

Fabrizio Bernardi  
Gioele Cerri  
Alessandra Di Nardo  
Gabriele Gabrielli  
Bruno Guindani  
Simone Polito  
Aron Wussler

# APPUNTI DI PROBABILITÀ

DALLE LEZIONI DEL PROF. MATTEO GREGORATTI

**fubini**  **tonelli**

© Gli autori. Alcuni diritti riservati

Questa opera è rilasciata sotto licenza Creative Commons BY-NC-SA 4.0.

<https://creativecommons.org/licenses/by-nc-sa/4.0/>

In particolare, senza il permesso degli autori non è consentito fare copie di questo libro (né cartacee né digitali) per rivenderle.

Il codice sorgente  $\text{\LaTeX}$  è disponibile su

<https://github.com/fubinitonelli/probabilita>

fubinitonelli.it - [info@fubinitonelli.it](mailto:info@fubinitonelli.it)

REVISIONE DEL 12 FEBBRAIO 2022 - d13df6f1f2009aff2c6475b6d59984782d7b5aae

DEVELOPED WITH GIT  BY:

FABRIZIO BERNARDI - [fabrizio@fubinitonelli.it](mailto:fabrizio@fubinitonelli.it)

GIOELE CERRI - [gioele@fubinitonelli.it](mailto:gioele@fubinitonelli.it)

ALESSANDRA DI NARDO - [alessandra@fubinitonelli.it](mailto:alessandra@fubinitonelli.it)

GABRIELE GABRIELLI - [gabriele@fubinitonelli.it](mailto:gabriele@fubinitonelli.it)

BRUNO GUINDANI - [bruno@fubinitonelli.it](mailto:bruno@fubinitonelli.it)

SIMONE POLITO - [simone@fubinitonelli.it](mailto:simone@fubinitonelli.it)

ARON WUSSLER - [aron@fubinitonelli.it](mailto:aron@fubinitonelli.it)

## Prefazione

Questo libro raccoglie gli appunti del corso di Probabilità per alunni ingegneri matematici, tenuto presso il Politecnico di Milano nell'anno accademico 2016-2017 dal prof. Matteo Gregoratti, con la collaborazione del dott. Andrea Cosso per le esercitazioni.

Abbiamo cominciato a redigere le prime lezioni a computer quasi per caso, spinti prima di tutto dalla curiosità e dalla voglia di sperimentare. Ben presto ci siamo resi conto che questo metodo ci consentiva di collaborare in maniera molto semplice: ciò che sarebbe stato troppo gravoso se fatto da uno solo di noi, ora diventava fattibile.

Settimana dopo settimana, il nostro progetto ha preso forma, fino a diventare la nostra prima edizione: un piccolo manuale di istruzioni alla probabilità, da studenti per studenti. Scrivendolo abbiamo avuto l'opportunità di poter capire fino in fondo alcuni degli aspetti meno intuitivi della materia; abbiamo dunque cercato di trasmettere in maniera chiara ciò che abbiamo colto, stimolando l'intuizione di voi lettori.

Speriamo che il nostro volume vi sia di aiuto nei vostri studi almeno quanto lo è stato per noi. Realizzarlo è stata un'avventura bellissima e istruttiva: per questo vogliamo ringraziare i compagni di corso che ci hanno stimolato e motivato ad andare avanti.

Ringraziamo Carlo Vitellio, *il tuo candidato per il CCS di Matematica*, cui contributo è stato di vitale importanza nella creazione del formulario e nella revisione finale dell'opera. Ringraziamo Gregorattibot, il fedele bot che ha compilato gli appunti centinaia di volte, rendendo sempre disponibile una versione aggiornata su Dropbox.

Ringraziamo infine il prof. Matteo Gregoratti e il dott. Andrea Cosso, perché ci hanno contagiato con la loro passione, ci hanno fatto sudare e faticare, ci hanno chiesto il massimo, riuscendoci a far apprezzare la materia e con qualche stregoneria trasformando un corso da dieci crediti in uno da quindici.

*Gli autori*

### **Nota di redazione**

L'errata corrige è disponibile sul nostro sito: [www.fubinitonelli.it/probabilita](http://www.fubinitonelli.it/probabilita);  
per segnalare refusi e inesattezze potete contattarci a [info@fubinitonelli.it](mailto:info@fubinitonelli.it).

Nell'intestazione di teoremi e corollari sono stati indicati i riferimenti al libro di testo del corso: Jean Jacod, Philip Protter, *Probability Essentials*, Springer-Verlag, Berlin 2004.

# Indice

<b>1</b>	<b>Definizione assiomatica di probabilità</b>	<b>9</b>
1.1	Eventi matematici . . . . .	9
1.2	Algebre e $\sigma$ -algebre . . . . .	11
1.3	Misure di probabilità . . . . .	15
1.4	Successioni di eventi . . . . .	17
1.5	Probabilità su spazi campionari discreti . . . . .	20
<b>2</b>	<b>Probabilità condizionata e indipendenza</b>	<b>23</b>
2.1	Probabilità condizionata . . . . .	23
2.2	Indipendenza di eventi . . . . .	24
2.3	Proprietà della probabilità condizionata . . . . .	27
<b>3</b>	<b>Costruzione di una probabilità</b>	<b>31</b>
3.1	Estensione di una probabilità . . . . .	31
3.2	Probabilità su spazi finiti e discreti . . . . .	36
<b>4</b>	<b>Probabilità su <math>\mathbb{R}</math></b>	<b>37</b>
4.1	Funzione di ripartizione . . . . .	37
4.2	Densità continua di probabilità . . . . .	44
<b>5</b>	<b>Variabili aleatorie</b>	<b>49</b>
5.1	Definizione . . . . .	49
5.2	Criteri di misurabilità . . . . .	52
5.3	Proprietà delle funzioni misurabili . . . . .	54
5.4	Legge di una variabile aleatoria . . . . .	58
5.5	Uguaglianza quasi certa . . . . .	60
<b>6</b>	<b>Integrazione rispetto alla probabilità</b>	<b>63</b>
6.1	Variabili aleatorie semplici . . . . .	63
6.2	Integrazione di variabili aleatorie reali . . . . .	65
6.3	Proprietà del valore atteso e dello spazio $\mathcal{L}^1$ . . . . .	72
6.4	Spazi $L^p$ . . . . .	81
6.5	Varianza . . . . .	83
6.6	Valore atteso di serie di variabili aleatorie . . . . .	86
6.7	Integrazione di funzioni misurabili . . . . .	88
<b>7</b>	<b>Variabili aleatorie reali discrete</b>	<b>93</b>
<b>8</b>	<b>Variabili aleatorie reali continue</b>	<b>95</b>
8.1	Misura di Lebesgue . . . . .	95
8.2	Densità continua di probabilità . . . . .	98
8.3	Considerazioni pratiche . . . . .	102

<b>9</b>	<b>Variabili aleatorie indipendenti</b>	<b>103</b>
9.1	Indipendenza di variabili aleatorie . . . . .	103
9.2	Condizioni equivalenti . . . . .	105
9.3	Prodotto cartesiano . . . . .	109
9.4	Il teorema di Fubini-Tonelli . . . . .	112
<b>10</b>	<b>Vettori aleatori</b>	<b>121</b>
10.1	Vettori aleatori discreti . . . . .	122
10.2	Vettori aleatori continui . . . . .	127
10.3	Considerazioni pratiche . . . . .	133
10.4	Covarianza . . . . .	135
10.5	Vettore media e matrice varianza . . . . .	139
<b>11</b>	<b>Funzione caratteristica</b>	<b>143</b>
11.1	Definizioni . . . . .	143
11.2	Proprietà . . . . .	144
11.3	Risultati notevoli . . . . .	149
11.4	Continuità e indipendenza di vettori aleatori . . . . .	150
<b>12</b>	<b>Variabili e vettori aleatori gaussiani</b>	<b>153</b>
12.1	Definizioni . . . . .	153
12.2	Media e varianza . . . . .	155
12.3	Vettori gaussiani e componenti . . . . .	156
12.4	Normalizzazione . . . . .	159
12.5	Leggi correlate alla gaussiana . . . . .	161
12.6	Media e varianza campionaria . . . . .	162
12.7	Considerazioni pratiche . . . . .	165
<b>13</b>	<b>Leggi condizionate</b>	<b>167</b>
13.1	Valori attesi rispetto a una legge condizionata . . . . .	167
13.2	Casi discreti . . . . .	169
<b>14</b>	<b>Caso congiuntamente continuo</b>	<b>172</b>
14.1	Caso generico . . . . .	172
14.2	Caso misto . . . . .	174
<b>15</b>	<b>Attesa condizionata</b>	<b>177</b>
15.1	Condizionamento data una variabile aleatoria . . . . .	177
15.2	Condizionamento data una $\sigma$ -algebra . . . . .	181
15.3	Approssimazione di variabili aleatorie . . . . .	184
15.4	Varianza condizionata . . . . .	185
<b>16</b>	<b>Convergenza di variabili aleatorie</b>	<b>187</b>
16.1	Convergenza certa e quasi certa . . . . .	187
16.2	Convergenza in $L^p$ . . . . .	189

16.3	Convergenza in probabilità . . . . .	190
16.4	Relazioni tra tipi di convergenza (1) . . . . .	192
16.5	Convergenza in legge . . . . .	194
16.6	Criteri di convergenza in legge . . . . .	195
16.7	Relazioni tra tipi di convergenza (2) . . . . .	201
<b>17</b>	<b>Teoremi limite</b>	<b>203</b>
17.1	Legge dei grandi numeri . . . . .	203
17.2	Teorema centrale del limite . . . . .	207
17.3	Comportamento asintotico di alcune distribuzioni . . . . .	208
17.4	Velocità di convergenza . . . . .	210
17.5	Stima dei parametri di una distribuzione . . . . .	216
17.6	Un controesempio: la legge di Cauchy . . . . .	220
<b>18</b>	<b>Catene di Markov</b>	<b>221</b>
18.1	Definizioni base . . . . .	221
18.2	Esempi notevoli . . . . .	226
18.3	Leggi congiunte . . . . .	227
18.4	Classificazione degli stati . . . . .	232
18.5	Problemi di assorbimento . . . . .	240
18.6	Comportamento asintotico . . . . .	243
	<b>Appendice A. Dispense integrative</b>	<b>251</b>
A.1	Richiami di analisi e insiemistica . . . . .	251
A.2	Calcolo combinatorio . . . . .	255
A.3	Integrazione complessa . . . . .	257
A.4	Logaritmo complesso . . . . .	257
A.5	Radice quadrata di una matrice . . . . .	258
	<b>Appendice B. Distribuzioni discrete</b>	<b>260</b>
B.1	Uniforme discreta . . . . .	260
B.2	Bernoulli . . . . .	261
B.3	Binomiale . . . . .	262
B.4	Poisson . . . . .	263
B.5	Geometrica . . . . .	264
B.6	Ipergeometrica . . . . .	265
B.7	Pascal . . . . .	266
B.8	Delta di Dirac . . . . .	267
	<b>Appendice C. Distribuzioni continue</b>	<b>268</b>
C.1	Uniforme continua . . . . .	268
C.2	Esponenziale . . . . .	269
C.3	Normale . . . . .	270
C.4	Gamma . . . . .	271

C.5 Chi-quadro . . . . .	273
C.6 <i>t</i> di Student . . . . .	274
C.7 Weibull . . . . .	275
C.8 Cauchy . . . . .	277
C.9 Beta . . . . .	278
C.10 Rayleigh . . . . .	279

<b>Indice analitico</b>	<b>281</b>
-------------------------	------------



---

# 1 Definizione assiomatica di probabilità

La probabilità, come ogni altra branca della matematica, si fonda su assiomi e definizioni. In questo capitolo saranno introdotti gli strumenti e i concetti alla base della costruzione e dello studio della teoria della probabilità, indispensabili per tradurre i fenomeni aleatori del mondo reale: gli *eventi* rappresentati come insiemi di oggetti, a cui il risultato di un esperimento può appartenere o non appartenere; le  $\sigma$ -algebre, collezioni (insiemi di insiemi) di eventi; la *probabilità*, funzione che associa ad ogni evento un grado di fiducia da 0 a 1.

## 1.1 Eventi matematici

### Definizione 1.1

Un **esperimento aleatorio** è un'osservazione nel mondo reale il cui risultato non è noto a priori e non è deterministico, ma influenzato dal caso (da cui l'aggettivo *aleatorio*).

