . Ens podem plantejar la qüestió següent: Quan direm que  $\mu_n$  convergeix cap a una probabilitat  $\mu$ ? La definició més natural seria dir que  $\lim_{n \rightarrow \infty} \mu_n(B) = \mu(B)$  per a tot  $B \in \mathcal{B}(\mathbb{R})$ . Ara bé aquesta definició és clarament inadequada pels nostres objectius. En efecte, ens interessarà en alguns casos (per exemple en l'aproximació de la llei binomial per la llei normal, o en l'aproximació de la llei geomètrica per una exponencial) poder dir que una successió de probabilitats discretes  $\mu_n$  convergeix cap a una llei contínua  $\mu$ . Això no és possible amb la definició anterior ja que si  $S$  és el conjunt numerable igual a la unió dels suports de les  $\mu_n$ , tindrem  $\mu_n(S) = 1$  per a tot  $n$  però  $\mu(S) = 0$ .

**Definició 11.1.1.** Sigui  $\{\mu_n, n \geq 1\}$  una successió de probabilitats en  $\mathbb{R}$ . Direm que aquesta successió *convergeix feblement* cap a una probabilitat  $\mu$  i escriurem  $w - \lim_{n \rightarrow \infty} \mu_n = \mu$ .

$$\lim_n \int_{\mathbb{R}} f d\mu_n = \int_{\mathbb{R}} f d\mu,$$

per a tota funció contínua i afitada  $f : \mathbb{R} \rightarrow \mathbb{R}$ .

Al llarg d'aquest capítol designarem per  $C_b(\mathbb{R})$  el conjunt de les funcions  $f : \mathbb{R} \rightarrow \mathbb{R}$  contínues i afitades.

El següent resultat caracteritza la convergència feble en termes de les funcions de distribució  $F_n$  i  $F$  de  $\mu_n$  i  $\mu$ , respectivament.

$$\mu_n((a, b]) = F_n(a) + 1 - F_n(b) \longrightarrow F(a) + 1 - F(b) = \mu((a, b]),$$

Subem que

$$\begin{aligned} \left| \int_{\mathbb{R}} f d\mu_n - \int_{\mathbb{R}} f d\mu \right| &= \left| \int_{(a, b)^c} f d\mu_n + \sum_{i=1}^r \int_{I_i} f d\mu_n - \int_{(a, b)^c} f d\mu - \sum_{i=1}^r \int_{I_i} f d\mu \right| \\ &\leq \|f\|_{\infty} (\mu_n((a, b]^c) + \mu((a, b]^c)) + \sum_{i=1}^r \left| \int_{I_i} f d\mu_n - \int_{I_i} f d\mu \right|. \end{aligned}$$

quan  $n$  tendeix a infinit.

D'altra banda,

$$\begin{aligned} \left| \int_{I_i} f d\mu_n - \int_{I_i} f d\mu \right| &\leq \left| \int_{I_i} (f - f(a_i)) d\mu_n \right| + \left| f(a_i) [\mu_n(I_i) - \mu(I_i)] \right| \\ &+ \left| \int_{I_i} (f - f(a_i)) d\mu \right| \leq \varepsilon (\mu_n(I_i) + \mu(I_i)) + \left| f(a_i) [\mu_n(I_i) - \mu(I_i)] \right|. \end{aligned}$$

i també sabem que  $\lim_{n \rightarrow \infty} \mu_n(I_i) = \mu(I_i)$  per a tot  $i = 1, \dots, r$ . En conseqüència obtenim

$$\begin{aligned} \limsup_n \left| \int_{\mathbb{R}} f d\mu_n - \int_{\mathbb{R}} f d\mu \right| &\leq 2 \|f\|_\infty \mu((a, b]) + \sum_{i=1}^r 2\varepsilon \mu(I_i) \\ &\leq \varepsilon (2 \|f\|_\infty + 1). \end{aligned}$$

Com que  $\varepsilon$  és arbitrari, tindrem  $\lim_n \int_{\mathbb{R}} f d\mu_n = \int_{\mathbb{R}} f d\mu$ , que és el que volem demostrar. ■

#### Observacions

- El límit feble d'una successió  $\mu_n$ , si existeix, és únic.

En efecte si  $\mu$  i  $\mu'$  són dos límits, pel teorema anterior les funcions de distribució de  $\mu$  i de  $\mu'$  coincideixen, llevat potser d'un conjunt numerable, i com que són contínues per la dreta, són iguals. Per tant,  $\mu = \mu'$ .

- De la demonstració de la primera part del teorema es dedueix que per a comprovar la convergència  $w - \lim_{n \rightarrow \infty} \mu_n = \mu$  només cal veure que  $\lim_n \int_{\mathbb{R}} f d\mu_n = \int_{\mathbb{R}} f d\mu$  per a tota funció  $f : \mathbb{R} \rightarrow \mathbb{R}$  uniformement contínua i afistada, ja que les funcions  $f_\varepsilon^+$  i  $f_\varepsilon^-$  són d'aquest tipus. En realitat només cal comprovar la convergència de les integrals per a tota funció  $f$  de classe  $C^\infty$ , tal que ella i totes les seves derivades estan afistades ( $f \in C_b^\infty(\mathbb{R})$ ). Per a veure això, prenem una funció  $\varphi \in C_b^\infty(\mathbb{R})$  tal que  $0 \leq \varphi \leq 1$ ,  $\varphi(x) = 1$  si  $x \geq 1$  i  $\varphi(x) = 0$  si  $x \leq 0$ . Per exemple podem agafar  $\varphi(x) = \frac{1}{c} \exp\left(-\frac{|x|}{c(1-x)}\right)$  per a tot  $x \leq 1$ , on  $c = \int_0^1 \exp\left(-\frac{|x|}{c(1-x)}\right) dx$ .

Llavors, definim  $f_\varepsilon^+(y) = \varphi\left(\frac{x-y+\varepsilon}{\varepsilon}\right)$ ,  $f_\varepsilon^-(y) = \varphi\left(\frac{x-y}{\varepsilon}\right)$  i procedim com en la primera part de la demonstració del teorema.

#### Exemples

- En el cas d'una successió de deltes de Dirac, la convergència feble equival a la convergència ordinària d'una successió de nombres reals. Més precisament,  $w - \lim_{n \rightarrow \infty} \delta_{x_n} = \delta_x$  equivalent a que  $\lim_{n \rightarrow \infty} x_n = x$ .

**DemOSTRACIÓ:** Suposem que hi ha convergència de les deltes de Dirac. Fixem  $\varepsilon > 0$ . sabem que

$$\begin{aligned} \lim_n F_n(x + \varepsilon) &= F(x + \varepsilon) = 1, \\ \lim_n F_n(x - \varepsilon) &= F(x - \varepsilon) = 0, \end{aligned}$$

ja que  $x + \varepsilon$  i  $x - \varepsilon$  són punts de continuitat de la funció de distribució  $F$  de  $\delta_x$ . Com que les funcions  $F_n$  només prenen els valors 0 i 1, existeix un  $n_0$  tal que  $F_n(x + \varepsilon) = 1$  i  $F_n(x - \varepsilon) = 0$ , per a tot  $n \geq n_0$ . És a dir, per a tot  $n \geq n_0$   $x - \varepsilon < x_n \leq x + \varepsilon$ . Per tant,  $\lim_n x_n = x$ .

Recíprocament, suposem que  $\lim_{n \rightarrow \infty} x_n = x$ . Per a tota funció real  $f$  contínua i afistada es complirà  $\lim_{n \rightarrow \infty} f(x_n) = f(x)$ , i en conseqüència  $w - \lim_{n \rightarrow \infty} \delta_{x_n} = \delta_x$ .

- Si  $\mu_n = B(n, p_n)$  són lleis binomials tals que  $\lim_{n \rightarrow \infty} n p_n = \lambda > 0$ , aleshores  $w - \lim_{n \rightarrow \infty} \mu_n = \mu$ , on  $\mu$  és la distribució de Poisson amb paràmetre  $\lambda$ .

**DemOSTRACIÓ:** En efecte, en el capítol 8 hem vist que  $\lim_n \mu_n(\{k\}) = \mu(\{k\})$  per a tot natural  $k$  i això implica clarament, la convergència de les funcions de distribució, en tot punt.

La convergència feble de probabilitats sobre la recta permet definir un concepte de convergència per a una successió de variables aleatories, donat que tota variable aleatòria inclou una probabilitat sobre la recta real.

**DefINICIÓ 11.3.** Considerem una successió  $\{X_n, n \geq 1\}$  de variables aleatories definides en un espai de probabilitat  $(\Omega, \mathcal{A}, P)$ . Direm que aquesta successió convergeix en llei cap a una variable  $X$  i escriurem  $\mathcal{L} - \lim_{n \rightarrow \infty} X_n = X$ , si

$$w - \lim_{n \rightarrow \infty} P \circ X_n^{-1} = P \circ X^{-1},$$

o equivalentment

$$\lim_{n \rightarrow \infty} E[f(X_n)] = E[f(X)],$$

per a tota funció  $f : \mathbb{R} \rightarrow \mathbb{R}$  contínua i afistada.

Cal observar que aquí el límit no és únic, ja que només podem determinar amb unicitat la llei de la variable límit. També es diu que la successió  $X_n$  convergeix en llei cap a una probabilitat  $\mu$  si  $w - \lim_{n \rightarrow \infty} P \circ X_n^{-1} = \mu$ . La definició de convergència feble es pot estendre al cas de probabilitats en  $\mathbb{R}^n$ , i es pot definir també la convergència en llei d'una successió de vectors aleatoris. Es podria demostrar també una versió multidimensional del Teorema 11.2.

### Exemple

Considerem una successió de variables aleatòries independents amb distribució de Bernoulli de paràmetre  $p_n$ . Per a cada natural  $n \geq 1$  sigui  $T_n$  el primer instant en que aquesta successió pren el valor 1. Sabem que  $T_n$  té una llei geomètrica de paràmetre  $p_n$ , és a dir,

$$P\{T_n = k\} = (1 - p_n)^{k-1} p_n, \quad k \geq 1.$$

Suposem que  $p_n = \frac{\lambda}{n}$  on  $\lambda > 0$  és una constant i definim  $S_n = \frac{T_n}{n}$ . Això vol dir, que hem fet un canvi d'escala en el temps, de forma que entre dues realitzacions seguides de l'experiència hi hagi un interval de temps de longitud  $1/n$ .

Aquest canvi d'escala compensa el fet de que la probabilitat de treure un 1, que és  $p_n = \frac{\lambda}{n}$ , és cada cop més petita quan  $n$  augmenta. Cal observar que l'esperança de  $S_n$  es manté constant, igual a  $\lambda^{-1}$ . La successió  $S_n$  convergeix en llei cap a una distribució exponencial de paràmetre  $\lambda$ . En efecte, per a tot nombre real  $a > 0$  tindrem

$$P\{S_n > a\} = P\{T_n > an\} = \sum_{k=\lceil a\eta \rceil + 1}^{\infty} p_n (1 - p_n)^{k-1} = \left(1 - \frac{\lambda}{n}\right)^{\lceil a\eta \rceil},$$

que tendeix cap a  $e^{-a\lambda}$  quan  $n$  tendeix cap a infinit.