### Definizione 1.2

Lo **spazio campionario**  $\Omega$  è l'insieme degli esiti elementari possibili ( $\omega$ ) di un esperimento aleatorio.

Esempi di esperimenti ripetibili:

- estrazione del Lotto:  $\Omega = \{1, \dots, 90\}$ ;
- estrazione di un numero reale nell'intervallo  $[0, 1]$ :  $\Omega = [0, 1]$ ;
- lancio di un dado:  $\Omega = \{1, \dots, 6\}$ ;
- lancio di due dadi:  $\Omega = \{1, \dots, 6\}^2$ .<sup>1</sup>

Esempi di esperimenti irripetibili:

- valore di un'azione di Autostrade per l'Italia S.p.A. in un dato tempo:  $\Omega = (0, +\infty)$ ;
- partito di maggioranza alle elezioni di domani:  $\Omega = \{\text{tutti i partiti}\}$ .

Lo spazio campionario  $\Omega$  non è definito univocamente dall'esperimento: per esempio, posso tirare due dadi ma tenerne in conto uno solo. In generale si può trovare l' $\Omega$  più semplice possibile, ma non è detto sia quello giusto (e in generale non lo sarà).

$\Omega$  può essere continuo o discreto, ma nel caso continuo non possiamo attribuire una probabilità ad ogni singolo esito. Si introduce quindi il concetto di *evento* a

---

<sup>1</sup>Si ricorda che  $\Omega = \{1, \dots, 6\}^2$  significa  $\{1, \dots, 6\} \times \{1, \dots, 6\}$ . Dati due insiemi  $A$  e  $B$ , con  $A \times B$  si intende il prodotto cartesiano tra di essi, ovvero l'insieme che ha per elementi tutte le coppie ordinate formate da un elemento di  $A$  e uno di  $B$ .

cui attribuiremo una probabilità.

### Definizione 1.3

È detto **evento** un fatto di cui, al termine di un esperimento aleatorio, si può stabilire il grado di verità (vero o falso).

L'**evento matematico**  $A$  rispetto a un evento dato è l'insieme degli esiti  $\omega$  con cui si afferma che tale evento si è verificato.

Vi è dunque corrispondenza tra l'evento *reale* e l'evento *matematico*  $A$ .

Esempi:

- Pesco un numero primo a tombola  $\longleftrightarrow A = \{2, 3, 5, \dots, 89\} \subseteq \Omega = \{1, \dots, 90\}$ ;
- Tiro due dadi, escono numeri uguali  $\longleftrightarrow A = \{(1, 1), (2, 2), \dots\} \subseteq \Omega = \{1, \dots, 6\}^2$ ;
- Pesco un razionale in  $[0, 1]$   $\longleftrightarrow A = \mathbb{Q} \cap [0, 1] \subseteq \Omega = [0, 1]$ .

Esiste una correlazione tra le operazioni logiche tra eventi reali e quelle insiemistiche:

- $\emptyset$ : evento *impossibile*;
- $\Omega$ : evento *certo*;
- $A^C$ : evento *contrario* di  $A$ ;
- $A \cup B$ : evento  $A$  oppure evento  $B$ ;
- $A \cap B$ : evento  $A$  ed evento  $B$ ;
- $A \cap B = \emptyset$ : eventi *incompatibili*;
- $A \subseteq B$ :  $A$  *implica*  $B$ .

Per un ripasso di alcune importanti proprietà degli insiemi che verranno largamente utilizzate nel corso delle prossime pagine, si veda la sezione dell'appendice A dedicata all'argomento, a pagina [251](#).

### Definizione 1.4

Si dice **insieme delle parti** di  $\Omega$  e si scrive  $2^\Omega$  la collezione di tutti i sottoinsiemi di  $\Omega$ , inclusi  $\Omega$  stesso e l'insieme vuoto  $\emptyset$ .

**NB:** Si noti che  $\#(2^\Omega) = 2^{\#\Omega}$ : questo spiega la notazione utilizzata.

**NB:** Non è possibile definire una probabilità su tutti i sottoinsiemi di  $[0, 1]$  tale che essa sia uguale alla lunghezza del sottoinsieme.

Un sottoinsieme dell'insieme delle parti, ovvero un insieme di eventi (che, peraltro, sono a loro volta insiemi di esiti), è detto **collezione di eventi** o **classe**.

## 1.2 Algebre e $\sigma$ -algebre

### Definizione 1.5

$\mathcal{A} \subseteq 2^\Omega$  si dice **algebra** se:

1.  $\emptyset \in \mathcal{A}, \Omega \in \mathcal{A}$
2.  $A \in \mathcal{A} \implies A^C \in \mathcal{A}$  ( $\mathcal{A}$  è **chiusa per complementazione**)
3.  $A_1, A_2, \dots, A_n \in \mathcal{A} \implies \bigcup_{i=1}^n A_i \in \mathcal{A}, \bigcap_{i=1}^n A_i \in \mathcal{A}$  ( $\mathcal{A}$  è **chiusa per unioni e intersezioni finite**)

### Definizione 1.6

$\mathcal{A} \subseteq 2^\Omega$  si dice  **$\sigma$ -algebra** se:

1.  $\emptyset \in \mathcal{A}, \Omega \in \mathcal{A}$
2.  $A \in \mathcal{A} \implies A^C \in \mathcal{A}$  ( $\mathcal{A}$  è chiusa per complementazione)
3.  $A_1, A_2, \dots, A_n \in \mathcal{A} \implies \bigcup_{i=1}^{+\infty} A_i \in \mathcal{A}, \bigcap_{i=1}^{+\infty} A_i \in \mathcal{A}$  ( $\mathcal{A}$  è chiusa per unioni e intersezioni **numerabili**<sup>2</sup>)

**NB:** Ponendo  $A = \emptyset \in \mathcal{A}$ , per la condizione (2) anche  $A^C = \Omega \in \mathcal{A}$ , quindi la prima condizione è parzialmente ridondante. Ovviamente vale anche il viceversa.

Introduciamo alcune  $\sigma$ -algebre notevoli su un generico  $\Omega$ :

- $\mathcal{A} = 2^\Omega$ :  $\mathcal{A}$  è la  $\sigma$ -algebra più grande possibile;
- $\mathcal{A} = \{\emptyset, \Omega\}$ :  $\mathcal{A}$  è la  $\sigma$ -algebra più piccola possibile;
- $\mathcal{A} = \{\emptyset, \Omega, A, A^C\}$ :  $\mathcal{A}$  è una generica  $\sigma$ -algebra.

*Esempio.* Si lanci di un dado e si consideri l'evento "esce 3". Possiamo considerare due spazi campionari:

- $\Omega = \{1, \dots, 6\}$ . Gli eventi a cui siamo interessati sono  $A = \{3\} \in \Omega$  e il suo complementare  $A^C = \{1, 2, 4, 5, 6\}$ .

---

<sup>2</sup>Si ricorda che un insieme *numerabile* è definito tale se la sua cardinalità coincide con quella di  $\mathbb{N}$ . Più formalmente si dice che l'insieme in questione ha *potenza del numerabile*, ovvero che è possibile creare una biiezione tra esso e l'insieme dei naturali. Anche i razionali  $\mathbb{Q}$  hanno la potenza del numerabile, ma non i reali (che hanno invece *potenza del continuo*).

La  $\sigma$ -algebra generata da essi è  $\mathcal{A} = \{\emptyset, \Omega, \{3\}, \{1, 2, 4, 5, 6\}\}$ , che è una collezione di 4 insiemi.

- $\Omega = \{3, -3\}$ : lo spazio campionario è composto dall'evento  $A$  e dal suo complementare.

Allora la  $\sigma$ -algebra generata da essi è  $\mathcal{A} = 2^\Omega = \{\emptyset, \Omega, \{3\}, \{-3\}\}$ .

Risulta quindi conveniente scegliere prima gli eventi e poi allargare  $\mathcal{A}$  di volta in volta, senza specificare  $\Omega$ ; infatti, si possono avere diversi  $\Omega$  a parità di eventi.

### Definizione 1.7

Siano  $\Omega$  spazio campionario e  $\mathcal{A}$   $\sigma$ -algebra su  $\Omega$  (con  $\mathcal{A} \subseteq 2^\Omega$ ). La coppia  $(\Omega, \mathcal{A})$  è detta **spazio misurabile**.

**NB:** Considereremo *eventi* solo gli insiemi che appartengono a una  $\sigma$ -algebra:  
 $A \subseteq \Omega : A \in \mathcal{A}$ .

### Definizione 1.8

Data la classe  $\mathcal{C} \subseteq 2^\Omega$ , si dice  **$\sigma$ -algebra generata** da  $\mathcal{C}$  e si scrive  $\sigma(\mathcal{C})$  la più piccola  $\sigma$ -algebra che contiene  $\mathcal{C}$ .

### Proposizione 1.9

$\sigma(\mathcal{C})$  è ben definita  $\forall \mathcal{C}$ .

### Dimostrazione

1. Considerando la  $\sigma$ -algebra delle parti  $2^\Omega$ , per ipotesi  $2^\Omega \supseteq \mathcal{C}$ .  
 Dunque esiste almeno una  $\sigma$ -algebra  $\mathcal{A} \supseteq \mathcal{C}$ .
2. Ne esiste una più piccola di tutte?  
 Sia  $\mathcal{A}_\alpha$  una famiglia di  $\sigma$ -algre e  $\mathcal{A} = \bigcap_\alpha \mathcal{A}_\alpha$  (con  $\alpha$  finito o infinito).  
 $\mathcal{A}$  è una  $\sigma$ -algebra. Infatti:
  - (a)  $\bigcap_\alpha \mathcal{A}_\alpha$  contiene  $\Omega$  e  $\emptyset$ :  $\emptyset \in \mathcal{A}_\alpha \forall \alpha$  e dunque  $\emptyset \in \bigcap_\alpha \mathcal{A}_\alpha$ .  
 Si procede analogamente per dimostrare l'appartenenza di  $\Omega$ .
  - (b)  $\bigcap_\alpha \mathcal{A}_\alpha$  è chiusa per complementazione: sia  $A \in \bigcap_\alpha \mathcal{A}_\alpha$ . Dunque  $A \in \mathcal{A}_\alpha \forall \alpha$ .  
 Ciò avviene se e solo se  $A^C \in \mathcal{A}_\alpha \forall \alpha$ , essendo  $\mathcal{A}_\alpha$  una  $\sigma$ -algebra.  
 Ma allora,  $A^C \in \bigcap_\alpha \mathcal{A}_\alpha$ .

(c)  $\bigcap_{\alpha} \mathcal{A}_{\alpha}$  è chiusa per unioni e intersezioni numerabili:

$$\begin{aligned} A_k \in \bigcap_{\alpha} \mathcal{A}_{\alpha} \quad \forall k \in \mathbb{N} &\iff A_k \in \mathcal{A}_{\alpha} \quad \forall k \in \mathbb{N}, \forall \alpha \\ &\implies \bigcup_k A_k \in \mathcal{A}_{\alpha} \quad \forall \alpha \\ &\implies \left( \bigcup_k A_k \right) \in \bigcap_{\alpha} \mathcal{A}_{\alpha} \end{aligned}$$

Allora  $\sigma(\mathcal{C}) = \bigcap_{\mathcal{A} \supseteq \mathcal{C}} \mathcal{A}$ . □

### Teorema 1.10

Se  $A_n \in \mathcal{A} \quad \forall n \in \mathbb{N}$  allora:

$$\limsup A_n \in \mathcal{A} \quad \text{e} \quad \liminf A_n \in \mathcal{A}$$

### Dimostrazione

Per definizione di limite superiore<sup>3</sup>:

$$\limsup A_n = \bigcap_{k=1}^{+\infty} \bigcup_{k \geq n} A_k$$

Per le proprietà di  $\mathcal{A}$  si ha che  $\bigcup_{k \geq n} A_k \in \mathcal{A}$ , e di conseguenza anche la loro intersezione numerabile su  $k$  sarà in  $\mathcal{A}$ .

La dimostrazione per  $\liminf A_n$  è del tutto identica. □

### Proposizione 1.11

$\mathcal{A}$   $\sigma$ -algebra su  $\Omega \implies \mathcal{A}$  algebra su  $\Omega$ .

### Dimostrazione

Siano  $A_1, \dots, A_n$  eventi appartenenti alla  $\sigma$ -algebra  $\mathcal{A}$ .

Definito  $A_{n+1} = A_{n+2} = \dots = A_n$ , si ha che:

$$\bigcap_{k=1}^n A_k = \bigcap_{k=1}^{+\infty} A_k \in \mathcal{A} \quad \text{e} \quad \bigcup_{k=1}^n A_k = \bigcup_{k=1}^{+\infty} A_k \in \mathcal{A} \quad \square$$

*Esempio.* Sia dato l'evento  $A \subseteq \Omega$ . Esso genera la  $\sigma$ -algebra  $\mathcal{A}$ :

$$\mathcal{A} = \{\emptyset, \Omega, A, A^C\} = \sigma(A, A^C) = \sigma(A) = \sigma(A^C)$$

Si nota dunque che la stessa  $\sigma$ -algebra può essere generata da collezioni di dimensioni diverse, come  $\mathcal{C} = \{A\}$  e  $\mathcal{C} = \{A, A^C\}$ .

<sup>3</sup>Le definizioni e proprietà di limite inferiore e superiore sono riprese nella già citata appendice A.

### 1.2.1 $\sigma$ -algebra di Borel

#### Definizione 1.12

Sia  $\Omega = \mathbb{R}$ . La **topologia euclidea di  $\mathbb{R}$**  è la collezione  $\tau := \{A \subseteq \mathbb{R} : A \text{ aperto}\}$ .

La  **$\sigma$ -algebra di Borel** è la  $\sigma$ -algebra generata dalla topologia euclidea di  $\mathbb{R}$ :  $\mathcal{B} = \sigma(\tau)$ .

Gli elementi della  $\sigma$ -algebra di Borel sono detti **boreliani**: in altre parole  $B \subseteq \mathbb{R}$ ,  $B \in \mathcal{B} \implies B$  boreliano.

**NB:** Intervalli aperti e chiusi sono tutti boreliani; tuttavia, non tutti i sottoinsiemi di  $\mathbb{R}$  sono boreliani, ovvero  $\mathcal{B} \subsetneq 2^{\mathbb{R}}$ .

#### Teorema 1.13 (*J-P Th 2.1*)

Sia  $\mathcal{B}$  la  $\sigma$ -algebra di Borel.

Dati  $\mathcal{C} = \{(a, b) : -\infty \leq a \leq b \leq +\infty\}$  e  $\tilde{\mathcal{C}} = \{(-\infty, q] : q \in \mathbb{Q}\}$ , allora:

$$\mathcal{B} = \sigma(\mathcal{C}) = \sigma(\tilde{\mathcal{C}})$$

L'insieme  $\tilde{\mathcal{C}}$  è la collezione delle semirette chiuse e limitate a un estremo razionale. Questo teorema ci dice che  $\mathcal{B}$ , insieme più che continuo, è generato da semirette, oggetti abbastanza semplici e di cardinalità meno che continua.

#### Dimostrazione

Dati  $n \in \mathbb{N}$  e  $k \in \mathbb{Z}$ , definiamo intervalli del tipo  $(\frac{k}{2^n}, \frac{k+1}{2^n}]$ . La loro unione copre tutto  $\mathbb{R}$ :  $\bigcup_k (\frac{k}{2^n}, \frac{k+1}{2^n}] = \mathbb{R} \quad \forall n$ .

Definiamo quindi:

$$\mathcal{C}_0 := \left\{ \left( \frac{k}{2^n}, \frac{k+1}{2^n} \right] : n \in \mathbb{N}, k \in \mathbb{Z} \right\} \subseteq \mathcal{B}$$

Essa è una classe numerabile, perché è unione di un numero numerabile di insiemi numerabili.

1. Sia  $A \subseteq \mathbb{R}$  con  $A \in \tau$ , ovvero  $A$  aperto. Allora esiste una successione  $(I_j)_{j \in \mathbb{N}} \subseteq \mathcal{C}_0$  tale che  $A = \bigcup_j I_j$ , dove  $I_j$  è un intervallo del tipo  $(\frac{k}{2^n}, \frac{k+1}{2^n}]$ . Infatti, preso  $x \in A$ , esiste un suo intorno  $I_x \in \mathcal{C}_0$  tale che  $x \in I_x$  e  $I_x \subseteq A$ . L'unione degli  $I_x$  copre tutto  $\mathbb{R}$ , quindi:

$$A = \bigcup_{x \in A} \{x\} = \bigcup_{x \in A} I_x = \bigcup_{j=1}^{+\infty} I_j$$

Da ciò segue che  $\sigma(\mathcal{C}_0) \supseteq \tau$ .

2. (a)  $\mathcal{C}_0 \subseteq \mathcal{B} \implies \sigma(\mathcal{C}_0) \subseteq \sigma(\mathcal{B}) = \mathcal{B}$  (una  $\sigma$ -algebra genera ovviamente sé stessa).
- (b) Poiché ora sappiamo che  $\sigma(\mathcal{C}_0) \supseteq \tau$ , si può scrivere  $\sigma(\sigma(\mathcal{C}_0)) \supseteq \sigma(\tau)$ , ovvero  $\sigma(\mathcal{C}_0) \supseteq \mathcal{B}$ .
- (c) Dunque necessariamente  $\sigma(\mathcal{C}_0) = \mathcal{B}$ , visto che si contengono l'un l'altra.
3. (a)  $\tilde{\mathcal{C}} \subseteq \tau \implies \sigma(\tilde{\mathcal{C}}) \subseteq \sigma(\tau) = \mathcal{B}$ .
- (b)  $\sigma(\tilde{\mathcal{C}}) \supseteq \mathcal{C}_0$  perché ogni intervallo di  $\mathcal{C}_0$  può anche essere ottenuto come unione o intersezione di elementi di  $\tilde{\mathcal{C}}$ : per esempio,  $(\frac{k}{2^n}, \frac{k+1}{2^n}] = (-\infty, \frac{k+1}{2^n}] \cap (-\infty, \frac{k}{2^n}]^C$ . Allora  $\sigma(\sigma(\tilde{\mathcal{C}})) \supseteq \sigma(\mathcal{C}_0)$ , ovvero  $\sigma(\tilde{\mathcal{C}}) \supseteq \mathcal{B}$  per quanto detto prima.
- (c) Si ottiene dunque  $\sigma(\tilde{\mathcal{C}}) = \mathcal{B}$ .
4. (a)  $\mathcal{C} \subseteq \sigma(\tau) = \mathcal{B} \implies \sigma(\mathcal{C}) \subseteq \mathcal{B}$ .
- (b)  $\sigma(\mathcal{C}) \supseteq \mathcal{C}_0$  perché ogni intervallo di  $\mathcal{C}_0$  può anche essere ottenuto come unione o intersezione di elementi di  $\mathcal{C}$ : per esempio,  $(\frac{k}{2^n}, \frac{k+1}{2^n}] = (\frac{k}{2^n}, +\infty) \cap (\frac{k+1}{2^n}, +\infty)^C$ . Allora  $\sigma(\sigma(\mathcal{C})) \supseteq \sigma(\mathcal{C}_0) = \mathcal{B} \implies \sigma(\mathcal{C}) \supseteq \mathcal{B}$ .
- (c) Dalle precedenti si ottiene  $\sigma(\mathcal{C}) = \mathcal{B}$ . □

## 1.3 Misure di probabilità

La probabilità  $\mathbb{P}(A)$  rappresenta un grado di fiducia, misurato da 0 a 1, circa il verificarsi dell'evento  $A$ . Si noti che questa è una valutazione *a priori*, cioè prima dell'effettivo svolgimento dell'esperimento aleatorio.

### 1.3.1 Interpretazioni modellistiche

- **Interpretazione soggettivistica:** secondo chi calcola la probabilità, specie in finanza, con eventi non ripetibili.
- **Interpretazione frequentista:** ripetizione degli eventi per calcolarne la frequenza.

$$\mathbb{P}(A) = \lim_{n \rightarrow +\infty} f_n(A)$$

Qui  $f_n(A)$  è la frequenza relativa; questo funziona grazie alla *legge dei grandi numeri*, come vedremo molto più avanti.

- **Estrazione da popolazioni finite non truccate:**

$$\mathbb{P}(A) = \frac{\#\text{Casi favorevoli}}{\#\text{Casi possibili}}$$

*Esempio.* Nel lotto:  $\mathbb{P}(\text{numero primo}) = \frac{\#\text{Numeri primi tra 1 e 90}}{90} = \frac{24}{90}$

### 1.3.2 Definizione di probabilità

#### Definizione 1.14

Dato uno spazio misurabile  $(\Omega, \mathcal{A})$ , una **misura di probabilità**, o, più semplicemente, **probabilità**, è una funzione  $\mathbb{P} : \mathcal{A} \rightarrow [0, 1]$  tale che:

1.  $\mathbb{P}(\Omega) = 1$
2.  $\forall$  successione  $A_n \in \mathcal{A}$  di eventi **disgiunti a coppie**, cioè tali che  $A_k \cap A_l = \emptyset \ \forall k \neq l$ , vale la seguente proprietà di  **$\sigma$ -additività**:

$$\mathbb{P}\left(\bigcup_{n=1}^{+\infty} A_n\right) = \sum_{n=1}^{+\infty} \mathbb{P}(A_n)$$

### 1.3.3 Proprietà

#### Teorema 1.15 (*J-P Th 2.2*)

Sia  $\mathbb{P}$  una (misura di) probabilità su  $\mathcal{A}$ . Allora:

1.  $\mathbb{P}(\emptyset) = 0$
2. Dati  $A_1, \dots, A_n \in \mathcal{A}$  eventi disgiunti a coppie, vale la seguente proprietà di **additività**:

$$\mathbb{P}\left(\bigcup_{k=1}^n A_k\right) = \sum_{k=1}^n \mathbb{P}(A_k)$$

#### Dimostrazione

1. Si crea una serie di  $A_n$  insiemi vuoti e si dimostra che la probabilità della loro unione è uguale a quella dell'insieme vuoto:

$$\begin{aligned} A_n = \emptyset \quad \forall n \in \mathbb{N} &\implies A_n \in \mathcal{A}, A_k \cap A_l = \emptyset \cap \emptyset = \emptyset \\ &\implies \mathbb{P}\left(\bigcup_{n=1}^{+\infty} A_n\right) = \sum_{n=1}^{+\infty} \mathbb{P}(A_n) \\ &\implies \mathbb{P}(\emptyset) = \sum_{n=1}^{+\infty} \mathbb{P}(\emptyset) \\ &\implies \mathbb{P}(\emptyset) = 0 \end{aligned}$$



2. Si può sfruttare quindi la  $\sigma$ -additività estendendo gli  $A_k$  fino all'infinito:

$$\begin{aligned}
 A_1, \dots, A_n \in \mathcal{A} \quad A_k \cap A_l = \emptyset \quad \forall k \neq l, \quad \tilde{A}_k &= \begin{cases} A_k & k = 1, \dots, n \\ \emptyset & k > n \end{cases} \\
 \implies \tilde{A}_k \cap \tilde{A}_l = \emptyset \quad \forall k \neq l \\
 \implies \mathbb{P} \left( \bigcup_{k=1}^{+\infty} \tilde{A}_k \right) &= \sum_{k=1}^{+\infty} \mathbb{P}(\tilde{A}_k) \\
 \implies \mathbb{P} \left( \bigcup_{k=1}^n A_k \right) &= \sum_{k=1}^n \mathbb{P}(\tilde{A}_k) + \sum_{k=n+1}^{+\infty} \mathbb{P}(\tilde{A}_k) = \sum_{k=1}^n \mathbb{P}(\tilde{A}_k) + 0 \\
 \implies \mathbb{P} \left( \bigcup_{k=1}^n A_k \right) &= \sum_{k=1}^n \mathbb{P}(A_k) \quad \square
 \end{aligned}$$

Introduciamo una nuova diffusa notazione di intersezione insiemistica:  $\mathbb{P}(A, B)$  è sinonimo di  $\mathbb{P}(A \cap B)$ .

**Corollario 1.16**

- Dati  $A, A' \in \mathcal{A}$  si ha che  $A \subseteq A' \implies \mathbb{P}(A) \leq \mathbb{P}(A')$ ;
- $A \in \mathcal{A} \implies \mathbb{P}(A^C) = 1 - \mathbb{P}(A)$ , dove 1 è  $\mathbb{P}(\Omega)$ ;
- $A, B \in \mathcal{A} \implies \mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A, B)$ .