Establirem tot seguit dues propietats interessants de la convergència en llei.

**Proposició 11.4.** Sigui  $\{X_n, n \geq 1\}$  una successió de variables aleatòries que convergeix en llei cap a una variable aleatòria  $X$ , i sigui  $f : \mathbb{R} \rightarrow \mathbb{R}$  una funció contínua. Aleshores la successió  $\{f(X_n), n \geq 1\}$  convergeix en llei cap a la variable aleatòria  $f(X)$ .

**Demostració:** Per a tota funció  $g \in C_b(\mathbb{R})$ , la composició  $g \circ f$  és també contínua i afistada, i per tant,

$$\lim_n E[g(f(X_n))] = E[g(f(x))]. \quad \blacksquare$$

**Proposició 11.5.** Si  $\{X_n, n \geq 1\}$  és una successió de variables aleatòries que convergeix en probabilitat cap a una variable aleatòria  $X$ , també convergeix en llei cap a  $X$ . El recíproc és cert si la variable  $X$  és constant q.s.

**Demostració:** Suposem primer que  $P - \lim_{n \rightarrow \infty} X_n = X$ . Sigui  $f : \mathbb{R} \rightarrow \mathbb{R}$  una funció afistada i uniformement contínua. Fixat  $\varepsilon > 0$ , sigui  $\delta > 0$  tal que si  $|x - y| \leq \delta$  aleshores  $|f(x) - f(y)| \leq \varepsilon$ .

Tindrem

$$\begin{aligned} |E(f(X_n)) - E(f(X))| &\leq \int_{\{|X_n - X| \leq \delta\}} |f(X_n) - f(X)| dP \\ &\quad + \int_{\{|X_n - X| > \delta\}} |f(X_n) - f(X)| dP \\ &\leq \varepsilon + 2\|f\|_\infty P\{|X_n - X| > \delta\}. \end{aligned}$$

En conseqüència,

$$\limsup_n |E(f(X_n)) - E(f(X))| \leq \varepsilon.$$

Com que  $\varepsilon$  és arbitrari,  $\lim_n E(f(X_n)) = E(f(X))$ , la qual cosa ens diu que  $L - \lim_{n \rightarrow \infty} X_n = X$ .

Suposem ara que  $X = a$  q.s. i que hi ha convergència en llei de la successió  $\{X_n, n \geq 1\}$  cap a  $X$ . Com que la funció  $f(x) = 1 \wedge |x - a|$  és contínua i afistada tindrem, per a tot  $0 < \varepsilon < 1$ ,

$$P\{|X_n - X| > \varepsilon\} \leq P\{1 \wedge |X_n - X| > \varepsilon\} \leq \frac{1}{\varepsilon} E[1 \wedge \{|X_n - X|\}].$$

expressió que tendeix a zero quan  $n$  tendeix a infinit. Per tant  $P - \lim_{n \rightarrow \infty} X_n = X$ . ■

La convergència feble de probabilitats en  $\mathbb{R}$  és metrizable. Si  $\mu$  i  $\nu$  són probabilitats sobre  $\mathbb{R}$  amb funcions de distribució  $F$  i  $G$  respectivament. Definim

$$d(\mu, \nu) = \inf\{s > 0 : F(x - s) - \varepsilon \leq G(x) \leq F(x + s) + \varepsilon, x \in \mathbb{R}\}.$$

Es pot demostrar que  $d$  és una distància en el conjunt  $\mathcal{P}$  de totes les probabilitats sobre  $\mathbb{R}$  i que hi ha equivalència entre  $w - \lim_{n \rightarrow \infty} \mu_n = \mu$  i  $\lim_{n \rightarrow \infty} d(\mu_n, \mu) = 0$ . A d se l'anomena distància de P. Lévy.

### 11.2. Compacitat feble i ajustament

**Definició 11.6.** Sigui  $m$  una família de probabilitats sobre  $\mathbb{R}$ .

(i) Direm que  $m$  és relativament compacte si tota successió d'elements de  $m$  té una subsuccessió feblement convergent.

(ii) Direm que  $m$  és ajustada si per a tot  $\varepsilon > 0$  existeix un  $a > 0$  tal que  $\mu([-a, a]^c) < \varepsilon$  per a tota  $\mu \in m$ .

Observis que si  $m$  és una família finita, aleshores  $m$  és ajustada ja que, si  $m = \{\mu_1, \mu_2, \dots, \mu_r\}$  hindrem

$$\sup_{\mu \in m} \mu([(-n, n]^c) \leq \sum_{i=1}^r \mu_i([(-n, n]^c),$$

que tendeix a zero si  $n$  tendeix a infinit.

L'objectiu d'aquest apartat és demostrar l'equivalència entre els dos conceptes introduïts en la definició anterior. Necessitem abans un resultat tècnic que donem tot seguit. El següent resultat estableix l'equivalència entre els dos conceptes que acabem d'introduir.

**Teorema 11.7. (Helly-Bray).** Sigui  $\{F_n, n \geq 1\}$  una successió de funcions de  $\mathbb{R}$  a valors en  $[0, M]$ , creixents i contínues per la dreta. Aleshores existeix una funció  $F : \mathbb{R} \rightarrow [0, M]$  creixent i contínua per la dreta i existeix una subsuccessió  $F_{n_k}$  tal que

$$\lim_n F_{n_k}(x) = F(x),$$

per a tot punt  $x$  de contínuitat de  $F$ .

**DemOSTRaciÓ:** Fixem un subconjunt  $D = \{x_n, n \geq 1\}$  numerable i dens en  $\mathbb{R}$ . Utilitzarem el mètode diagonal de Cantor per a construir una subsuccessió  $F_{n_k}$  tal que  $\{F_{n_k}(x_j), k \geq 1\}$  convergeix cap a un cert límit  $y_j$ , per a tot  $j \geq 1$ :

Com que  $0 \leq F_n(x_1) \leq M$ , per a tot  $n \geq 1$ , existeix una subsuccessió  $\{F_{1,n}(x_1)\}$  convergent cap a un límit  $y_1 \in [0, M]$ .

Donat que  $0 \leq F_{1,n}(x_2) \leq M$ , per a tot  $n \geq 1$ , existeix una subsuccessió  $\{F_{2,n}(x_2)\}$  convergent cap a un límit  $y_2 \in [0, M]$ .

En general, com que  $0 \leq F_{m,n}(x_{m+1}) \leq M$ , per a tot  $n \geq 1$ , existeix una subsuccessió  $\{F_{m+1,n}(x_{m+1})\}$  convergent cap un límit  $y_{m+1} \in [0, M]$ .

Considerem la successió diagonal  $F_{n_k}(x) = F_{k,k}(x)$ . Per a cada  $x_j \in D$ , la successió  $F_{n_k}(x_j) = F_{k,k}(x_j)$  és una parcial de  $\{F_{j,k}(x_j), k \geq 1\}$  per  $k \geq j$  i, per tant,  $\lim_n F_{n_k}(x_j) = y_j$ .

Definim la funció  $F_D : D \rightarrow [0, M]$  per  $F_D(x_j) = y_j$ . Observis que  $\lim_k F_{n_k}(x) = F_D(x)$  per a tot  $x \in D$ .

La funció  $F_D$  és creixent. En efecte, si  $x \leq y$ ,  $F_{n_k}(x) \leq F_{n_k}(y)$  per a tot  $k \geq 1$ . Per tant  $F_D(x) \leq F_D(y)$ .

Sigui ara  $F : \mathbb{R} \rightarrow [0, M]$  definida per  $F(x) = \inf\{F_D(y); y \in D, y > x\}$ . Aquesta funció té les propietats següents:

- (1)  $F$  és creixent. En efecte, si  $x_1 \leq x_2$ , tenim  $F(x_1) = \inf\{F_D(y); y \in D, y > x_1\} \leq \inf\{F_D(y); y \in D, y > x_2\} = F(x_2)$ .
- (2)  $F$  és contínua per la dreta. En efecte, sigui  $\{z_n, n \geq 1\}$  una successió decreixent cap a  $x$ . Aleshores  $F(z_n) \downarrow b \geq F(x)$ . Suposem que  $b > F(x)$ . Per la definició de  $F(x)$  existeix un  $y_0 \in D$  tal que  $y_0 > x$  i  $F_D(y_0) > b$ . Aleshores, per  $n$  prou gran,  $x \leq z_n \leq y_0$ , i en conseqüència  $F(z_n) \leq F_D(y_0) < b$  (per la definició de  $F(z_n)$ ). Per tant  $b = \lim_{n \rightarrow \infty} F(z_n) \leq F_D(y_0) < b$  i obtenim una contradicció.
- (3) En tot punt  $x$  de contínuitat de  $F$ ,  $\lim_n F_{n_k}(x) = F(x)$ . En efecte, sigui  $y \in D, y > x$ . Tindrem

$$\limsup_k F_{n_k}(x) \leq \limsup_k F_{n_k}(y) = F_D(y),$$

i per tant

$$\limsup_k F_{n_k}(x) \leq F(x).$$

D'altra banda, si  $x' < y < x$ ,  $y \in D$ , tindrem

$$\liminf_k F_{n_k}(x) \geq \liminf_k F_{n_k}(y) = F_D(y) \geq F(x'),$$

per la definició de  $F(x')$ . Com que això val per a tot  $x' < x$  i  $F$  és contínua en el punt  $x$ , obtenim  $\liminf_k F_{n_k}(x) \geq F(x)$ , i en conseqüència,  $F(x) = \lim_k F_n(x)$ . ■

Observacions

- 1.- Es pot demostrar de manera anàloga una versió del teorema de Helly-Bray per a funcions  $F : \mathbb{R}^n \rightarrow [0, M]$  contínues per la dreta i amb increments rectangulars no negatius.
- 2.- Si la successió  $\{F_n, n \geq 1\}$  està formada per funcions de distribució és a dir,  $M = 1$ ,  $\lim_{r \rightarrow -\infty} F_n(x) = 0$  i  $\lim_{r \rightarrow +\infty} F_n(x) = 1$ , pot ocórrer que la funció  $F$  no sigui una funció de distribució. És a dir, donada una successió de probabilitats  $\{\mu_n, n \geq 1\}$  en  $\mathbb{R}$ , el teorema de Helly-Bray no ens proporciona, en general, una subseqüència feblement convergent, ja que la funció  $F$  que apareix en l'enunciat del teorema pot no complir les condicions  $\lim_{r \rightarrow -\infty} F(x) = 0$  i  $\lim_{r \rightarrow +\infty} F(x) = 1$ .

Per exemple, si  $\mu_n = \delta_{-n}$ ,  $F_n(x) = 1_{[n, +\infty)}(x)$  i  $\lim_n F_n(x) = 0$  per a tot  $x$ . Analogament, si  $\mu_n = \delta_{-n}$ ,  $F_n(x) = 1_{[-n, +\infty)}(x)$  i  $\lim_n F_n(x) = 1$  per a tot  $x$ .

En aquests dos exemples, la successió  $\{\mu_n, n \geq 1\}$  no té cap parcial feblement convergent. Intuitivament, això és degut a que  $\mathbb{R}$  no és compacte i la massa s'escapa cap a l'infinít.

Les definicions donades en 11.6 proporcionen condicions per a que una família de probabilitats sobre  $\mathbb{R}$  sigui tal que tota successió d'elements de la família contingui una subsuccessió feblement convergent.