**1.4 Successioni di eventi**

**Definizione 1.17**

Dato  $A_n \in \mathcal{A} \quad \forall n$ :

- $A_n \uparrow A$  significa che  $A_n$  è una successione di eventi **crescente** verso  $A$ :

$$A_n \subseteq A_{n+1}, \quad \bigcup_{n=1}^{+\infty} A_n = A$$

- $A_n \downarrow A$  significa che  $A_n$  è una successione di eventi **decrescente** verso  $A$ :

$$A_n \supseteq A_{n+1}, \quad \bigcap_{n=1}^{+\infty} A_n = A$$

In entrambi i casi,  $A$  è detto l'**insieme limite** della successione.

**NB:**  $A_n$  decresce verso  $A \iff A_n^C$  cresce verso  $A^C$  (legge di De Morgan).

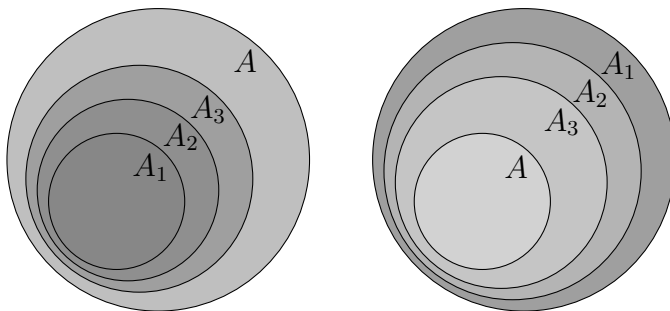


Figura 1.1: eventi  $A_n$  crescenti e decrescenti verso  $A$

**Teorema 1.18** (*J-P Th 2.3*)

Siano  $(\Omega, \mathcal{A})$  spazio misurabile, e  $\mathbb{P} : \mathcal{A} \rightarrow [0, 1]$  funzione additiva e tale che  $\mathbb{P}(\Omega) = 1$ . Allora le seguenti affermazioni sono equivalenti:

1.  $\mathbb{P}$  è  $\sigma$ -additiva (ovvero  $\mathbb{P}$  è una probabilità)
2.  $A_n \in \mathcal{A}, A_n \downarrow \emptyset \implies \mathbb{P}(A_n) \downarrow 0$
3.  $A_n \in \mathcal{A}, A_n \downarrow A \implies \mathbb{P}(A_n) \downarrow \mathbb{P}(A)$
4.  $A_n \in \mathcal{A}, A_n \uparrow \Omega \implies \mathbb{P}(A_n) \uparrow 1$
5.  $A_n \in \mathcal{A}, A_n \uparrow A \implies \mathbb{P}(A_n) \uparrow \mathbb{P}(A)$

Dai punti (3) e (5) si può dedurre lo *scambio di limiti*  $\mathbb{P}(A) = \mathbb{P}(\lim_n A_n) = \lim_n \mathbb{P}(A_n)$ .

**Dimostrazione**

La dimostrazione si svolgerà così:

- (a) casi ovvi;
  - (b) (4)  $\implies$  (5);
  - (c) (5)  $\implies$  (1);
  - (d) (1)  $\implies$  (5);
- 
- (a)
    - (3)  $\iff$  (5) per le leggi di De Morgan (complementarità);
    - (2)  $\iff$  (4) per le leggi di De Morgan (complementarità);
    - (3)  $\implies$  (2) e (5)  $\implies$  (4) perché sono casi particolari.

- (b) (4)  $\implies$  (5). Siano  $A_n \in \mathcal{A}$  tali che  $A_n \uparrow A$ . Definiamo allora  $B_n = A_n \cup A^C$ .  $B_{n+1} \supseteq B_n$ . Allora  $B_n \uparrow \Omega$  e dunque  $\mathbb{P}(B_n) \rightarrow 1$ . Inoltre, vale:

$$\bigcup_n B_n = \bigcup_n (A_n \cup A^C) = \left( \bigcup_n A_n \right) \cup A^C = A \cup A^C = \Omega$$

Passando alle probabilità:

$$\begin{aligned} \mathbb{P}(B_n) &= \mathbb{P}(A_n \cup A^C) \\ &= \mathbb{P}(A_n) + \mathbb{P}(A^C) - \mathbb{P}(A_n \cap A^C) \\ &= \mathbb{P}(A_n) + \mathbb{P}(A^C) \longrightarrow \mathbb{P}(A) + \mathbb{P}(A^C) \end{aligned}$$

Allora, sottraendo da entrambi i membri  $\mathbb{P}(A^C)$ , si ha che  $\mathbb{P}(A_n) \rightarrow \mathbb{P}(A)$ .

- (c) (5)  $\implies$  (1). Siano  $A_n \in \mathcal{A}$  tali che  $A_k \cap A_l = \emptyset \ \forall k \neq l$ .

Definiamo dunque  $B_n = \bigcup_{k=1}^n A_k$ . Questo garantisce che:

$$B_n \subseteq B_{n+1} \quad \text{e} \quad B_n = \bigcup_{k=1}^n B_k = \bigcup_{k=1}^n A_k$$

Possiamo quindi passare alla probabilità:

$$\begin{aligned} \mathbb{P}(B_n) &= \mathbb{P}\left(\bigcup_{k=1}^n A_k\right) \implies \sum_{k=1}^n (\mathbb{P}(A_k)) = \mathbb{P}\left(\bigcup_{k=1}^n A_k\right) \\ &\implies \sum_{k=1}^{+\infty} (\mathbb{P}(A_k)) = \mathbb{P}\left(\bigcup_{k=1}^{+\infty} A_k\right) \end{aligned}$$

- (d) (1)  $\implies$  (5). Siano  $A_n \uparrow A$ ,  $A_n \in \mathcal{A}$ . Si definisca la successione  $B_n$  nel seguente modo:

$$\begin{cases} B_1 = A_1 \\ B_2 = A_2 \setminus A_1 = A_2 \cap A_1^C \\ \dots \\ B_n = A_n \setminus A_{n-1} \end{cases}$$

Visto che gli  $A_n \uparrow A$  allora i  $B_n$  risultano disgiunti e si può sfruttare la  $\sigma$ -additività, ripetendo la procedura del punto (c):

$$\mathbb{P}\left(\bigcup_{n=1}^{+\infty} B_n\right) = \sum_{n=1}^{+\infty} \mathbb{P}(B_n) \quad \square$$

**Proposizione 1.19: proprietà di subadditività**

Data una successione numerabile di eventi  $A_n \in \mathcal{A} \ \forall n$ , vale la seguente disuguaglianza:

$$\mathbb{P}\left(\bigcup_{n=1}^{+\infty} A_n\right) \leq \sum_{n=1}^{+\infty} \mathbb{P}(A_n)$$

Citiamo, inoltre, una disuguaglianza di verso opposto che tratta invece dell'intersezione finita di eventi.

**Proposizione 1.20: disuguaglianza di Bonferroni**

Data una collezione finita di eventi  $A_1, \dots, A_n \in \mathcal{A}$ , si ha che:

$$\mathbb{P}\left(\bigcap_{i=1}^n A_i\right) \geq \sum_{i=1}^n \mathbb{P}(A_i) - (n - 1)$$

**1.5 Probabilità su spazi campionari discreti**

**Teorema 1.21: probabilità su spazi discreti (J-P Th 4.1)**

Sia  $\Omega$  discreto (ovvero con cardinalità al più numerabile).

1.  $\mathbb{P}$  su  $(\Omega, 2^\Omega)$  è caratterizzata dai valori sugli *atomi* (ossia elementi costituiti da un solo esito), cioè:

$$p_\omega = \mathbb{P}(\{\omega\}), \ \omega \in \Omega; \quad \text{infatti} \quad \mathbb{P}(A) = \sum_{\omega \in A} p_\omega \quad \text{con} \quad p : \Omega \rightarrow [0, 1]$$

2. Sia  $p : \Omega \rightarrow [0, 1]$ . Allora:

$$\exists! \mathbb{P} : 2^\Omega \rightarrow [0, 1] : \mathbb{P}(\{\omega\}) = p_\omega \iff \begin{cases} p_\omega \geq 0 \quad \forall \omega \\ \sum_{\omega \in \Omega} p_\omega = 1 \end{cases}$$

**Dimostrazione**

1. Siano  $\mathbb{P} : 2^\Omega \rightarrow [0, 1]$  e  $A \in 2^\Omega, A \subseteq \Omega$ .

Allora  $A = \bigcup_{x \in A} \{x\}$  e essendo gli atomi eventi disgiunti, per  $\sigma$ -additività vale:

$$\mathbb{P}(A) = \mathbb{P}\left(\bigcup_{x \in A} \{x\}\right) = \sum_{x \in A} \mathbb{P}(\{x\}) = \sum_{x \in A} p_x$$

2. **Dimostriamo** ( $\implies$ ):

Per ipotesi  $\exists! \mathbb{P} : 2^\Omega \rightarrow [0, 1]$ , si può allora porre  $p_\omega = \mathbb{P}(\{\omega\}) \in [0, 1]$ . Inoltre:

$$\sum_{\omega \in \Omega} p_\omega = \sum_{\omega \in \Omega} \mathbb{P}(\{\omega\}) = \mathbb{P}\left(\bigcup_{\omega \in \Omega} \{x\}\right) = \mathbb{P}(\Omega) = 1$$

**Dimostriamo** ( $\Leftarrow$ ):

Sia  $\mathbb{P} : \Omega \rightarrow \mathbb{R}$  tale che  $p_\omega \geq 0$  e  $\sum_{\omega \in \Omega} p_\omega = 1$ .

Posto  $\mathbb{P}(A) = \sum_{\omega \in A} p_\omega$ , allora  $p_\omega = \mathbb{P}(\{\omega\})$  ed è possibile mostrare che  $\mathbb{P}$  è una probabilità:

- $\mathbb{P}(\Omega) = \sum_{\omega \in \Omega} p_\omega = 1$ ;
- $\mathbb{P}$  è  $\sigma$ -additiva, infatti dati  $A_k \in 2^\Omega$ ,  $k \in \mathbb{N}$ ,  $A_k \cap A_l = \emptyset \quad \forall k \neq l$ :

$$\mathbb{P}\left(\bigcup_{k=1}^{+\infty} A_k\right) = \sum_{\omega \in \bigcup_{k=1}^{+\infty} A_k} p_\omega = \sum_{k=1}^{+\infty} \left[ \sum_{\omega \in A_k} p_\omega \right] = \sum_{k=1}^{+\infty} \mathbb{P}(A_k)$$

Allora  $\exists! \mathbb{P} : 2^\Omega \rightarrow [0, 1]$  tale che  $\mathbb{P}(\{\omega\}) = p_\omega$ . □

*Esempio.* Dati  $\Omega = \{0, 1, \dots, n\}$  e  $p \in [0, 1]$  fissato, si definisca la *distribuzione binomiale* nel modo che segue:

$$p_k = \binom{n}{k} p^k (1-p)^{n-k} \implies p_k \geq 0 \quad \forall k = 0, \dots, n$$

Per lo sviluppo binomiale, o formula di Newton:

$$\sum_{k=0}^n \binom{n}{k} p^k (1-p)^{n-k} = [p + (1-p)]^n = 1$$

La somma delle probabilità sugli atomi è dunque pari a 1, come richiesto a una probabilità.

*Esempio.* Dati  $\Omega = \mathbb{Z}^+ = \{0, 1, 2, \dots\}$  e  $\lambda > 0$ , si definisca la *distribuzione di Poisson* di parametro  $\lambda$ :

$$p_k = (e^{-\lambda}) \frac{\lambda^k}{k!} \implies \sum_{k=0}^{+\infty} p_k = \sum_{k=0}^{+\infty} (e^{-\lambda}) \frac{\lambda^k}{k!} = (e^{-\lambda}) \sum_{k=0}^{+\infty} \frac{\lambda^k}{k!} = e^{-\lambda} e^\lambda = 1$$

Anche qui la probabilità totale è pari a 1.

### Definizione 1.22

Siano  $(\Omega, \mathcal{A})$  uno spazio misurabile e  $\mathbb{P}$  una probabilità su  $\mathcal{A}$ .

La tripletta  $(\Omega, \mathcal{A}, \mathbb{P})$  è detta **spazio di probabilità**.

**Definizione 1.23**

Dato uno spazio di probabilità  $(\Omega, \mathcal{A}, \mathbb{P})$ , sia l'evento  $A \in \mathcal{A}$ .

Se  $\mathbb{P}(A) = 0$ ,  $A$  è detto **improbabile** (o trascurabile). Se invece  $\mathbb{P}(A) = 1$ ,  $A$  è detto **quasi certo**.

---

## 2 Probabilità condizionata e indipendenza

La probabilità, che è per definizione il grado di fiducia del verificarsi di un evento, deve essere flessibile a informazioni aggiuntive accumulate nel corso dell'esperimento aleatorio. Per questo motivo introduciamo la *probabilità condizionata*, che assimila l'informazione data dal verificarsi di un certo evento. Questo consentirà anche di parlare di *eventi indipendenti*, ovvero che non si influenzano tra di loro. Infine studieremo le proprietà di questi due nuovi concetti, che permetteranno di enunciare alcuni teoremi e formule molto utili nello studio e nell'applicazione pratica della probabilità.

### 2.1 Probabilità condizionata

#### Definizione 2.1

Sia  $(\Omega, \mathcal{A}, \mathbb{P})$  uno spazio di probabilità,  $A, B \in \mathcal{A}$  con  $\mathbb{P}(B) > 0$ .

Definiamo la **probabilità di  $A$  condizionata da  $B$**  come:

$$\mathbb{P}(A|B) := \frac{\mathbb{P}(A, B)}{\mathbb{P}(B)}$$

Siano due eventi  $A$  e  $B$  di cui si scoprono gli esiti in tempi separati, prima  $B$  poi  $A$ . Come si può ricalcolare la probabilità di  $A$  una volta venuto a conoscenza del risultato di  $B$ ?

<i>A priori</i>	<i>In medias res</i>	<i>A posteriori</i>
$\mathbb{P}(B)$	$B$	$B$
$\mathbb{P}(A)$	$\mathbb{P}(A B)$	$A$ oppure $A^C$
All'inizio si può solo predire l'esito	Se si verifica $B$ si aggiornano le probabilità	Alla fine è noto l'esito

#### 2.1.1 Interpretazioni modellistiche

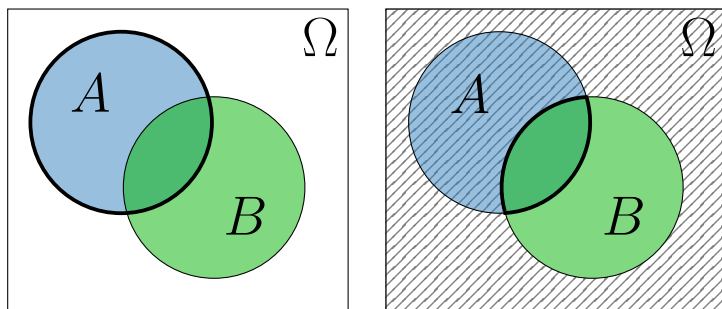
- **Interpretazione soggettivistica:**

*A priori* si considera tutto  $\Omega$ . Considerato un suo punto questo può essere all'interno di  $A$ ,  $B$ , in  $A \cap B$ , oppure in  $(A \cup B)^C$ .

$$\mathbb{P}(A) = f_r(A) = \frac{\#A}{\#\Omega} \quad \text{e} \quad \mathbb{P}(B) = f_r(B) = \frac{\#B}{\#\Omega}$$

Se *in medias res* si viene a conoscenza che si verifica  $B$ , allora l'insieme  $B$  diventa il “nuovo”  $\Omega$ , quindi bisogna *aggiornare* il valore della probabilità di  $A$ :

$$\mathbb{P}(A|B) = \frac{\#(A \cap B)}{\#B} = \frac{\#(A \cap B)}{\#\Omega} \cdot \frac{\#\Omega}{\#B} = \frac{\mathbb{P}(A, B)}{\mathbb{P}(B)}$$



Scoprendo che è avvenuto  $B$  si escludono automaticamente tutti gli eventi che non sono in  $A \cap B$ . La probabilità condizionata cambia il dominio da  $\Omega$  a  $B$  e considera possibili solo gli eventi nell'intersezione.

- **Interpretazione frequentista:**

Ripetendo  $n$  volte (con  $n \rightarrow +\infty$ ) l'esperimento:

$$\mathbb{P}(A|B) \approx \frac{f_r(A, B)}{f_r(B)} = \frac{\mathbb{P}(A, B)}{\mathbb{P}(B)}$$

Infatti vale la seguente relazione con le frequenze<sup>4</sup>:

$$\mathbb{P}(A) \approx f_r(A) = \frac{F(A)}{n} \quad \text{e} \quad \mathbb{P}(A, B) = \frac{F(A, B)}{n}$$

Analizzando i casi degeneri, quando  $A$  e  $B$  sono disgiunti o  $B$  è un sottoinsieme di  $A$ :

$$A \cap B = \emptyset \implies \mathbb{P}(A|B) = 0 \quad \text{e} \quad B \subseteq A \implies \mathbb{P}(A|B) = 1$$

## 2.2 Indipendenza di eventi

### Definizione 2.2

Dati  $A, B \in \mathcal{A}$ , essi si dicono **eventi indipendenti** (e si scrive  $A \perp\!\!\!\perp B$ ) se  $\mathbb{P}(A, B) = \mathbb{P}(A) \cdot \mathbb{P}(B)$ .

---

<sup>4</sup>Si ricorda che con  $F$  si indica la *frequenza assoluta*, mentre con  $f_r$  la *frequenza relativa*.



La definizione è intuitivamente giustificata: dati  $A, B$  tali che  $\mathbb{P}(B) > 0$ ,  $\mathbb{P}(A) > 0$ , per la definizione di proprietà condizionata:

$$\mathbb{P}(A, B) = \mathbb{P}(A) \mathbb{P}(B) \iff \begin{cases} \mathbb{P}(A|B) = \mathbb{P}(A) \\ \mathbb{P}(B|A) = \mathbb{P}(B) \end{cases}$$

In altre parole, i due insiemi non influenzano le rispettive probabilità condizionate, come ci si aspetta da due eventi indipendenti. Si noti inoltre che indipendenza e incompatibilità (intersezione vuota) *non* sono sinonimi.

**Casi degeneri** ( $\mathbb{P}(A) = 0$  oppure  $\mathbb{P}(B) = 0$ ):

- $A$  improbabile:  $A \perp\!\!\!\perp B \quad \forall B \in \mathcal{A} \quad (\mathbb{P}(A) = 0)$
- $A$  quasi certo:  $A \perp\!\!\!\perp B \quad \forall B \in \mathcal{A}$ . Infatti  $\mathbb{P}(B) = \mathbb{P}(B, A) + \mathbb{P}(B, A^C)$ , ma  $\mathbb{P}(B, A^C) = \mathbb{P}(A^C) = 0$ .

*Esempio.* Si pesca una carta da un mazzo di salentine<sup>5</sup>. Si definiscono quindi gli eventi:

$$A = \{\text{esce zappe}\} \quad \text{e} \quad B = \{\text{esce Santo}\}$$

In termini probabilistici questo significa che:  $\Omega = \{\text{Mazzo di salentine}\}$ ,  $\#\Omega = 40$ ,  $\mathcal{A} = 2^\Omega$ . Se  $\mathbb{P}$  è uniforme su  $\mathcal{A}$ , allora  $p_\omega = \frac{1}{40} \quad \forall \omega \in \Omega$ .

Quindi, calcolando le probabilità dei due eventi e facendone il prodotto possiamo mostrarne l'indipendenza:

$$\begin{aligned} \mathbb{P}(A) &= \frac{\#A}{\#\Omega} = \frac{10}{40}, & \mathbb{P}(B) &= \frac{\#B}{\#\Omega} = \frac{4}{40}, & \mathbb{P}(A, B) &= \frac{\#(A \cap B)}{\#\Omega} = \frac{1}{40} \\ \implies \mathbb{P}(A, B) &= \mathbb{P}(A)\mathbb{P}(B) \implies A \perp\!\!\!\perp B \end{aligned}$$

*Esempio:* lanciati due dadi, calcolare la probabilità che la somma sia 3.

Si definisce lo spazio campionario  $\Omega = \{1, \dots, 6\}^2$  e gli eventi  $A = \{\text{esce } x \text{ sul primo dado}\}$  e  $B = \{\text{esce } y \text{ sul secondo dado}\}$ .

Gli eventi  $A$  e  $B$  sono indipendenti e con probabilità  $\frac{1}{6}$ , quindi la probabilità dell'atomo è  $\frac{1}{36}$ . Gli atomi il cui risultato è 3 sono rispettivamente  $(1, 2)$  e  $(2, 1)$ ; possiamo quindi calcolare la probabilità della somma:

$$\mathbb{P}(\text{somma} = 3) = \frac{\#(\text{somma} = 3)}{\#\Omega} = \frac{2}{36} = \frac{1}{18}$$

---

<sup>5</sup>Mazzo tradizionale composto da 40 carte. I semi sono tarante, zappe, lecci e brocche; le figure sono fante, asino e santo

**Teorema 2.3: indipendenza e complementarità (J-P Th 3.1)**

$$A \perp\!\!\!\perp B \iff A \perp\!\!\!\perp B^C \iff A^C \perp\!\!\!\perp B \iff A^C \perp\!\!\!\perp B^C$$

**Dimostrazione**

Si dimostra che  $A \perp\!\!\!\perp B \implies A \perp\!\!\!\perp B^C$ , poi il resto è una banale applicazione della complementarità.

$$\begin{aligned} \mathbb{P}(A, B^C) &= \mathbb{P}(A) - \mathbb{P}(A, B) = \mathbb{P}(A) - \mathbb{P}(A)\mathbb{P}(B) \\ &= \mathbb{P}(A)(1 - \mathbb{P}(B)) = \mathbb{P}(A)\mathbb{P}(B^C) \end{aligned} \quad \square$$

*Esempio:* lancio di due monete.

Definito  $(\Omega, \mathcal{A}, \mathbb{P})$  spazio di probabilità,  $\Omega = \text{“Non ve lo dico”}$ <sup>6</sup>,  $\mathcal{A} \subseteq 2^\Omega$ ,  $\mathcal{A} = \sigma(T_1, T_2)$  Dove  $T_k$  è testa al lancio  $k$ . Qual è la probabilità su  $\mathcal{A}$ ?

$$\begin{cases} \mathbb{P}(T_1) = \mathbb{P}(T_2) = \frac{1}{2} \\ T_1 \perp\!\!\!\perp T_2 \end{cases}$$

**2.2.1 Famiglie di eventi indipendenti**

**Definizione 2.4**

Una famiglia di eventi  $\{A_i\}_{i \in I}$  è detta **mutuamente indipendente** se,  $\forall J \subseteq I$  (con  $J$  insieme finito):

$$\mathbb{P}\left(\bigcap_{i \in J} A_i\right) = \prod_{i \in J} \mathbb{P}(A_i)$$

*Esempio.* Sia  $I = \{1, 2, 3\}$ . Gli insiemi  $\{A_1, A_2, A_3\}$  sono mutuamente indipendenti se e solo se:

$$\begin{cases} \mathbb{P}(A_1, A_2) = \mathbb{P}(A_1) \cdot \mathbb{P}(A_2) \\ \mathbb{P}(A_2, A_3) = \mathbb{P}(A_2) \cdot \mathbb{P}(A_3) \\ \mathbb{P}(A_1, A_3) = \mathbb{P}(A_1) \cdot \mathbb{P}(A_3) \\ \mathbb{P}(A_1, A_2, A_3) = \mathbb{P}(A_1) \cdot \mathbb{P}(A_2) \cdot \mathbb{P}(A_3) \end{cases}$$

La quarta proprietà non è ricavabile dalle altre e comporta, ad esempio,  $A_1 \cap A_2 \perp\!\!\!\perp A_3$ . Non basta dunque l'indipendenza tra i singoli insiemi per garantire la mutua indipendenza della famiglia.

Similmente, l'ultima proprietà non implica necessariamente le prime tre.

---

<sup>6</sup>Questa brillante provocazione dell'esimio prof. Gregoratti, citata testualmente, ci ricorda che la probabilità non dipende direttamente dall'insieme degli atomi  $\Omega$ , bensì dalla probabilità sugli eventi della  $\sigma$ -algebra  $\mathcal{A}$ . Non è dunque necessario avere informazioni sul dominio. Così tanto significato in così poche parole.