**Teorema 11.8. (Prokhorov).** Sigui  $m$  una família de probabilitats en  $\mathbb{R}$ . Aleshores  $m$  és ajustada si i només si  $m$  és relativament compacte.

*DemOSTRació:* (1) Suposem primer que  $m$  és ajustada i considerem una successió  $\{\mu_n, n \geq 1\} \subset m$ . Per a cada  $n \geq 1$ , sigui  $F_n$  la funció de distribució de  $\mu_n$ . Pel teorema de Helly-Bray, existeix una subsuccessió  $\{F_{n_k}, k \geq 1\}$  tal que  $\lim_k F_{n_k}(x) = F(x)$ , en tot punt  $x$  de continuïtat de  $F$ , on la funció  $F : \mathbb{R} \rightarrow [0, 1]$  és creixent i continua per la dreta. Llavors només cal veure que

$$\lim_{x \rightarrow -\infty} F(x) = 0 \quad \text{i} \quad \lim_{x \rightarrow \infty} F(x) = 1. \quad (11.1)$$

Fixem un  $\epsilon > 0$  i sigui  $a > 0$  tal que  $\mu_n([ -a, a]^c) < \epsilon$ , per a tot  $n \geq 1$ . Sigui  $b \geq a$  i  $c < -a$  punts de continuïtat de  $F$ . Tindrem

$$F_n(b) \geq \mu_n([ -a, a]) > 1 - \epsilon,$$

$$F_n(c) \leq \mu_n([ -a, a]^c) < \epsilon.$$

Fent  $n \rightarrow \infty$  obtenim  $F(b) > 1 - \epsilon$  i  $F(c) < \epsilon$ , i, en conseqüència, (11.1) és cert.

(2) Recíprocament, suposem que la família  $m$  és relativament compacte. Si no fos ajustada existiria un  $\epsilon > 0$  tal que  $\forall n \geq 1$  existeix  $\mu_n \in m$  amb  $\mu_n([ -n, n]^c) \geq \epsilon$ . Per hipòtesi, existirà una subsuccessió  $\{\mu_{n_k}, k \geq 1\}$  tal que  $w = \lim_{k \rightarrow \infty} \mu_{n_k} = \mu$ .

Considerem la funció  $f^m(x) = [|x| - (m - 1)]^+$ . Per a tot  $n \geq m$  tindrem

$$\begin{aligned} 0 < \epsilon &\leq \limsup_k \mu_{n_k}([ -n_k, n_k]^c) \leq \limsup_k \int_{\mathbb{R}} f^m d\mu_{n_k} = \int_{\mathbb{R}} f^m d\mu \\ &\leq \mu([ -m + 1, m - 1]^c), \end{aligned}$$

i fent tendir  $m$  a infinit obtenim una contradicció. ■

Del teorema de Prokhorov es dedueix el següent criteri de convergència feble:

**Teorema 11.9.** Sigui  $\{\mu_n, n \geq 1\}$  una successió ajustada de probabilitats en  $\mathbb{R}$  tal que totes les seves subsuccessions convergents tenen el mateix límit  $\mu$ . Aleshores  $w = \lim_{n \rightarrow \infty} \mu_n = \mu$ .

*DemOSTRació:* Suposem que  $\{\mu_n, n \geq 1\}$  no convergeixi feblement cap a  $\mu$ . Existeix una funció  $f \in C_b(\mathbb{R})$  tal que  $\{\int_{\mathbb{R}} f d\mu_n, n \geq 1\}$  no convergeix cap a  $\int_{\mathbb{R}} f d\mu$ . Això ens diu que existeix un  $\epsilon > 0$  i una subsuccessió  $\{\mu_{n_k}, k \geq 1\}$  tal que  $|\int_{\mathbb{R}} f d\mu_{n_k} - \int_{\mathbb{R}} f d\mu| \geq \epsilon$ ,  $\forall k \geq 1$ . Pel teorema de Prokhorov, ha d'existir una subsuccessió  $\{\mu_{n_{k_i}}, i \geq 1\}$  de la successió  $\{\mu_{n_k}, k \geq 1\}$  feblement convergent i amb límit  $\mu$ , la qual cosa contradiu la desigualtat anterior. ■

## 12. FUNCTIONS CARACTERÍSTIQUES

En aquest capítol estudarem una tècnica per a tractar la convergència feble de probabilitats: Les funcions característiques introduïdes per Paul Lévy.

Sigui  $\mu$  una probabilitat en  $\mathbb{R}$ . La funció característica de  $\mu$  es defineix com l'aplicació  $\varphi_\mu : \mathbb{R} \rightarrow \mathbb{C}$  donada per

$$\varphi_\mu(t) = \int_{\mathbb{R}} e^{itx} \mu(dx) = \int_{\mathbb{R}} \cos(tx) \mu(dx) + i \int_{\mathbb{R}} \sin(tx) \mu(dx).$$

La funció  $\varphi_\mu$  està ben definida, ja que les funcions sinus i cosinus són contínues i afinitades.

Si  $X$  és una variable aleatòria en un cert espai de probabilitat  $(\Omega, \mathcal{A}, P)$ , la funció característica de  $X$  serà, per definició, la funció característica de la seva llei, és a dir,

$$\varphi_X(t) = \int_{\mathbb{R}} e^{itx} P_X(dx) = E(e^{itX}).$$

Anàlogament, si  $\mu$  és una probabilitat en  $\mathbb{R}^n$ , la funció característica de  $\mu$  es defineix com  $\varphi_\mu : \mathbb{R}^n \rightarrow \mathbb{C}$ ,  $\varphi_\mu(t) = \int_{\mathbb{R}^n} e^{i\langle t, x \rangle} \mu(dx)$  i la funció característica d'un vector aleatori  $X = (X_1, \dots, X_n)$  serà la funció característica de la seva llei.

### 12.1. Propietats fonamentals de les funcions característiques

En tot aquest apartat  $\mu$  serà una probabilitat en  $\mathbb{R}^n$ . Es compleixen les següents propietats.

$$(1) \quad \varphi_\mu(0) = 1.$$

$$(2) \quad |\varphi_\mu(t)| \leq 1, \text{ per a tot } t \in \mathbb{R}^n.$$

Aquesta propietat es dedueix fàcilment de la identitat  $|e^{i\langle t, x \rangle}| = 1$ .

$$(3) \quad \varphi_\mu(-t) = \overline{\varphi_\mu(t)}.$$

En efecte,

$$\varphi_\mu(-t) = \int_{\mathbb{R}^n} e^{-i\langle -t, x \rangle} \mu(dx) = \int_{\mathbb{R}^n} \overline{e^{i\langle t, x \rangle}} \mu(dx) = \overline{\int_{\mathbb{R}^n} e^{i\langle t, x \rangle} \mu(dx)} = \overline{\varphi_\mu(t)}.$$

(4)  $\varphi_\mu$  és una funció uniformement contínua.

*Demostració:* Qualsevol que siguin  $s, t \in \mathbb{R}^n$  tenim

$$\begin{aligned} |\varphi_\mu(t) - \varphi_\mu(s)| &= \left| \int_{\mathbb{R}^n} (e^{-i\langle t, x \rangle} - e^{i\langle s, x \rangle}) \mu(dx) \right| \\ &\leq \int_{\mathbb{R}^n} |e^{i\langle t-s, x \rangle} - 1| \mu(dx) \end{aligned}$$

Aquest darrer integrand està afusat per 2 i tendix a zero quan  $|t-s|$  tendix a zero. Aleshores només cal aplicar convergència dominada per a deduir el resultat.

(5) Sigui  $X$  un vector aleatori  $n$ -dimensional,  $A$  una matríg  $m \times n$  i  $b \in \mathbb{R}^m$ . Aleshores per a tot  $t \in \mathbb{R}^n$ ,

$$\varphi_{AX+b}(t) = e^{i\langle t, AX+b \rangle}.$$

En efecte,

$$\varphi_{AX+b}(t) = E(e^{i\langle t, AX+b \rangle}) = e^{i\langle t, AX+b \rangle} \varphi_X(A^* t).$$

#### (6) Propietat fonamental d'injectivitat.

Si  $\mu_1$  i  $\mu_2$  són dues probabilitats en  $\mathbb{R}^n$  tals que  $\varphi_{\mu_1} = \varphi_{\mu_2}$ , necessàriament  $\mu_1 = \mu_2$ .

*Demostració:* Aquest resultat es pot deduir de la fórmula de inversió que demostrarem més endavant. Ara donarem les idees bàsiques d'una demostració directa. Suposarem per simplificar que  $n = 1$ .

Fixem un interval  $[-T, T]$ . Pel teorema de Stone-Weierstrass, les combinacions lineals finites (a coeficients complexos) de les funcions  $e^{ikx}/T$ ,  $k \in \mathbb{Z}$ , són denses (respecte la norma del suprem) en el conjunt de les funcions contínues sobre  $[-T, T]$  a valors complexes. En efecte, aquestes combinacions lineals formen una àlgebra de funcions, que conté les constants, és estable per conjugació, i separa punts.

Cal doncs veure que  $\int_{\mathbb{R}} f d\mu_1 = \int_{\mathbb{R}} f d\mu_2$  per a tota funció real contínua  $f$  amb suport compacte. Fixem un  $\varepsilon > 0$  i prenem  $T > 0$  de forma que el suport de  $f$  estigui contingut en  $[-T, T]$  i que  $\mu_1([-T, T]^c) \leq \varepsilon$ ,  $\mu_2([-T, T]^c) \leq \varepsilon$ .

Per l'observació anterior, existirà una funció

$$\hat{f}(x) = \sum_{j=1}^m a_j \exp(i\pi k_j x/T),$$

amb  $a_j \in \mathbb{C}$ , tal que  $\sup_{|x| \leq T} |f(x) - \hat{f}(x)| < \varepsilon$ . Per hipòtesi sabem que  $\int_{\mathbb{R}} \hat{f} d\mu_1 = \int_{\mathbb{R}} \hat{f} d\mu_2$ . D'altra banda, com que la funció  $\hat{f}$  és periòdica, amb període  $2T$  tindrem que

$$\|\hat{f}\|_\infty = \sup_{x \in [-T, T]} |\hat{f}(x)| \leq \varepsilon + \|f\|_\infty.$$

En conseqüència,

$$\begin{aligned} \left| \int_{\mathbb{R}} f d\mu_1 - \int_{\mathbb{R}} f d\mu_2 \right| &\leq \left| \int_{\mathbb{R}} f d\mu_1 - \int_{\mathbb{R}} \hat{f} d\mu_1 \right| + \left| \int_{\mathbb{R}} \hat{f} d\mu_1 - \int_{\mathbb{R}} \hat{f} d\mu_2 \right| \\ &\leq \varepsilon (\mu_1([-T, T]) + \mu_2([-T, T])) + 2\varepsilon (\varepsilon + 2\|f\|_\infty) \\ &\leq 2\varepsilon (1 + \varepsilon + 2\|f\|_\infty), \end{aligned}$$

i com que  $\varepsilon > 0$  és arbitrari deduirem el que volem.

(7) Direm que una probabilitat  $\mu$  sobre  $\mathbb{R}^n$  és simètrica si  $\mu(B) = \mu(-B)$  per a tot borelià  $B$ .

B. La propietat de que  $\mu$  sigui simètrica és equivalent a que la seva funció característica  $\varphi_\mu$  sigui real.

**Demostació:** Considerem l'aplicació  $\tau : \mathbb{R}^n \rightarrow \mathbb{R}^n$  definida per  $\tau(x) = -x$ . Per a tota probabilitat  $\mu$  en  $\mathbb{R}^n$  es compleix la següent relació

$$\overline{\varphi_\mu} = \varphi_{\mu \circ \tau^{-1}}. \quad (12.1)$$

En efecte, pel teorema de la mesura imatge, per a tota funció  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  mesurable i afinitada, es compleix que

$$\int_{\mathbb{R}^n} f(-x)\mu(dx) = \int_{\mathbb{R}^n} (f \circ \tau)(x)\mu(dx) = \int_{\mathbb{R}^n} f(x)(\mu \circ \tau^{-1})(dx).$$

En conseqüència,

$$\overline{\varphi_\mu}(-t) = \varphi_\mu(-t) = \int_{\mathbb{R}^n} e^{-i\langle t, x \rangle} \mu(dx) = \int_{\mathbb{R}^n} e^{i\langle t, x \rangle} (\mu \circ \tau^{-1})(dx) = \varphi_{\mu \circ \tau^{-1}}(t).$$

Aleshores, tenint en compte la propietat (6), tindrem que  $\mu$  és simètrica si i únicament si  $\mu = \mu \circ \tau^{-1}$ , i així equival a que  $\varphi_\mu = \varphi_{\mu \circ \tau^{-1}} = \overline{\varphi_\mu}$ .

(8) Sigui  $X = (X_1, \dots, X_n)$  un vector aleatori. Les variables aleatories  $X_1, \dots, X_n$  són independents si i només si

$$\varphi_X(t_1, \dots, t_n) = \varphi_{X_1}(t_1) \cdots \varphi_{X_n}(t_n). \quad (12.2)$$

**Demostació:** Sabem que  $X_1, \dots, X_n$  són independents si i només si  $P \circ X^{-1} = P$ ,  $X_1^{-1} \times \dots \times P \circ X_n^{-1}$  i, per la propietat (6) això equival a la igualtat de les funcions característiques d'aquestes probabilitats. La funció característica de  $P \circ X^{-1}$  es  $\varphi_X$  i la de  $P \circ X_1^{-1} \times \dots \times P \circ X_n^{-1}$  val

$$\begin{aligned} \int_{\mathbb{R}^n} e^{i\langle t, x \rangle} (P \circ X_1^{-1})(dx_1) \cdots (P \circ X_n^{-1})(dx_n) &= \left( \int_{\mathbb{R}} e^{it_1 X_1} (P \circ X_1^{-1})(dx_1) \right) \cdots \\ &\quad \left( \int_{\mathbb{R}} e^{it_n X_n} (P \circ X_n^{-1})(dx_n) \right) \\ &= \varphi_{X_1}(t_1) \cdots \varphi_{X_n}(t_n). \end{aligned}$$

Queda doncs estabert el resultat. ■

(9) Si  $X_1, \dots, X_n$  són variables aleatories independents, aleshores

$$\varphi_{X_1 + \dots + X_n}(t) = \varphi_{X_1}(t) \cdots \varphi_{X_n}(t).$$

**Demostració:** Pel la independència

$$\begin{aligned} \varphi_{X_1 + \dots + X_n}(t) &= E\left(e^{i(t(X_1 + \dots + X_n))}\right) = E\left(e^{itX_1}\right) \cdots E\left(e^{itX_n}\right) \\ &= \varphi_{X_1}(t) \cdots \varphi_{X_n}(t). \end{aligned}$$

## 12.2. Funció característica i moments

En aquest apartat relacionarem l'existència de moments per a una probabilitat  $\mu$  sobre la recta real amb el comportament a l'origen de la funció característica de  $\mu$ .  
Establirem primer un resultat tècnic sobre derivació sota el signe integral.

**Lema 12.1.** Sigui  $f(t, x) : \mathbb{R}^2 \rightarrow \mathbb{R}$  una funció contínua amb derivada parcial contínua  $\frac{\partial f}{\partial t}$ . Suposem que  $|f(t, x)| + |\frac{\partial f}{\partial t}(t, x)| \leq g(x)$  on  $g$  és una funció integrable respecte una probabilitat  $\mu$  en  $\mathbb{R}$ . Aleshores, la funció  $F(t) = \int_{\mathbb{R}} f(t, x)\mu(dx)$  és derivable i

$$F'(t) = \int_{\mathbb{R}} \frac{\partial f}{\partial t}(t, x)\mu(dx).$$

**Demostació:** Fixem  $t \in \mathbb{R}$ ,  $h \neq 0$  i calculem

$$\frac{1}{h} [F(t+h) - F(t)] = \int_{\mathbb{R}} \frac{1}{h} [f(t+h, x) - f(t, x)]\mu(dx).$$

Si prenem una successió  $\{h_n, n \geq 0\}$  que convergeix cap a zero, per a cada  $x$  tindrem

$$\lim_n \frac{1}{h_n} [f(t+h_n, x) - f(t, x)] = \frac{\partial f}{\partial t}(t, x).$$

Pel teorema de convergència dominada podem commutar aquest límit puntual amb la integral respecte  $\mu$ . En efecte, pel teorema del valor mig

$$\left| \frac{1}{h_n} [f(t+h_n, x) - f(t, x)] \right| = \left| \frac{\partial f}{\partial t}(t', x) \right| \leq g(x).$$

Per tant:

$$\lim_n \frac{1}{h_n} [F(t+h_n) - F(t)] = \frac{\partial f}{\partial t}(t, x). \blacksquare$$

Recordem que una funció  $\varphi : \mathbb{R} \rightarrow \mathbb{C}$  és derivable en un punt si i només si ho són la seva part real i la seva part imaginària. Observi's també que el lema anterior és igualment cert per a funcions que prenen valors en  $\mathbb{C}$ .

**Teorema 12.2.** Sigui  $\mu$  una probabilitat sobre  $\mathbb{R}$  amb moment d'ordre  $n \geq 1$  finit. Aleshores la funció característica  $\varphi_\mu$  és  $n$  vegades derivable i

$$\varphi_\mu^{(k)}(t) = i^k \int_{\mathbb{R}} x^k e^{itx} \mu(dx)$$

per  $k = 1, \dots, n$ .

En particular, si designem per  $m_i, i = 1, \dots, n$  als moments de la probabilitat  $\mu$ , es compleix

$$\varphi_\mu^{(k)}(0) = i^k m_k,$$

per  $k = 1, \dots, n$ .

**DemOSTRACIÓ:** Apliquem el lema 12.1 a la funció  $e^{itx}$  que compleix  $\left| \frac{\partial}{\partial t} e^{itx} \right| = |xe^{itx}| = |x|$ . Sabem que  $\int_{\mathbb{R}} |x| \mu(dx) < \infty$ . En conseqüència,  $\varphi_\mu(t) = \int_{\mathbb{R}} e^{itx} \mu(dx)$  és derivable i  $\varphi'_\mu(t) = i \int_{\mathbb{R}} xe^{itx} \mu(dx)$ . Finalment, aplicant iterativament el lema 12.1 a les funcions  $\varphi'_\mu, \varphi''_\mu, \dots, \varphi_{\mu}^{(n-1)}$  demostrarem el teorema. ■

El resultat següent estableix un tipus de recíproc d'aquest teorema.

**Teorema 12.3.** Sigui  $\mu$  una probabilitat sobre  $\mathbb{R}$ . Suposem que la funció característica  $\varphi_\mu$  és  $k$  vegades derivable en un entorn del 0; aleshores  $\mu$  té moments d'ordre  $2n$ , amb  $2n \leq k$ .

**DemOSTRACIÓ:** Volem veure que  $\int_{\mathbb{R}} x^{2n} \mu(dx) < \infty$ . Suposem primer que  $k = 2$  i  $n = 1$ . Tindrem

$$\begin{aligned} \varphi''_\mu(0) &= \lim_{h \rightarrow 0} \frac{\varphi_\mu(h) - 2\varphi_\mu(0) + \varphi_\mu(-h)}{h^2} = \lim_{h \rightarrow 0} \int_{\mathbb{R}} \frac{e^{ihx} - 2 + e^{-ihx}}{h^2} \mu(dx) \\ &= -2 \lim_{h \rightarrow 0} \int_{\mathbb{R}} \frac{1 - \cos hx}{h^2} \mu(dx). \end{aligned}$$

Pel lemma de Fatou

$$\begin{aligned} \int_{\mathbb{R}} x^2 \mu(dx) &= 2 \int_{\mathbb{R}} \liminf_{h \rightarrow 0} \frac{1 - \cos hx}{h^2} \mu(dx) \leq 2 \liminf_{h \rightarrow 0} \int_{\mathbb{R}} \frac{1 - \cos hx}{h^2} \mu(dx) \\ &= -\varphi''_\mu(0) < \infty. \end{aligned}$$

El cas general es demostra per inducció. Suposem que hem demostrat que  $\int_{\mathbb{R}} x^{2(n-1)} \mu(dx) < \infty$  i sabem que  $2n \leq k$ , és a dir  $\varphi_\mu$  és  $2n$  vegades derivable en un entorn del zero. Pel teorema anterior sabem que

$$\varphi_\mu^{(2n-2)}(t) = (-1)^{n-1} \int_{\mathbb{R}} x^{2n-2} e^{itx} \mu(dx).$$

Per tot  $B \in \mathcal{B}(\mathbb{R})$  posem  $\nu(B) = \int_{\mathbb{R}} x^{2n-2} \mu(dx)$ . Aleshores,  $\nu$  és una probabilitat en  $\mathbb{R}$  que té per funció característica

$$\varphi_\nu(t) = \frac{1}{K} (-1)^{k-1} \varphi_\mu^{(2n-2)}(t),$$

on  $K = \int_{\mathbb{R}} x^{2n-2} \mu(dx)$ . Per hipòtesi  $\varphi_\nu''(t)$  existeix en un entorn del zero i utilitzant el resultat en el cas  $k = 2$  obtenim

$$-\varphi_\nu''(0) = \int_{\mathbb{R}} x^2 \nu(dx) = \frac{1}{K} \int_{\mathbb{R}} x^{2n} \mu(dx),$$

és a dir  $\mu$  té un moment finit d'ordre  $2n$ .

Hem suposat que  $K > 0$ . Si fos  $K = 0$ , aleshores  $\mu = \delta_0, \varphi_\mu = 1$  i el teorema és trivial. ■

### 12.3. Càlcul de funcions característiques

(1) *Llei degenerada.* Sigui  $\mu = \delta_a, a \in \mathbb{R}$ . Es immediat comprovar que  $\varphi_{\delta_a}(t) = e^{iat}$ .