## 2.3 Proprietà della probabilità condizionata

### Teorema 2.5 (*J-P Th 3.2*)

Siano  $B \in \mathcal{A}$  e  $\mathbb{P}$  probabilità su  $\mathcal{A}$ , con  $\mathbb{P}(B) > 0$ .

Allora la *funzione*  $\mathbb{P}(\cdot|B)$  è una misura di probabilità su  $\mathcal{A}$ :

$$\mathbb{P}(\cdot|B) : \mathcal{A} \rightarrow [0, 1], \quad A \mapsto \mathbb{P}(A|B) = \frac{\mathbb{P}(A, B)}{\mathbb{P}(B)}$$

### Dimostrazione

La probabilità dell'intero dominio è:

$$\mathbb{P}(\Omega|B) = \frac{\mathbb{P}(\Omega, B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B)}{\mathbb{P}(B)} = 1$$

Dimostriamo ora che  $\mathbb{P}(\cdot|B)$  è  $\sigma$ -additiva. Siano  $A_n \in \mathcal{A}$ ,  $n \in \mathbb{N}$ ,  $A_k \cap A_l = \emptyset \forall k \neq l$ . Allora:

$$\begin{aligned} \mathbb{P}\left(\bigcup_{n=1}^{+\infty} A_n \mid B\right) &= \frac{\mathbb{P}(\bigcup A_n, B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(\bigcup (A_n, B))}{\mathbb{P}(B)} \\ &= \frac{\sum \mathbb{P}(A_n, B)}{\mathbb{P}(B)} = \sum_{n=1}^{+\infty} \mathbb{P}(A_n|B) \end{aligned}$$

$\mathbb{P}$  rispetta dunque le proprietà di una probabilità, come da tesi.  $\square$

**NB:** Essendo  $\mathbb{P}(\cdot|B)$  una probabilità, vale la proprietà  $\mathbb{P}(A^C|B) = 1 - \mathbb{P}(A|B)$ .

Tuttavia, in generale,  $\mathbb{P}(A|B^C) \neq 1 - \mathbb{P}(A|B)$ , perché  $\mathbb{P}(\cdot|B)$  è definito con un  $B$  fissato.

### Corollario 2.6

Se  $\mathbb{P}(A), \mathbb{P}(B) > 0$ , allora si ha:

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A, B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(A, B)}{\mathbb{P}(B)} \cdot \frac{\mathbb{P}(A)}{\mathbb{P}(A)} = \mathbb{P}(B|A) \cdot \frac{\mathbb{P}(A)}{\mathbb{P}(B)}$$

In questo modo è possibile scambiare l'ordine degli eventi, trasformando  $A$  da evento condizionato a evento condizionante.

*Esempio.* Consideriamo un test clinico somministrato a una popolazione in cui la malattia ha una probabilità dell'1% di comparire. Il test non è affidabile al 100%: si ha una probabilità di falsi positivi del 5% e del 2% di falsi negativi. Qual è la probabilità che un individuo sia malato se il test ha dato esito positivo?

Siano  $S$  l'evento "individuo sano",  $M$  l'evento "individuo malato",  $N$  l'evento "esito del test negativo",  $P$  l'evento "esito del test positivo". Considerata la  $\sigma$ -algebra  $\mathcal{A} = \sigma(M, P)$ , i dati allora si scrivono come:  $\mathbb{P}(M) = 0.01$ ,  $\mathbb{P}(P|S) = 0.05$ ,  $\mathbb{P}(N|M) = 0.02$ , da cui  $\mathbb{P}(S) = 0.99$ ,  $\mathbb{P}(N|S) = 0.95$ ,  $\mathbb{P}(P|M) = 0.98$ . Dunque:

$$\mathbb{P}(M|P) = \frac{\mathbb{P}(M, P)}{\mathbb{P}(P)} = \frac{\mathbb{P}(M) \cdot \mathbb{P}(P|M)}{\mathbb{P}(M) \cdot \mathbb{P}(P|M) + \mathbb{P}(S) \cdot \mathbb{P}(P|S)} = 0.1653 = 16.53\%$$

**Definizione 2.7**

Siano  $A_k \in \mathcal{A}$ ,  $k \in I$  con  $I$  discreto.  $(A_k)_{k \in I}$  è detto **partizione** di  $\Omega$  se:

- $A_k \cap A_l = \emptyset \quad \forall k \neq l$ ;
- $\bigcup_k A_k = \Omega$ ;
- $\mathbb{P}(A_k) > 0 \quad \forall k$ .

**Teorema 2.8** (*J-P Th 3.3*)

Siano  $A_1, \dots, A_n \in \mathcal{A}$ , con  $\mathbb{P}(A_1, \dots, A_n) > 0$ . Allora si ha che:

$$\mathbb{P}(A_1, \dots, A_n) = \mathbb{P}(A_n | A_1, \dots, A_{n-1}) \cdot \mathbb{P}(A_{n-1} | A_1, \dots, A_{n-2}) \cdots \mathbb{P}(A_2 | A_1) \cdot \mathbb{P}(A_1)$$

**Dimostrazione**

1.  $\mathbb{P}(A_k | A_{k-1}, \dots, A_1)$  è ben definito  $\forall k = 1, \dots, n$ .

Infatti  $A_1 \cap \dots \cap A_{k-1} \supseteq A_1 \cap \dots \cap A_n$ , e dunque:

$$\mathbb{P}(A_1, \dots, A_{k-1}) \geq \mathbb{P}(A_1, \dots, A_n) > 0$$

2. Dimostriamo ora l'uguaglianza.

$$\begin{aligned} & \mathbb{P}(A_n | A_1, \dots, A_{n-1}) \cdot \mathbb{P}(A_{n-1} | A_1, \dots, A_{n-2}) \cdots \mathbb{P}(A_2 | A_1) \cdot \mathbb{P}(A_1) \\ &= \frac{\mathbb{P}(A_n, \dots, A_1)}{\cancel{\mathbb{P}(A_{n-1}, \dots, A_1)}} \cdot \frac{\cancel{\mathbb{P}(A_{n-1}, \dots, A_1)}}{\cancel{\mathbb{P}(A_{n-2}, \dots, A_1)}} \cdots \frac{\cancel{\mathbb{P}(A_2, A_1)}}{\mathbb{P}(A_1)} \cdot \mathbb{P}(A_1) \\ &= \mathbb{P}(A_n, \dots, A_1) \end{aligned} \quad \square$$

**Teorema 2.9: formula delle probabilità totali** (*J-P Th 3.4*)

Siano  $(\Omega, \mathcal{A}, \mathbb{P})$  spazio di probabilità,  $(E_n)_n$  partizione discreta di  $\Omega$ ,  $A \in \mathcal{A}$ .

Allora:

$$\mathbb{P}(A) = \sum_n \mathbb{P}(A, E_n) = \sum_n \mathbb{P}(A | E_n) \cdot \mathbb{P}(E_n)$$

**Dimostrazione**

Poiché  $(A \cap E_k) \cap (A \cap E_l) = \emptyset \ \forall k \neq l$ :

$$\begin{aligned} \mathbb{P}(A) &= \mathbb{P}(A \cap \Omega) = \mathbb{P}\left(A \cap \left(\bigcup_n E_n\right)\right) = \mathbb{P}\left(\bigcup_n (A \cap E_n)\right) \\ &= \sum_n \mathbb{P}(A \cap E_n) = \sum_n \mathbb{P}(A | E_n) \cdot \mathbb{P}(E_n) \end{aligned} \quad \square$$

**Teorema 2.10: formula di Bayes** (*J-P Th 3.5*)

Siano  $(\Omega, \mathcal{A}, \mathbb{P})$  spazio di probabilità,  $(E_n)_n$  partizione discreta di  $\Omega$ ,  $A \in \mathcal{A}$ ,  $\mathbb{P}(A) > 0$ . Allora:

$$\mathbb{P}(E_k | A) = \frac{\mathbb{P}(A | E_k) \cdot \mathbb{P}(E_k)}{\mathbb{P}(A)} = \frac{\mathbb{P}(A | E_k) \cdot \mathbb{P}(E_k)}{\sum_n \mathbb{P}(A | E_n) \cdot \mathbb{P}(E_n)}$$

La prima uguaglianza è il corollario 2.6. La seconda uguaglianza è una semplice conseguenza della formula delle probabilità totali.

**Teorema 2.11**

Sia  $(A_n)_n$  una partizione discreta di  $\Omega$ , e sia  $\mathcal{A} = \sigma(A_n : n \in I)$ , con  $I$  insieme arbitrario di indici.

Allora  $\exists!$   $\mathbb{P}$  su  $(\Omega, \mathcal{A})$ , con  $\mathbb{P}(A_n) = p_n \ \forall n \in I$ , tale che  $p_n \geq 0$  e  $\sum_{n \in I} p_n = 1$ ; ovvero, una misura di probabilità  $\mathbb{P}$  su  $(\Omega, \mathcal{A})$  è *caratterizzata* dai  $\mathbb{P}(A_n)$  con  $n \in I$ .

**Teorema 2.12**

Siano  $(A_n)_n$  e  $(B_k)_k$  partizioni discrete di  $\Omega$ .

Siano inoltre:

- $p_n$  tale che  $p_n \geq 0 \ \forall n$  e  $\sum_n p_n = 1$ ;
- $q_{k|n}$  tale che,  $\forall n \in \mathbb{N}$  fissato,  $q_{k|n} \geq 0 \ \forall k$  e  $\sum_k q_{k|n} = 1$ .

Allora  $\exists!$   $\mathbb{P}$  su  $\sigma((A_n)_n, (B_n)_n)$  tale che,  $\forall n, k \in \mathbb{N}$ , si ha che:

$$\mathbb{P}(A_n) = p_n \quad \text{e} \quad \mathbb{P}(B_k | A_n) = q_{k|n}$$

*Esempio.* Tre carte: la carta 1, nera su entrambi i lati, la carta 2, bianca su entrambi i lati, e la carta 3, con un lato bianco e uno nero. Pesca una carta e la poggio sul tavolo e il lato superiore è nero. Qual è la probabilità che il lato inferiore sia nero?

Considero gli eventi  $B$  “lato superiore bianco”,  $N = B^C$  “lato superiore nero”,  $A_i$  “esce la carta  $i$ -esima”.  $\{N, B\}$  e  $\{A_1, A_2, A_3\}$  sono due partizioni di  $\Omega$ . Definiamo

$\mathcal{A} = \sigma(A_1, A_2, A_3, N, B)$ . Allora:

$$\begin{array}{lll}
 \mathbb{P}(A_1) = \frac{1}{3} & \mathbb{P}(B|A_1) = 0 & \mathbb{P}(A_1, B) = 0 \\
 & \mathbb{P}(N|A_1) = 1 & \mathbb{P}(A_1, N) = \frac{1}{3} \\
 \mathbb{P}(A_2) = \frac{1}{3} & \mathbb{P}(B|A_2) = 1 & \mathbb{P}(A_2, B) = \frac{1}{3} \\
 & \mathbb{P}(N|A_2) = 0 & \mathbb{P}(A_2, N) = 0 \\
 \mathbb{P}(A_3) = \frac{1}{3} & \mathbb{P}(B|A_3) = \frac{1}{2} & \mathbb{P}(A_3, B) = \frac{1}{6} \\
 & \mathbb{P}(N|A_3) = \frac{1}{2} & \mathbb{P}(A_3, N) = \frac{1}{6}
 \end{array}$$

Quindi possiamo calcolare  $\mathbb{P}(A_1|N)$  come:

$$\begin{aligned}
 \mathbb{P}(A_1|N) &= \frac{\mathbb{P}(N|A_1) \cdot \mathbb{P}(A_1)}{\mathbb{P}(N)} \\
 &= \frac{\mathbb{P}(N|A_1) \cdot \mathbb{P}(A_1)}{\mathbb{P}(N|A_1) \cdot \mathbb{P}(A_1) + \mathbb{P}(N|A_2) \cdot \mathbb{P}(A_2) + \mathbb{P}(N|A_3) \cdot \mathbb{P}(A_3)}
 \end{aligned}$$

Sostituendo i valori numerici si ottiene  $\frac{1/3}{1/3} = \frac{2}{3}$ .