(2) *Llei uniforme en el conjunt  $\{-1, 1\}$ .* En aquest cas  $\mu = \frac{1}{2}\delta_{-1} + \frac{1}{2}\delta_1$ . Per tant

$$\varphi_\mu(t) = \frac{1}{2}(e^{it} + e^{-it}) = \cos t.$$

(3) *Llei binomial.* La llei  $\mu$  corresponent a una distribució binomial  $B(r, p)$  pot obtenir-se a partir de  $n$  distribucions independents de Bernoulli  $\mu' = p\delta_1 + q\delta_0$  ( $p+q=1$ ). Aleshores, utilitzant la propietat (9) de la secció anterior s'obté

$$\varphi_\mu(t) = (\varphi_{p\delta_1 + q\delta_0}(t))^n = (pe^{it} + q)^n.$$

(4) *Llei uniforme en un interval  $[a, b]$ .* Mitjançant una integració és fàcil comprovar que

$$\varphi_\mu(t) = \frac{1}{b-a} \int_a^b e^{itx} dx = \frac{e^{itb} - e^{ita}}{it(b-a)}.$$

(6) *Llei de Poisson.* La funció característica d'una llei de Poisson  $\mu$  de paràmetre  $\lambda > 0$  val

$$\int_{\mathbb{R}} e^{-x^2/2} \sin t x dx. \text{ Tindrem}$$

$$\varphi_\mu(t) = \sum_{k=0}^{\infty} e^{-\lambda} e^{itk} \frac{\lambda^k}{k!} = e^{\lambda(e^{it}-1)}.$$

$$\varphi_\mu(t) = (\alpha(t) + i\beta(t)) \frac{1}{\sqrt{2\pi}}.$$

#### 12.4. Fòrmules de inversió

L'objectiu d'aquest apartat és establir fòrmules que permetin de calcular la funció de distribució o la densitat d'una probabilitat a partir de la seva funció característica. El resultat fonamental és el següent.

**Teorema 12.4.** Considerem una probabilitat  $\mu$  sobre  $\mathbb{R}$  amb funció característica  $\varphi$  i funció de distribució  $F$ . Aleshores si  $\alpha < \beta$  són dos punts de continuïtat de  $F$ , es compleix

$$F(\beta) - F(\alpha) = \lim_{\sigma \rightarrow 0} \frac{1}{2\pi} \int_{\mathbb{R}} \varphi(t) e^{-\frac{x^2}{2}} \frac{e^{-it\beta} - e^{-it\alpha}}{-it} dt.$$

#### Observacions

- (1) El pas al límit i el factor  $e^{\sigma^2 t^2/2}$  estan motivades pel fet que la funció  $\varphi$  pot no ser integrable.

- (2) Hi ha altres fòrmules de inversió que s'obtenen per exemple, integrant en un interval  $[-N, N]$  i sent  $N \rightarrow \infty$ , o bé utilitzant altres funcions en lloc de  $e^{-\sigma^2 t^2/2}$ .

*DemOSTRaciÓ del Teorema 12.4:* Sigui  $X$  i  $Y_\sigma$  dues variables aleatòries independents amb lleis  $\mu$  i  $N(0, \sigma^2)$ , respectivament. La funció característica de  $Z_\sigma = X + Y_\sigma$  valdrà

$$\varphi_{Z_\sigma}(t) = \varphi(t) e^{-\sigma^2 t^2/2}.$$

Observem que  $|\varphi_{Z_\sigma}(t)| \leq e^{-\sigma^2 t^2/2}$ , i per tant,  $\varphi_{Z_\sigma}$  és integrable sobre  $\mathbb{R}$  respecte la mesura de Lebesgue. Endemés

$$\int_{\mathbb{R}} e^{-itx} \varphi_{Z_\sigma}(t) dt = \int_{\mathbb{R}} e^{-itx} e^{-\sigma^2 t^2/2} \left( \int_{\mathbb{R}} e^{ity} \mu(dy) \right) dt.$$

Aplicant el teorema de Fubini a la funció  $(t, y) \mapsto \exp(-itx - \frac{\sigma^2 t^2}{2} + ity)$ , que és integrable en  $\mathbb{R}^2$  respecte la mesura producte  $\mu(dy)dt$ , obtenim

$$\begin{aligned} \int_{\mathbb{R}} e^{itx} \varphi_{Z_\sigma}(t) dt &= \int_{\mathbb{R}} \left( \int_{\mathbb{R}} e^{ity(y-x)-\frac{\sigma^2 t^2}{2}} dt \right) \mu(dy) = \\ &= \frac{\sqrt{2\pi}}{\sigma} \int_{\mathbb{R}} e^{-\frac{(y-x)^2}{2\sigma^2}} \mu(dy). \end{aligned} \quad (12.1)$$

En particular, per  $a = -1$ ,  $b = 1$  s'obté  $\varphi_\mu(t) = \frac{\sin t}{t}$ .

(5) *Llei normal.* Sigui  $\mu$  la llei  $N(0, 1)$ . Sigui  $\alpha(t) = \int_{\mathbb{R}} e^{-x^2/2} \cos tx dx$  i  $\beta(t) =$

Integrant per parts obtenim

$$\begin{aligned} \alpha'(t) &= - \int_{\mathbb{R}} x e^{-x^2/2} \sin tx dx = \left[ e^{-x^2/2} \sin tx \right]_{-\infty}^{+\infty} \\ &\quad - \int_{\mathbb{R}} e^{-x^2/2} t \cos tx dx = -i\alpha(t). \end{aligned}$$

Com que  $\alpha(0) = \sqrt{2\pi}$ , això ens diu que  $\alpha(t) = \sqrt{2\pi} e^{-t^2/2}$ , i en conclusió,

$$\varphi(t) = e^{-t^2/2}.$$

Observem que la llei  $N(0, 1)$  és simètrica i en conseqüència la seva funció característica és real. Considerem la variable  $Y = m + \sigma X$ , amb  $m \in \mathbb{R}$  i  $\sigma > 0$ .  $Y$  té per densitat  $f_Y(y) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y-m)^2}{2\sigma^2}}$ . Aplicant la propietat (5) de la secció 12.1 resulta que la funció característica de  $Y$  és  $\varphi_Y(t) = e^{itm - \frac{\sigma^2 t^2}{2}}$ .

La funció característica de la llei  $N(0, 1)$  és infinitament diferenciable. El teorema 12.2 ens diu aleshores que aquesta llei té moments finits de tots els ordres. Podem calcular aquests moments utilitzant les derivades a l'origen de la funció característica. Tenint en compte el desenvolupament

$$e^{-t^2/2} = \sum_{n=0}^{\infty} \frac{(-1)^n t^{2n}}{n! 2^n}$$

$$\varphi^{(k)}(0) = \begin{cases} 0, & \text{si } k = 2n+1 \\ \frac{(-1)^n (2n)!}{n! 2^n}, & \text{si } k = 2n. \end{cases}$$

En conseqüència, els moments de la llei  $N(0, 1)$  valen

$$m_k = \begin{cases} 0, & \text{si } k = 2n+1 \\ \frac{(2n)!}{2^n n!}, & \text{si } k = 2n. \end{cases}$$

ja que  $\frac{\sigma}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{iy-t} - e^{-it} dt$  és el valor de la funció característica de la llei  $N(0, \frac{1}{\sigma^2})$  en el punt  $y - x$ .

Si integrarem respecte de  $x$  en l'interval  $[\alpha, \beta]$  els dos membres de la igualtat (12.1), i apliquem de nou el teorema de Fubini, vindrem

$$\begin{aligned} \frac{1}{2\pi} \int_{\mathbb{R}} \varphi(t) e^{-t^2/\sigma^2/2} \frac{e^{-it\beta} - e^{-ita}}{-it} dt &= \int_{\mathbb{R}} \left( \int_{\alpha}^{\beta} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y-z)^2}{2\sigma^2}} dz \right) \mu(dy) \\ &= \int_{\mathbb{R}} \left( \int_{\alpha-y}^{\beta-y} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-z^2/2\sigma^2} dz \right) \mu(dy) = F_{\sigma}(\beta) - F_{\sigma}(\alpha), \end{aligned}$$

on  $F_{\sigma}$  és la funció de distribució de la variable aleatòria  $Z_{\sigma}$ . En efecte aquesta funció de distribució ve donada per

$$\begin{aligned} F_{\sigma}(\beta) &= P\{X + Y_{\sigma} \leq \beta\} = \int_{\mathbb{R}} \int_{\{y+z \leq \beta\}} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-z^2/2\sigma^2} \mu(dy) = \\ &= \int_{\mathbb{R}} \left( \int_{-\infty}^{\beta-y} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-z^2/2\sigma^2} dz \right) \mu(dy) \end{aligned}$$

Finalment, només cal comprovar que

$$\lim_{\sigma \rightarrow 0} F_{\sigma}(x) = F(x)$$

en tot punt  $x$  de continuïtat de  $F$ . Però això és immediat ja que

$$\lim_{\sigma \rightarrow 0} E(Y_{\sigma}^2) = 0.$$

Per tant  $L^2 - \lim_{\sigma \rightarrow 0} [X + Y_{\sigma}] = X$ , i en conseqüència  $\mathcal{L} - \lim_{\sigma \rightarrow 0} Z_{\sigma} = X$ . Això acaba la demostració del teorema. ■

Com a aplicació d'aquest teorema es pot demostrar la propietat d'injectivitat (6) de l'apartat 12.1 que abans hem provat de forma directa. En efecte, si  $F_1$ , i  $F_2$  són les funcions de distribució de senyals probabilitats en  $\mathbb{R}$ ,  $\mu_1, \mu_2$ , que tenen la mateixa funció característica. Pel teorema anterior vindrem que  $F_1(\beta) - F_1(\alpha) = F_2(\beta) - F_2(\alpha)$ , per a tota parella  $\alpha < \beta$  de punts on les funcions  $F_1$  i  $F_2$  són contínues. Llavors, fent  $\alpha \rightarrow -\infty$  obtenim  $F_1(\beta) = F_2(\beta)$  en un conjunt dens, i per tant  $F_1 = F_2$ .

Si la funció característica és integrable ens podem estalviar el pas al límit en la fórmula de inversió i podem trobar la densitat com a transformada de Fourier inversa de la funció característica. El resultat següent precisa aquest comentari.

**Proposició 12.5.** Si la funció característica  $\varphi$  d'una probabilitat  $\mu$  sobre  $\mathbb{R}$  és integrable respecte la mesura de Lebesgue, aleshores  $\mu$  és absolutament contínua, i la seva densitat és

$$f(x) = \frac{1}{2\pi} \int_{\mathbb{R}} e^{-itx} \varphi(t) dt.$$

*Demostració:* Tenint en compte que

$$|\varphi(t)e^{-\frac{t^2}{2}} \frac{e^{-it\beta} - e^{-ita}}{-it}| \leq |\varphi(t)|(\beta - a),$$

podem aplicar el teorema de convergència dominada en el límit del Teorema 12.4 i obtenim

$$\begin{aligned} F(\beta) - F(a) &= \frac{1}{2\pi} \int_{\mathbb{R}} \varphi(t) \frac{e^{-it\beta} - e^{-ita}}{-it} dt = \frac{1}{2\pi} \int_{\mathbb{R}} \varphi(t) \left( \int_a^{\beta} e^{-itx} dx \right) dt \\ &= \frac{1}{2\pi} \int_a^{\beta} \left( \int_{\mathbb{R}} e^{-itx} \varphi(t) dt \right) dx = \int_a^{\beta} f(x) dx, \end{aligned}$$

de seguit al teorema de Fubini. Això ens dona el resultat que buscavem. ■

### 12.5. Teoremes de continuïtat

En aquest apartat establirem la relació entre la convergència feble de probabilitats estudiada en el capítol 11 i la convergència de les corresponents funcions característiques. En primer lloc demostrarem un resultat auxiliar.