---

### 3 Costruzione di una probabilità

L'obiettivo di questo capitolo è quello di definire una probabilità su un'algebra semplice e di mostrare, con l'aiuto di alcuni teoremi, che la sua *estensione* a una  $\sigma$ -algebra esiste ed è unica. Da qui il concetto di *costruzione* di una probabilità: partendo da una collezione ristretta di eventi (come un'algebra), corrispondenti per esempio agli eventi verificabili sperimentalmente nel mondo reale, si è in grado di estendere le loro proprietà a una classe più ampia di eventi (come una  $\sigma$ -algebra).

#### 3.1 Estensione di una probabilità

Riformuliamo la definizione di probabilità su un'algebra invece che su una  $\sigma$ -algebra:

##### Definizione 3.1

Siano  $(\Omega, \mathcal{A})$  spazio misurabile,  $\mathcal{A}_0 \subseteq 2^\Omega$  algebra su  $\Omega$ .

$\mathbb{P} : \mathcal{A}_0 \rightarrow [0, 1]$  è una probabilità se valgono le seguenti condizioni:

1.  $\mathbb{P}(\Omega) = 1$ ;
2.  $\mathbb{P}$  è  $\sigma$ -additiva: dati  $A_n \in \mathcal{A}_0 \forall n \in \mathbb{N}$ , allora:

$$A_k \cap A_l = \emptyset \quad \forall k \neq l, \quad \bigcup_n A_n \in \mathcal{A}_0 \implies \mathbb{P}\left(\bigcup_n A_n\right) = \sum_n \mathbb{P}(A_n)$$

##### Teorema 3.2 (J-P Th 6.1)

Siano  $\Omega$  spazio campionario,  $\mathcal{A}_0$  algebra su  $\Omega$ ,  $\mathbb{P}_0$  probabilità su  $\mathcal{A}_0$ .

Allora  $\exists!$  probabilità  $\mathbb{P}$  su  $\mathcal{A} = \sigma(\mathcal{A}_0)$  che estende  $\mathbb{P}_0$ , ovvero tale che:

$$\mathbb{P}(A) = \mathbb{P}_0(A) \quad \forall A \in \mathcal{A}_0$$

Il teorema è un'immediata conseguenza del corollario 3.5, che introdurremo poco più avanti.

##### Definizione 3.3

Una collezione  $\mathcal{C} \subseteq 2^\Omega$ :

- è chiusa per intersezioni finite se:

$$A_1, \dots, A_n \in \mathcal{C} \implies \bigcap_{k=1}^n A_k \in \mathcal{C}$$

- è chiusa per limiti crescenti se:

$$\{A_n\}_{n \in \mathbb{N}} \subseteq \mathcal{C}, \quad A_n \uparrow A \implies A \in \mathcal{C}$$

- è chiusa per differenze se:

$$A, B \in \mathcal{C}, A \subseteq B \implies B \setminus A = B \cap A^C \in \mathcal{C}$$

**NB:** Una  $\sigma$ -algebra è chiusa per intersezioni finite, limiti crescenti e differenze, mentre un'algebra è chiusa per intersezioni finite e differenze, ma in generale non per limiti crescenti.

### 3.1.1 Classi monotone e corollari

**Teorema 3.4: classi monotone** (*J-P Th 6.2*)

Sia  $\mathcal{C} \subseteq 2^\Omega$  chiusa per intersezioni finite, con  $\Omega \in \mathcal{C}$ .

Sia inoltre  $\mathcal{D} \subseteq 2^\Omega$  la più piccola collezione di insiemi tale che  $\mathcal{D} \supseteq \mathcal{C}$ , chiusa per limiti crescenti e per differenze. Allora:

$$\mathcal{D} = \sigma(\mathcal{C})$$

Il teorema è anche noto come *lemma di Dynkin*.

#### Dimostrazione

La dimostrazione si svolgerà secondo i seguenti punti:

1.  $\mathcal{D}$  esiste ed è ben definita;
  2. definizione della classe  $\mathcal{D}_B$ ;
  3.  $\mathcal{D}_B$  è chiusa per limiti crescenti;
  4.  $\mathcal{D}_B$  è chiusa per differenze;
  5.  $\mathcal{D}$  è chiusa per intersezioni finite particolari;
  6.  $\mathcal{D}$  è chiusa per intersezioni finite;
  7.  $\mathcal{D}$  è una  $\sigma$ -algebra;
  8. conclusione.
1. Sia  $\mathcal{D}_i$  (con  $i \in I$ ) un gruppo di collezioni contenute in  $2^\Omega$  che rispettano tutte le proprietà richieste a  $\mathcal{D}$  nelle ipotesi (ovvero  $\mathcal{D}_i \supseteq \mathcal{C}$  e  $\mathcal{D}_i$  chiusa per limiti crescenti e per differenze,  $\forall i$ ). Analizziamo le proprietà della loro intersezione  $\mathcal{D} := \bigcap_i \mathcal{D}_i$ :
    - (a) Sia  $A \in \mathcal{C}$ . Allora  $A \in \mathcal{D}_i \forall i \implies A \in \mathcal{D} \implies \mathcal{D} \supseteq \mathcal{C}$ .
    - (b) Siano  $A_n \in \mathcal{D} \forall n \in \mathbb{N}$ , tali che  $A_n \uparrow A$ . Allora  $A_n \in \mathcal{D}_i \forall i \forall n \implies A = \bigcup_n A_n \in \mathcal{D}_i \forall i \implies A \in \mathcal{D}$  Dunque  $\mathcal{D}$  è chiusa per limiti crescenti.
    - (c) Sia  $A \subseteq B$  con  $A, B \in \mathcal{D}$ . Allora  $B \setminus A \subseteq A \in \mathcal{D} \implies B \setminus A \in \mathcal{D}$ .



Posso dunque prendere  $\mathcal{D} = \bigcap_i \mathcal{D}_i$ , che esiste e rispetta le tre proprietà richieste.

2. Sia  $B \in \mathcal{D}$  fissato arbitrariamente. Definiamo la seguente sottoclasse:

$$\mathcal{D}_B := \{A \in \mathcal{D} : A \cap B \in \mathcal{D}\}$$

Notiamo che  $\mathcal{D}_B \subseteq \mathcal{D}$ .

3. Siano  $A_n \in \mathcal{D}_B \forall n$  tali che  $A_n \uparrow A$ . Poiché  $A_n \in \mathcal{D}_B \subseteq \mathcal{D}$ , allora  $A_n \in \mathcal{D}$ . Ma essendo  $\mathcal{D}$  chiusa per limiti crescenti,  $A \in \mathcal{D}$ . Inoltre:

$$A \cap B = \left( \bigcup_n A_n \right) \cap B = \bigcup_n (A_n \cap B) \in \mathcal{D}$$

Perché  $\mathcal{D}$  è chiusa per limiti crescenti.

Dunque  $A \in \mathcal{D}_B$  per definizione di  $\mathcal{D}_B$ : in altre parole,  $\mathcal{D}_B$  è chiusa per limiti crescenti.

4. Siano  $A_1, A_2 \in \mathcal{D}_B$  tali che  $A_1 \subseteq A_2$ .  
 $A_2 \setminus A_1 \in \mathcal{D}$ , perché  $A_2 \in \mathcal{D}_B \subseteq \mathcal{D}$ ,  $A_1 \in \mathcal{D}_B \subseteq \mathcal{D}$  e  $\mathcal{D}$  è chiusa per differenze. Inoltre:

$$(A_2 \setminus A_1) \cap B = (A_2 \cap A_1^C) \cap B = (A_2 \cap B) \setminus (A_1 \cap B) \implies (A_2 \setminus A_1) \cap B \in \mathcal{D}$$

Quindi entrambi gli insiemi appartengono a  $\mathcal{D}_B \subseteq \mathcal{D}$ , la quale è chiusa per differenze. Pertanto,  $(A_2 \setminus A_1) \in \mathcal{D}_B$  per definizione di  $\mathcal{D}_B$ , ovvero  $\mathcal{D}_B$  è chiusa per differenze.

5. Dimostriamo ora che  $\mathcal{D} = \mathcal{D}_C \forall C \in \mathcal{C}$ , ovvero che  $\mathcal{D}$  è chiusa per intersezioni del tipo  $D \cap C$ , con  $D \in \mathcal{D}$  e  $C \in \mathcal{C}$ .

Sia  $A \in \mathcal{C}$ . Per definizione di  $\mathcal{D}$ ,  $A \in \mathcal{D}$ ; inoltre  $C \in \mathcal{C} \subseteq \mathcal{D}$ . Allora  $A \cap C \in \mathcal{D}$ , e di conseguenza  $A \in \mathcal{D}_C$  per definizione di  $\mathcal{D}_C$ . Poiché  $A$  è un generico elemento di  $\mathcal{C}$ , troviamo che  $\mathcal{C} \subseteq \mathcal{D}_C$ , e come già sappiamo,  $\mathcal{D}_C \subseteq \mathcal{D}$ .  $\mathcal{D}_C$  rispetta le tre proprietà richieste per  $\mathcal{D}$  ( $\mathcal{D}_C \subseteq \mathcal{C}$ , chiusura per limiti crescenti e differenze). Ma allora deve essere  $\mathcal{D}_C = \mathcal{D}$ .

6. Sia  $A \in \mathcal{C}$ . Per definizione di  $\mathcal{D}$ ,  $A \in \mathcal{D}$ ; inoltre per il punto precedente  $A \cap B \in \mathcal{D}$ ; ma allora  $A \in \mathcal{D}_B$ . Ciò significa che  $\mathcal{C} \subseteq \mathcal{D}_B \subseteq \mathcal{D}$ . Ma  $\mathcal{D}_C$  rispetta le tre proprietà di  $\mathcal{D}$ , e dunque deve essere  $\mathcal{D}_C = \mathcal{D}$ .

7. Verifichiamo le tre proprietà caratteristiche delle  $\sigma$ -algebre:

(a)  $\Omega \in \mathcal{D}$ , poiché  $\Omega \in \mathcal{C}$  per ipotesi, e  $\mathcal{C} \in \mathcal{D}$ .  $\emptyset \in \mathcal{D}$ , poiché  $\emptyset = \Omega \setminus \Omega$  e  $\mathcal{D}$  è chiusa per differenze.

(b)  $A \in \mathcal{D} \implies A^C = \Omega \setminus A \in \mathcal{D}$ , poiché  $\mathcal{D}$  è chiusa per differenze.

(c) ( $\sigma$ -additività) Siano  $A_n \in \mathcal{D} \quad \forall n \in \mathbb{N}$ .

Dopo aver definito la successione crescente  $B_k = \bigcup_{n=1}^k A_n$ , abbiamo che  $\bigcup_n A_n = \bigcup_k B_k$ . Ma  $B_k \in \mathcal{D}$ , poiché:

$$B_k = \bigcup_{n=1}^k A_n = \left( \bigcap_{n=1}^k A_n^C \right)^C$$

e  $\mathcal{D}$  è chiusa per intersezioni finite.

Abbiamo poi che:

$$B_k \uparrow \bigcup_{k=1}^{+\infty} B_k = \bigcup_{n=1}^{+\infty} A_n$$

Questa unione appartiene a  $\mathcal{D}$  poiché esso è chiuso per limiti crescenti. È dunque dimostrata la chiusura per unione numerabile; quella per intersezione numerabile si dimostra in maniera analoga.

$\mathcal{D}$  è dunque una  $\sigma$ -algebra.

8.  $\mathcal{D}$  deve per forza coincidere con  $\sigma(\mathcal{C})$ : infatti, se  $\mathcal{D}$  è una  $\sigma$ -algebra, allora  $\mathcal{D} \supseteq \sigma(\mathcal{C})$  per definizione di  $\sigma(\mathcal{C})$  e  $\mathcal{D} \subseteq \sigma(\mathcal{C})$  per definizione di  $\mathcal{D}$ .  $\square$

### Corollario 3.5

Siano  $(\Omega, \mathcal{A})$  spazio misurabile,  $\mathbb{P}$  e  $\mathbb{Q}$  probabilità su  $\mathcal{A}$ , e  $\mathcal{C} \subseteq \mathcal{A}$  classe chiusa per intersezioni finite e tale che  $\sigma(\mathcal{C}) = \mathcal{A}$ . Se  $\mathbb{P}(A) = \mathbb{Q}(A) \quad \forall A \in \mathcal{C}$  (ovvero, usando una notazione alternativa,  $\mathbb{P}|_{\mathcal{C}} = \mathbb{Q}|_{\mathcal{C}}$ ), allora  $\mathbb{P}(A) = \mathbb{Q}(A) \quad \forall A \in \mathcal{A}$ , ovvero  $\mathbb{P} \equiv \mathbb{Q}$ .