**Proposició 12.6. (Desigualtat de truncació).** Sigui  $\mu$  una probabilitat sobre  $\mathbb{R}$  amb funció característica  $\varphi$ . Aleshores, per a tot  $a > 0$ .

$$-\quad - \quad - \quad - \quad - \quad \frac{1}{a} \int_{-a}^a (1 - \varphi(t)) dt \geq \mu\{|x| \geq \frac{2}{a}\}.$$

*Demostració:* Utilitzant el teorema de Fubini i el fet que la funció sinus és impar, vindrem

$$\begin{aligned} \frac{1}{a} \int_{-a}^a (1 - \varphi(t)) dt &= \frac{1}{a} \int_{-a}^a \left( \int_{\mathbb{R}} (1 - e^{itx}) \mu(dx) \right) dt \\ &= \frac{1}{a} \int_{\mathbb{R}} \left( \int_{-a}^a (1 - \cos tx) dt \right) \mu(dx) \\ &= \int_{\mathbb{R}} \left( 1 - \frac{\sin ax}{ax} \right) \mu(dx) \\ &\geq \frac{2}{a} \inf_{|u| \geq 2} \left( 1 - \frac{\sin u}{u} \right) \int_{\{|x| \geq 2\}} \mu(dx) \\ &\geq \mu\{|x| \geq \frac{2}{a}\}, \end{aligned}$$

ja que  $\left| \frac{\sin u}{u} \right| \leq 1$  i  $\left| \frac{\sin u}{u} \right| \leq \frac{1}{2}$  si  $|u| \geq 2$ . El resultat queda així demostrat. ■

**Teorema 12.7.** (Teorema de continuïtat de P. Lévy). Considerem una successió  $\{\mu_n, n \geq 1\}$  de probabilitats en  $\mathbb{R}$  i denotem per  $\{\varphi_n, n \geq 1\}$  la successió de funcions característiques associades. Aleshores,

- Si  $\mu_n$  convergeix feblement cap a una probabilitat  $\mu$  quan  $n$  tendeix a infinit, aleshores  $\lim_n \varphi_n(t) = \varphi(t)$ , per a tot  $t \in \mathbb{R}$ .
- Si  $\lim_n \varphi_n(t) = \varphi(t)$ , on  $\varphi$  és una funció contínua en el zero, aleshores  $\varphi$  és la funció característica d'una probabilitat  $\mu$ , i  $w - \lim_{n \rightarrow \infty} \mu_n = \mu$ .

*Demonstració:* L'apartat (i) és una conseqüència immediata de la definició de convergència feble, ja que les funcions  $\cos tx$ ,  $i \sin tx$ , són continues i afitades.

Per a la demostració de (ii) utilitzarem la desigualtat de truncació (Proposició 12.6) i el teorema de convergència dominada (ja que  $|\varphi_n(t)| \leq 1$ ). Obtenim així

$$\mu_n\{x : |x| \geq \frac{2}{a}\} \leq \frac{1}{a} \int_{-a}^a (1 - \varphi_n(t)) dt.$$

Aquesta última expressió convergeix cap a  $\frac{1}{a} \int_{-a}^a (1 - \varphi(t)) dt$  quan  $n$  tendeix a infinit.

D'altra banda,

$$\lim_{a \rightarrow 0} \frac{1}{a} \int_{-a}^a (1 - \varphi(t)) dt = 2(1 - \varphi(0)) = 0,$$

ja que  $\varphi$  és contínua en el zero.

Fixem un  $\varepsilon > 0$  i signi  $a > 0$  tal que  $0 \leq \frac{1}{a} \int_{-a}^a (1 - \varphi(t)) dt \leq \frac{\varepsilon}{2}$ . A partir d'un cert  $n_0$  tindrem

$$\begin{aligned} \mu_n\left(\left[-\frac{2}{a}, \frac{2}{a}\right]\right) &\leq \mu_n\{x : |x| \geq \frac{2}{a}\} \leq \frac{1}{a} \int_{-a}^a (1 - \varphi_n(t)) dt \leq \\ &\leq \frac{1}{a} \int_{-a}^a (1 - \varphi(t)) dt + \frac{\varepsilon}{2} \leq \varepsilon. \end{aligned}$$

D'altra banda, per  $1 \leq n \leq n_0$  podem trobar  $a_n > 0$  tals que  $\mu_n([ -a_n, a_n ]^c) \leq \varepsilon$ . En conclusió, si  $b = \max\{a_1, \dots, a_{n_0}, \frac{2}{a}\}$ , es compleix que  $\sup_n \mu_n([ -b, b ]^c) \leq \varepsilon$ , i això ens diu que la successió de probabilitats  $\{\mu_n, n \geq 1\}$  és ajustada.

Tenint en compte el criteri de convergència feble (Teorema 11.9), per a demostrar que  $\mu_n$  convergeix feblement cap a  $\mu$  només cal comprovar que totes les subsuccessions convergents tenen el mateix límit, igual a  $\mu$ . Suposem que  $\{\mu_{n_i}, i \geq 1\}$  és una subsuccessió convergent cap a una probabilitat  $\nu$ . Tindrem, per la part (i),

$$\varphi_\nu(t) = \lim \varphi_{n_i}(t) = \varphi(t).$$

Això ens diu que  $\varphi$  és la funció característica d'una probabilitat (ja que sempre podem trobar una subsuccessió feblement convergent) que designarem per  $\mu = \nu$  i totes les parcials convergents ténen límit  $\mu$ . Acaba així la demostració del teorema. ■

**Observació:** En l'observació (2) que segueix al Teorema 11.2 varem indicar que una condició suficient per a la convergència feble d'una successió de probabilitats és que convergeixim les integrals de les funcions de la classe  $C_b^\infty(\mathbb{R})$  (infinitament diferenciables i amb totes les derivades afitades). Aquesta família de funcions és molt més petita que el conjunt de les funcions continues i afitades que apareix a la definició de convergència feble. El teorema de continuïtat de P. Lévy que acabem de demostrar ens diu que una condició suficient per a la convergència feble (part (ii) del teorema) és la convergència de les integrais de les funcions  $\sin tx$ ,  $\cos tx$ ,  $t \in \mathbb{R}$ . Aquestes funcions trigonomètriques constitueixen una família de funcions encara molt més petita que  $C_b^\infty(\mathbb{R})$ .

De l'estudi que acabem de fer de les funcions característiques se'n desprenden els fets següents:

- 1) Es equivalent conèixer la funció característica d'una probabilitat  $\mu$  que la pròpia  $\mu$  (vegi's la fórmula de inversió).
- 2) Les funcions característiques proporcionen un tractament senzill de la propietat de independència (vegi's propietat (8) de l'apartat 12.1).
- 3) Les funcions característiques proporcionen una eina adequada pel tractament de la convergència feble (vegi's Teorema 12.7).

En el Llenguatge de l'Anàlisi, la funció característica d'una probabilitat  $\mu$  és la transformada de Fourier de la mesura finita  $\mu$ . Com hem dit al començament del capítol, s'ou Paul Lévy qui a començaments d'aquest segle va introduir la tècnica de la transformada de Fourier en l'estudi del problema del límit central.

De fet idees similars havien estat utilitzades al llarg del segle XIX. Per exemple, el que avui coneixem com a transformada de Laplace, fou una tècnica desenvolupada per Laplace amb el mateix objectiu.

Acabem aquest capítol amb una breu secció destinada a la transformada de Laplace per a probabilitats discretes.

### 13. EL TEOREMA CENTRAL DEL LÍMIT

**12.6. Funció generatriu**  
 En el cas de lleis de probabilitat discretes, és de vegades més interessant utilitzar la funció generatriu que la funció característica. Considerem una probabilitat  $\mu$  sobre el conjunt dels nombres naturals  $\mathbb{N} = \{0, 1, 2, \dots\}$ . La probabilitat  $\mu$  ve caracterizada per una successió  $p_i = \mu(\{i\})$  tal que  $p_i \geq 0$  i  $\sum_{i=0}^{\infty} p_i = 1$ .

Observem que la sèrie  $\sum_{n=0}^{\infty} p_n z^n$  és convergent per a tot complex  $z$  tal que  $|z| \leq 1$ , és a dir, el radi de convergència d'aquesta sèrie serà més gran o igual a 1 i la sèrie convergeix uniformement en tot disc  $\{z \in \mathbb{C} : |z| \leq r\}$  de radi  $r < 1$ .

La funció  $G_{\mu}(z) = \sum_{n=0}^{\infty} p_n z^n$  definida en el disc  $\{z \in \mathbb{C} : |z| \leq 1\}$  s'anomena la funció generatriu de  $\mu$ . Aquesta funció és analítica en el disc obert  $\{z \in \mathbb{C} : |z| < 1\}$  i determina la probabilitat  $\mu$ . Més exactament, les probabilitats  $p_n$  són els coeficients del desenvolupament en sèrie de potències de  $G_{\mu}$  a l'origen:

$$p_n = \frac{1}{n!} G_{\mu}^{(n)}(0).$$

La relació entre la funció generatriu i la funció característica ve donada per la següent fórmula:

$$G_{\mu}(e^{it}) = \varphi_{\mu}(t), \quad \text{per a tot } t \in \mathbb{R}.$$

Com en el cas de les funcions característiques, si  $X$  és una variable aleatòria a valors naturals, s'anomena funció generatriu de  $X$  a la funció generatriu de la seva llei, és a dir,

$$G_X(z) = E(z^X).$$

Moltes de les propietats que hem estudiat en el cas de les funcions característiques es compleixen també per les funcions generatríus. Per exemple, si  $X$  i  $Y$  són dues variables aleatòries independents a valors naturals, tindrem que  $G_{X+Y} = G_X \cdot G_Y$ . En efecte,

$$G_{X+Y}(z) = E(z^{X+Y}) = E(z^X)E(z^Y) = G_X(z)G_Y(z).$$

Considerem una successió  $\{X_n, n \geq 1\}$  de variables aleatòries independents i amb lleis de Bernoulli de paràmetre  $p$ . Posem  $S_n = X_1 + \dots + X_n$ . Les  $\frac{S_n}{n}$  prenen els valors  $\frac{k}{n}, 0 \leq k \leq n$  amb probabilitat

$$P\left\{\frac{S_n}{n} = \frac{k}{n}\right\} = \binom{n}{k} p^k (1-p)^{n-k}.$$

Com que  $\sigma^2\left(\frac{S_n}{n}\right) = \frac{p(1-p)}{n}$  tendeix a zero quan  $n \rightarrow \infty$ , la distribució de probabilitat de  $\frac{S_n}{n}$  tendeix a concentrar-se en el punt  $p = E\left(\frac{S_n}{n}\right)$ . La llei fortia dels grans nombres ens diu que  $\lim_{n \rightarrow \infty} \frac{S_n}{n} = p$  quasi segurament i també en  $L^2$ .

Una altra idea per tal d'estudiar el comportament asymptòtic de la successió  $S_n$  consisteix en fer una normalització. És a dir, definim

$$S'_n = \frac{S_n - E(S_n)}{\sigma(S_n)} = \frac{S_n - np}{\sqrt{np(1-p)}}.$$

Les variables  $S'_n$  tenen mitjana zero i variància 1. En aquest cas, la successió  $S'_n$  no convergirà en probabilitat o quasi segurament però podem esperar que convergeixi en llei ja que la seva distribució manté un valor mig i una variància constants. La llei de  $S'_n$  és una distribució binomial normalitzada. En el llibre "The Doctrine of Chances" de A. De Moivre publicat l'any 1756, es va demostrar que si  $p = \frac{1}{2}$ ,  $S'_n$  convergeix en llei cap a una distribució normal  $N(0, 1)$ , és a dir,

$$\lim_n P\left\{a \leq \frac{S_n - np}{\sqrt{np(1-p)}} \leq b\right\} = \int_a^b \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx. \quad (13.1)$$

La funció de distribució de la llei normal  $F(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy$  es pot trobar a les taules, i aleshores una probabilitat del tipus  $P\{\alpha \leq S_n \leq \beta\}$  es podrà aproximar pel increment

$$F\left(\frac{\beta - np}{\sqrt{np(1-p)}}\right) - F\left(\frac{\alpha - np}{\sqrt{np(1-p)}}\right).$$

Dit d'una altra manera, la llei binomial  $B(n, p)$  es pot aproximar per una llei normal  $N(np, np(1-p))$ . En general, acceptarem aquesta aproximació quan  $np(1-p) > 15$ .

La taula següent (W. Feller, vol. 1) ens ofereix una comparació de la distribució binomial  $B(100, 0.3)$  amb la seva aproximació normal

Sabem que  $\varphi_n$  és dues vegades derivable amb continuïtat i

$$\begin{aligned}\varphi'_n(0) &= i E\left(\frac{X_1 - m}{\sigma \sqrt{n}}\right) = 0, \\ \varphi''_n(t) &= - \int_{\mathbb{R}} x^2 e^{itx} \left[ P_o\left(\frac{X_1 - m}{\sigma \sqrt{n}}\right)^{-1} \right] (dx) = -E\left(\left(\frac{X_1 - m}{\sigma \sqrt{n}}\right)^2 \exp(it \frac{X_1 - m}{\sigma \sqrt{n}})\right).\end{aligned}$$

En particular,

$$\varphi''_n(0) = -E\left(\left(\frac{X_1 - m}{\sigma \sqrt{n}}\right)^2\right) = -\frac{1}{n}.$$

La fórmula de Taylor aplicada a la funció  $\varphi_n(t)$  en el punt  $t = 0$  ens diu que

$$\varphi_n(t) = 1 - \frac{t^2}{2n} + \frac{t^2}{2} [\varphi''_n(\theta t) - \varphi''_n(0)],$$

on  $|\theta| \leq 1$ . llavors,

$$n[\varphi''_n(\theta t) - \varphi''_n(0)] = -n E\left[\left(\frac{X_1 - m}{\sigma \sqrt{n}}\right)^2 \left(e^{it\theta \frac{X_1 - m}{\sigma \sqrt{n}}} - 1\right)\right],$$

$$= -\frac{1}{\sigma^2} E[(X_1 - m)^2 (e^{it\theta \frac{X_1 - m}{\sigma \sqrt{n}}} - 1)],$$

Pel teorema de convergència dominada, aquesta expressió tendeix a zero quan  $n \rightarrow \infty$ . En efecte,

$$\lim_{n \rightarrow \infty} (X_1 - m)^2 \left( e^{it\theta \frac{X_1 - m}{\sigma \sqrt{n}}} - 1 \right) = 0,$$

per a tot  $t \in \mathbb{R}$ .  
Pel teorema de continuïtat de P. Lévy resulta que la successió  $\{Y_n, n \geq 1\}$  convergeix en llei cap a una variable aleatòria amb distribució  $N(0, 1)$ , que és el resultat que volem demostrar. ■

És interessant poder disposar d'una versió  $n$ -dimensional del teorema central del límit de Lévy-Lindeberg, per les seves aplicacions a l'Estadística. Per això, cal introduir primer la *llei normal multidimensional*.

	Probabilitat	Aproximació normal	Percentatge d'error
$0 \leq S_n \leq 11$	0.000 006	0.000 03	+ 400
$12 \leq S_n \leq 14$	0.000 15	0.000 33	+ 100
$15 \leq S_n \leq 17$	0.002 01	0.002 83	+ 40
$18 \leq S_n \leq 20$	0.014 30	0.015 99	+ 12
$21 \leq S_n \leq 23$	0.059 07	0.058 95	0
$24 \leq S_n \leq 26$	0.148 87	0.144 47	- 3
$27 \leq S_n \leq 29$	0.237 94	0.234 05	- 2
$31 \leq S_n \leq 33$	0.230 13	0.234 05	+ 2
$34 \leq S_n \leq 36$	0.140 86	0.144 47	+ 3
$37 \leq S_n \leq 39$	0.058 89	0.058 95	0
$40 \leq S_n \leq 42$	0.017 02	0.015 99	- 6
$43 \leq S_n \leq 45$	0.003 43	0.002 83	- 18
$46 \leq S_n \leq 48$	0.000 49	0.000 33	- 33
$49 \leq S_n \leq 51$	0.000 05	0.000 03	- 40

El resultat de De Moivre va ésser generalitzat per Laplace (1810) al cas de variables discretes i simètriques. Constitueix la versió més senzilla del *teorema central del límit*. A continuació donarem una versió del *teorema central de límit* deguda a Lévy-Lindeberg que generalitza el teorema de De Moivre-Laplace.

**Teorema 13.1** (*Teorema central del límit de Lévy-Lindeberg*). Sigui  $\{X_n, n \geq 1\}$  una successió de v.a. independents, idènticament distribuïdes i de quadrat integrable, amb mitjana  $m$  i variància  $\sigma^2$ . Si  $S_n = X_1 + \dots + X_n$ , aleshores

$$\frac{S_n - nm}{\sigma \sqrt{n}} \xrightarrow[n \rightarrow \infty]{\mathcal{L}} N(0, 1).$$

**Demostració:** Sigui  $Y_n = \frac{S_n - nm}{\sigma \sqrt{n}}$ . Observem que  $E(Y_n) = 0$  i  $\sigma^2(Y_n) = 1$ . Calclem la funció característica de la variable  $Y_n$ ,

$$\begin{aligned}E[e^{itY_n}] &= E\left[\exp\left(it \frac{S_n - nm}{\sigma \sqrt{n}}\right)\right] = E\left[\exp i \left(\sum_{j=1}^n \frac{X_j - m}{\sigma \sqrt{n}}\right)t\right] \\ &= \left(E\left[\exp i \left(\frac{X_1 - m}{\sigma \sqrt{n}}\right)t\right]\right)^n = \varphi_n(t)^n,\end{aligned}$$

on  $\varphi_n$  és la funció característica de la variable  $\frac{\sum_{j=1}^n X_j - nm}{\sigma \sqrt{n}}$ .

**Proposició 13.2.** Sigui  $m \in \mathbb{R}^n$  i  $\Lambda$  una matriu simètrica d'ordre  $n$  definida no negativa. Existix una probabilitat en  $\mathbb{R}^n$ , que designarem per  $N(m, \Lambda)$  i anomenarem llei normal  $n$ -dimensional, que té per funció característica

$$\varphi(t) = \exp\left(it^*m - \frac{1}{2}t^*\Lambda t\right),$$

per a tot  $t \in \mathbb{R}^n$ .

**DemOSTació:** Sigui  $C$  una matriu ortogonal tal que  $C\Lambda C^* = D$ , on  $D$  és una matriu diagonal. Els elements de la diagonal de  $D$ , que designarem per  $\lambda_1, \dots, \lambda_n$ , coincideixen amb els valors propis de  $\Lambda$  i en conseqüència són no negatius. Tindrem  $\Lambda = C^*DC$ .

Considerem un vector aleatori  $Y = (Y_1, \dots, Y_n)$  amb les components independents i amb lleis  $N(0, \lambda_i)$  si  $\lambda_i \neq 0$  i  $Y_i = 0$  si  $\lambda_i = 0$  ( $i = 1, \dots, n$ ). Definim  $X = C^*Y + m$ .

La funció característica del vector  $Y$  val

$$\varphi_Y(t) = E\left[\exp\left(i\sum_{j=1}^n t_j Y_j\right)\right] = \prod_{j=1}^n e^{-\frac{1}{2}t_j^2 \lambda_j} = e^{-\frac{1}{2}t^*D t}.$$

Per tant, la funció característica del vector  $X$  serà

$$\varphi_X(t) = e^{it^*m} \varphi_Y(Ct) = e^{it^*m} e^{-\frac{1}{2}(Ct)^*DCt} = e^{it^*m - \frac{1}{2}t^*\Lambda t},$$

i la llei del vector  $X$  és la probabilitat que busquem. ■

### Propietats de les lleis normals multidimensionals

1.- Sigui  $X$  un vector amb llei  $N(m, \Lambda)$ . Si  $C$  és una matriu ortogonal tal que  $\Lambda = C^*DC$  (on  $D$  és la matriu diagonal que hem utilitzat abans), el vector  $Y = C(X - m)$  té les components independents i amb lleis  $N(0, \lambda_i)$  si  $\lambda_i \neq 0$  o bé zero si  $\lambda_i = 0$ .

En efecte,

$$\begin{aligned}\varphi_Y(t) &= e^{it^*(-Cm)} \varphi_X(C^*t) = \exp\left(-it^*Cm + it^*Cm - \frac{1}{2}t^*C\Lambda C^*t\right) \\ &= \exp\left(-\frac{1}{2}t^*Dt\right).\end{aligned}$$

2.- Si  $X$  és un vector aleatori amb llei  $N(m, \Lambda)$ , aleshores  $m$  i  $\Lambda$  representen, respectivament, el vector de mitjanes i la matriu de variàncies i covariàncies del vector  $X$ .

En efecte, fent servir les mateixes notacions que abans, tindrem

$$\begin{aligned}E(X) &= C^*E(Y) + m = m \\ E[(X - m)(X - m)^*] &= E[C^*Y Y^* C] = C^*DC = \Lambda.\end{aligned}$$

3.- Sigui  $X$  un vector aleatori  $n$ -dimensional amb llei  $N(m, \Lambda)$ . Aleshores, si  $A$  és una matriu d'ordre  $r \times n$ , el vector  $AX$  té llei  $N(Am, A\Lambda A^*)$ .

En efecte,

$$\begin{aligned}\varphi_{AX}(t) &= \varphi_X(A^*t) = \exp\left(it^*Am - \frac{1}{2}(A^*t)^*\Lambda(A^*t)\right) \\ &= \exp\left(it^*(Am) - \frac{1}{2}t^*(A\Lambda A^*)t\right),\end{aligned}$$

per a tot  $t \in \mathbb{R}^r$ .

En particular, tot vector aleatori de la forma  $(X_1, \dots, X_m)$ , amb  $m \leq n$ , té llei normal, i tota combinació lineal  $\sum_{i=1}^m a_i X_i$  té llei normal.

Recíprocament, si  $X = (X_1, \dots, X_n)$  és un vector aleatori tal que tota combinació lineal  $\sum_{i=1}^m a_i X_i$  és normal, aleshores  $X$  té una llei normal multidimensional. En efecte, com que

$$E(e^{it^*X}) = \varphi_X(1) = e^{-\frac{1}{2}\sigma^2(t^*X) + E(t^*X)},$$

tindrem

$$E(t^*X) = t^*E(X),$$

$$\begin{aligned}\sigma^2(t^*X) &= E\left[\left(t^*(X - E(X))\right)^2\right] \\ &= E\left[t^*(X - E(X))(X - E(X))^*t\right] = t^*\Lambda t,\end{aligned}$$

on  $\Lambda$  és la matriu de variàncies i covariàncies del vector  $X$ .

Per tant,  $X$  té una llei normal multidimensional amb paràmetres  $m = E(X)$  i  $\Lambda$ .

4.- Sigui  $X = (X_1, \dots, X_n)$  un vector amb llei normal  $n$ -dimensional  $N(m, \Lambda)$ . La independència de les variables  $X_1, \dots, X_n$  és equivalent a que la matriu  $\Lambda$  sigui diagonal, és a dir. les variables  $X_1, \dots, X_n$  són incorrelacionades.

En efecte, la independència de les variables  $X_1, \dots, X_n$  implica que  $cov(X_i, X_j) = 0$ , per a  $i \neq j$ .

Recíprocament, si la matriu  $\Lambda$  és diagonal i designem per  $\lambda_1, \dots, \lambda_n$  els elements de la diagonal, tindrem

$$\varphi_X(t) = e^{it^*m - \frac{1}{2}t^*\Lambda t} = \prod_{j=1}^n e^{it_j m_j - \frac{1}{2}t_j^2} = \prod_{j=1}^n \varphi_{X_j}(t_j),$$

i això ens diu que les variables  $X_1, \dots, X_n$  són independents.

5.- Si la matriu  $\Lambda$  és regular (és a dir,  $\det \Lambda > 0$ ), direm que la llei normal  $N(m, \Lambda)$  és no degenerada.

En aquest cas,  $\lambda_i > 0$ ,  $i = 1, \dots, n$  i aquesta llei és absolutament contínua amb una densitat que es pot calcular mitjançant un canvi de variable. En efecte amb les notacions anteriors definim  $\varphi(y) = C^*y + m = x$ . Com que  $X = \varphi(Y)$ , tindrem

$$\begin{aligned} f_X(x) &= f_Y(\varphi^{-1}(x)) |J_\varphi(y)|^{-1} = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\lambda_i}} e^{-\frac{1}{2\lambda_i}y_i^2} \\ &= [(2\pi)^n \det \Lambda]^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(x-m)^* \Lambda^{-1} (x-m)\right). \end{aligned}$$

Per a justificar aquesta darrera igualtat cal tenir en compte que  $|J_\varphi(y)| = |\det C^*| = 1$ , i  $\lambda_1 \dots \lambda_n = \det \Lambda$ . Endemés

$$y^* D^{-1} y = (x-m)^* C^*(x-m) = (x-m)^* \Lambda (x-m).$$

Si la matriu  $\Lambda$  és singular, direm que la llei  $N(m, \Lambda)$  és degenerada. Suposem que el rang de  $\Lambda$  és  $r < n$  i siguin  $\lambda_1, \dots, \lambda_r$  els vectors propis no nuls de  $\Lambda$ . Aleshores  $y_{r+1} = \dots = y_n = 0$  i el vector  $Z = (Y_1, \dots, Y_r)$  té una llei normal  $r$ -dimensional no degenerada.

Signi  $W_1$  la matriu d'ordre  $n \times r$  formada per les  $r$  primeres columnes de la matriu  $C^*$ . Es compleix que  $X = W_1 Z + m$  i en conseqüència, la distribució de probabilitat induïda per  $X$  sobre  $\mathbb{R}^n$  ( $X$  és un vector aleatori amb llei  $N(m, \Lambda)$ ) està concentrada en la següent varietat lineal de dimensió  $r$

$$\{x = W_1 z + m, z \in \mathbb{R}^r\} = \{x = C^*y + m, z \in \mathbb{R}^r\}.$$

6.- Analitzem el cas de dimensió 2. Signi  $(X, Y)$  un vector aleatori amb llei normal. Suposem que  $X$  i  $Y$  tenen lleis  $N(m_1, \sigma_1^2)$ ,  $N(m_2, \sigma_2^2)$ , respectivament i sigui  $\rho$  el seu coeficient de correlació. Aleshores

$$\Lambda = \begin{pmatrix} \sigma_1^2 & \rho \sigma_1 \sigma_2 \\ \rho \sigma_1 \sigma_2 & \sigma_2^2 \end{pmatrix}$$

i la llei de  $(X, Y)$  és no degenerada si i només si  $|\rho| < 1$ . En aquest cas, la densitat del vector  $(X, Y)$  és

$$\begin{aligned} f_{(X,Y)}(x, y) &= \frac{1}{2\pi \sigma_1 \sigma_2 \sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)} \left[ \frac{(x-m_1)^2}{\sigma_1^2} \right.\right. \\ &\quad \left.\left. - 2 \frac{(x-m_1)(y-m_2)}{\sigma_1 \sigma_2} \rho + \frac{(y-m_2)^2}{\sigma_2^2} \right]\right\}. \end{aligned}$$

Acabarem el capítol amb la versió anunciatada del teorema central del límit multidimensional.

**Teorema 13.3.** Sigui  $\{X_n, n \geq 1\}$  una successió de vectors aleatoris  $k$ -dimensionals, independents i idènticament distribuïts. Posem  $S_n = X_1 + \dots + X_n$ . Suposem que les components de  $X_1$  són de quadrat integrable i posem  $E(X_1) = m$ ,  $E[(X_1 - m)(X_1 - m)^*] = \Lambda$ . Aleshores

$$\frac{S_n - n m}{\sqrt{n}} \xrightarrow{\mathcal{L}} N(0, \Lambda),$$

quan  $n$  tendeix a infinit.

**Demostració:** Posem  $Y_n = \frac{S_n - n m}{\sqrt{n}}$ . Fixem  $t \in \mathbb{R}^k$ . Aleshores  $\{t^* Y_n, n \geq 1\}$  és una successió de variables aleatories independents, idènticament distribuïdes, amb mitjana  $t^* m$  i variància

$$\sigma^2(t^* X_1) = E[t^*(X_1 - m)(X_1 - m)^*] = t^* \Lambda t.$$

En conseqüència, el teorema central del límit en dimensió  $u$  ens diu que

$$t^* Y_n = \frac{1}{\sqrt{n}} \left[ \sum_{j=1}^n t^* X_j - n t^* m \right] \xrightarrow[n \rightarrow \infty]{\mathcal{L}} N(0, t^* \Lambda t).$$

Per tant,

$$\varphi_{Y_n}(t) = E[e^{it^* Y_n}] = \varphi_{t^* Y_n}(1) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \varphi_{N(0, t^* \Lambda t)}(1) = e^{-\frac{1}{2} t^* \Lambda t} = \varphi_{N(0, \Lambda)}(t).$$

És a dir, la funció característica de  $Y_n$  convergeix cap a la funció característica de la llei  $N(0, \Lambda)$ , en tot punt  $t \in \mathbb{R}^k$ . La versió multidimensional del teorema de continuïtat de P. Lévy implica aleshores que  $Y_n \xrightarrow[n \rightarrow \infty]{\mathcal{L}} N(0, \Lambda)$  i el teorema queda demostrat. ■

La versió multidimensional del resultat de De Moivre ens porta a introduir la llei multínomial. Creiem interessant presentar aquest cas particular del Teorema 13.3 per la seva relació amb el fonament teòric del test de la  $\chi^2$ .

Considerem una experiència aleatòria amb  $k$  resultats possibles,  $A_1, \dots, A_k$ , de probabilitats  $p_1, \dots, p_k$ , respectivament ( $\sum_{j=1}^k p_j = 1$ ). Repetim aquesta experiència aleatòria  $n$  vegades de manera independent, i denotem per  $X_i$ ,  $i = 1, \dots, n$  a la variable aleatoria que ens dóna el nombre de vegades que s'ha produït el resultat  $A_i$  en aquesta repetició de l'experiència. El vector aleatori  $X = (X_1, \dots, X_k)$  té una distribució de probabilitat discreta donada per

$$P\{X_1 = n_1, \dots, X_k = n_k\} = \frac{n!}{n_1! \dots n_k!} p_1^{n_1} \dots p_k^{n_k},$$

on  $n_i \in \mathbb{N}$ ,  $i = 1, \dots, k$  i  $\sum_{i=1}^k n_i = n$ .

Direm que la distribució del vector  $X$  és una *multinomial* amb paràmetres  $n, k, p_1, \dots, p_k$ , i escriurem  $M(n; k, p_1, \dots, p_k)$ . Observi's que per  $k = 2$  obtenim la llei  $B(n, p)$  on  $p = p_1$ . Les variables aleatories  $X_i$ ,  $i = 1, \dots, k$  poden expressar-se com  $\sum_{j=1}^n X_{ij}$ , on  $X_{ij} = 1_{E_{ij}}$ , essent  $E_{ij}$  l'esdeveniment "obtenir el resultat  $A_i$  en la  $j$ -essima experiència". Així doncs  $X_i$ ,  $i = 1, \dots, k$ , és una suma de variables aleatories independents i idènticament distribuïdes amb valor mig igual a  $p_i$ . És fàcil comprovar que la matríg de variancias-covariancias de  $X$  és  $n \Lambda$  amb

$$\Lambda = \begin{pmatrix} p_1 - p_1^2 & -p_1 p_2 & \dots & -p_1 p_k \\ \dots & \dots & \dots & \dots \\ -p_1 p_k & -p_2 p_k & \dots & p_k - p_k^2 \end{pmatrix}.$$

Aleshores, el Teorema 13.3 estableix que el vector aleatori  $Y_n = (Y_1^n, \dots, Y_k^n)$ , on  $Y_i^n = \frac{X_i - np_i}{\sqrt{n}}$ ,  $i = 1, \dots, k$ , convergeix en llei quan  $n$  tendeix a infinit cap a un vector aleatori amb distribució  $N(0, \Lambda)$ .

#### Referències

1. Ash, R. Basic Probability Theory. Wiley, New York, 1970.
2. Ash, R. Real Analysis and Probability. Academic Press, New York, 1972.
3. Billingsley, P. Probability and Measure. Wiley, New York, 1979.
4. Chung, K. L. A Course in Probability Theory. Academic Press, New York, 1974.
5. Cramér, H. Métodos matemáticos en Estadística. Aguilar, Madrid, 1970.
6. Daucunha, D. et Duflo, M. Probabilités et Statistiques. Vol. I. Masson, Paris, 1982.
7. Dudley, R. Real Analysis and Probability. Wadsworth and Brooks/Cole, Belmont, California, 1989.
8. Feller, W. Introducción a la Teoría de Probabilidades y sus aplicaciones. Vol. I. Limusa, Mexico, 1975.
9. Métivier, M. Notions fondamentales de la théorie des Probabilités. Dunod, Paris, 1972.