In altre parole, è garantita l'*unicità dell'estensione* di una probabilità su una classe che rispetti le opportune ipotesi: se i valori di  $\mathbb{P}$  e  $\mathbb{Q}$  coincidono sugli insiemi della suddetta classe (la quale è un sottoinsieme del dominio), allora coincideranno anche su tutto il dominio  $\mathcal{A}$ , ovvero le due probabilità devono necessariamente essere la stessa.

### Dimostrazione

Sia  $\tilde{\mathcal{C}} = \mathcal{C} \cup \{\Omega\}$  che è ovviamente chiusa per intersezioni finite. Si definisca ora:

$$\tilde{\mathcal{D}} := \{A \in \mathcal{A} : \mathbb{P}(A) = \mathbb{Q}(A)\}$$

1.  $\tilde{\mathcal{D}} \supseteq \tilde{\mathcal{C}}$  perché  $\tilde{\mathcal{D}} \supseteq \mathcal{C}$  per ipotesi e  $\Omega \in \tilde{\mathcal{D}}$  in quanto  $\mathbb{P}(\Omega) = 1 = \mathbb{Q}(\Omega)$ .
2. Siano  $A_n \in \tilde{\mathcal{D}} \quad \forall n$ , con  $A_n \uparrow A$ . Allora  $\mathbb{P}(A_n) = \mathbb{Q}(A_n) \quad \forall n$ . Prendendo il limite per  $n \rightarrow +\infty$  di entrambi i membri,  $\mathbb{P}(A) = \mathbb{Q}(A) \implies A \in \tilde{\mathcal{D}}$ . Ergo,  $\tilde{\mathcal{D}}$  è chiusa per limiti crescenti.

3. Siano  $A, B \in \tilde{\mathcal{D}}$  tali che  $A \subseteq B$ . Allora:

$$\mathbb{P}(B \setminus A) = \mathbb{P}(B) - \mathbb{P}(A) = \mathbb{Q}(B) - \mathbb{Q}(A) = \mathbb{Q}(B \setminus A) \implies B \setminus A \in \tilde{\mathcal{D}}$$

Quindi  $\tilde{\mathcal{D}}$  è chiusa per differenze.

Grazie a queste 3 proprietà è possibile invocare il teorema delle classi monotone:  $\tilde{\mathcal{D}} \supseteq \sigma(\tilde{\mathcal{C}})$ . Il segno di inclusione è necessario in quanto sappiamo che  $\tilde{\mathcal{D}}$  rispetta le 3 proprietà, ma non è detto che sia essa stessa la classe più piccola che le soddisfa, ovvero  $\sigma(\tilde{\mathcal{C}})$ : in ogni caso la prima contiene sicuramente la seconda. Ora, poiché  $\sigma(\tilde{\mathcal{C}}) = \sigma(\mathcal{C}) = \mathcal{A}$ , si ottiene che  $\mathbb{P}(A) = \mathbb{Q}(A) \forall A \in \mathcal{A}$ .  $\square$

**Corollario 3.6: criterio di Caratheodory**

Siano  $\Omega$  spazio campionario,  $\mathcal{A}_0$  algebra su  $\Omega$ ,  $\mathcal{A} = \sigma(\mathcal{A}_0)$ ,  $\mathbb{P}$  e  $\mathbb{Q}$  probabilità su  $\mathcal{A}$ . Se  $\mathbb{P}(A) = \mathbb{Q}(A) \forall A \in \mathcal{A}_0$ , allora  $\mathbb{P}(A) = \mathbb{Q}(A) \forall A \in \mathcal{A}$ .

Questo è un'immediata conseguenza del corollario precedente (3.5) per il fatto che qualsiasi algebra è, per definizione, chiusa per intersezioni finite. A parte questa minima variazione nelle ipotesi, i due corollari sono identici: quest'ultimo afferma che l'estensione di una probabilità da un'algebra alla corrispondente  $\sigma$ -algebra è unica. È dunque sufficiente essere in possesso di un'informazione su un ristretto gruppo di insiemi (due probabilità sono identiche su un'algebra) per poter affermare una proprietà generale dell'intera  $\sigma$ -algebra generata (esse sono identiche su tutto il dominio).

*Controesempio.* Mostriamo che l'ipotesi della chiusura di  $\mathcal{C}$  per intersezioni finite è essenziale per l'applicazione dei corollari mediante un controesempio in cui questa ipotesi è mancante.

Considerando una  $\Omega$  qualsiasi e delle sue partizioni  $A_1, A_2, B_1, B_2$  e dati  $\mathcal{A} = \sigma(A_1, A_2, B_1, B_2)$ ,  $\mathbb{P}$  e  $\mathbb{Q}$  probabilità su  $\mathcal{A}$ , si ha:

$$\mathbb{P}(A_k) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2} \stackrel{\downarrow}{=} \frac{1}{2} + 0 = \mathbb{Q}(A_k) \quad \forall k = 1, 2$$

$\mathbb{P}$	$B_1$	$\frac{1}{4}$	$\frac{1}{4}$
	$B_2$	$\frac{1}{4}$	$\frac{1}{4}$
		$A_1$	$A_2$

$\mathbb{Q}$	$B_1$	0	$\frac{1}{2}$
	$B_2$	$\frac{1}{2}$	0
		$A_1$	$A_2$

In questo caso dunque si può effettivamente affermare che  $\mathbb{P}(A) = \mathbb{Q}(A) \forall A \in \mathcal{C}$  e che  $\mathbb{P}(B) = \mathbb{Q}(B) \forall B \in \mathcal{C}$ , ma il corollario alle classi monotone è inapplicabile

perché  $\mathcal{C}$  non è chiusa per intersezioni finite (come richiesto dalle ipotesi): per esempio  $A_1 \cap B_2 \notin \mathcal{C}$ . Infatti in questo caso si può facilmente verificare che  $\mathbb{P} \neq \mathbb{Q}$ .

**NB:** Possono esistere esperimenti diversi che hanno la stessa probabilità, il fatto che  $\mathbb{P} : \mathcal{A} \rightarrow [0, 1]$  sia uguale non garantisce che per ogni esito elementare la probabilità sia uguale.

*Esempio:* lancio di due dadi. Preso un d8<sup>7</sup> non truccato, si consideri la  $\sigma$ -algebra  $\mathcal{A} = \{\emptyset, \text{pari}, \text{dispari}, \Omega\}$ , di conseguenza la probabilità  $\mathbb{P}$  è:

$\mathcal{A}$	$\emptyset$	pari	dispari	$\Omega$
$\mathbb{P}$	0	$\frac{1}{2}$	$\frac{1}{2}$	1

Si consideri un secondo d8, truccato, in cui escono uniformemente solo gli esiti da 1 a 4, mentre gli altri hanno probabilità zero. Scegliendo la stessa  $\sigma$ -algebra si ottiene la probabilità  $\mathbb{Q}$  che è identica a  $\mathbb{P}$ .

Nonostante i due esperimenti abbiano esiti elementari  $\omega$  diversi, la probabilità è uguale perché questa non dipende direttamente da  $\Omega$ .

## 3.2 Probabilità su spazi finiti e discreti

### Definizione 3.7

Sia dato uno spazio probabilistico  $(\Omega, \mathcal{A}, \mathbb{P})$  in cui:

- $\Omega$  è uno spazio campionario finito e discreto tale che  $\#\Omega = n$ ;
- $\mathcal{A}$  è la  $\sigma$ -algebra definita su  $\Omega$  come  $\mathcal{A} = 2^\Omega$ ;
- $\mathbb{P}$  è la probabilità costruita sulla  $\sigma$ -algebra  $\mathcal{A}$  definita come:

$$\mathbb{P}(\{\omega\}) = \frac{1}{\#\Omega} \text{ e con } \mathbb{P}(\{A\}) = \frac{\#A}{\#\Omega}.$$

Tale spazio probabilistico si dice **spazio probabilistico discreto e uniforme**, cioè lo spazio è discreto, finito e tutti gli esiti sono equiprobabili.

---

<sup>7</sup>Ovvero “dado a 8 facce”, per i tre lettori che non sanno cosa siano i giochi di ruolo.

---

## 4 Probabilità su $\mathbb{R}$

L'insieme dei numeri reali  $\mathbb{R}$  è, senza troppe sorprese, un caso particolarmente importante nello studio della probabilità, e quello verso il quale sono orientate gran parte dell'attenzione e della teoria. Ciò è dovuto innanzitutto alla comodità dell'utilizzo dei numeri (strumento ormai ben conosciuto e sviluppato) per rappresentare un generico insieme, pratica peraltro utilizzata pressoché in ogni branca della matematica. In questo capitolo si analizzeranno le altre ragioni di questa scelta, ovvero le utili proprietà che caratterizzano  $\mathbb{R}$  e la possibilità di introdurre nuovi strumenti di calcolo, tra loro legati: la *funzione di ripartizione*, che “raccolle” la probabilità totale al crescere del suo argomento, e la *densità di probabilità*, che descrive il peso di ogni singolo punto di  $\mathbb{R}$  nel determinare la probabilità di un evento. Saranno infine forniti alcuni esempi notevoli di distribuzioni molto usate nello studio della materia.

### 4.1 Funzione di ripartizione

Per ricapitolare i teoremi dei capitoli precedenti:

1. per generare  $\mathcal{A}$ : è sufficiente definire  $\mathcal{C} \subset \mathcal{A}$  tale che  $\sigma(\mathcal{C}) = \mathcal{A}$ ;
2. per caratterizzare  $\mathbb{P}$  su  $\mathcal{A}$ : è sufficiente definire  $\mathcal{C} \subset \mathcal{A}$  tale che  $\sigma(\mathcal{C}) = \mathcal{A}$  con  $\mathcal{C}$  chiuso per intersezioni finite;
3. per costruire  $\mathbb{P}$  su  $\mathcal{A}$ : è sufficiente definire  $\mathcal{C} \subset \mathcal{A}$  tale che  $\sigma(\mathcal{C}) = \mathcal{A}$  con  $\mathcal{C}$  algebra.

#### Definizione 4.1

Sia  $(\Omega, \mathcal{A}) = (\mathbb{R}, \mathcal{B})$  e sia  $\mathbb{P} : \mathcal{B} \rightarrow [0, 1]$  una probabilità su  $\mathcal{B}$ .

La **funzione di ripartizione** di  $\mathbb{P}$  è definita come:

$$F : \mathbb{R} \rightarrow [0, 1], \quad F(x) := \mathbb{P}((-\infty, x]).$$

#### Teorema 4.2 (J-P Th 7.1)

Siano  $\mathbb{P}$  probabilità su  $(\mathbb{R}, \mathcal{B})$  e  $F$  funzione di ripartizione, allora  $F$  caratterizza  $\mathbb{P}$ , ovvero dati  $\mathbb{P}_1$  e  $\mathbb{P}_2$  probabilità su  $\mathcal{B}$  allora:

$$\mathbb{P}_1 = \mathbb{P}_2 \iff F_1 = F_2$$

#### Dimostrazione

La dimostrazione di  $(\implies)$  è ovvia.

Dimostriamo  $(\impliedby)$ : Sia  $\mathcal{C} = \{(-\infty, q], q \in \mathbb{Q}\}$ . Abbiamo già dimostrato che  $\sigma(\mathcal{C}) = \mathcal{B}$ .

Mostriamo che  $\mathcal{C}$  è chiuso per intersezioni finite:

$$\underbrace{(-\infty, q_1]}_{\in \mathcal{C}} \cap \underbrace{(-\infty, q_2]}_{\in \mathcal{C}} = \underbrace{(-\infty, \min(q_1, q_2)]}_{\in \mathcal{C}}$$

Allora:

$$\begin{aligned} F_1 = F_2 &\iff F_1(x) = F_2(x) \quad \forall x \in \mathbb{R} \\ &\implies F_1(q) = F_2(q) \quad \forall q \in \mathbb{Q} \\ &\iff \mathbb{P}_1((-\infty, q]) = \mathbb{P}_2((-\infty, q]) \quad \forall q \in \mathbb{Q} \\ &\iff \mathbb{P}_1(B) = \mathbb{P}_2(B) \quad \forall B \in \mathcal{C} \\ &\iff \mathbb{P}_1 = \mathbb{P}_2 \text{ su } B \in \sigma(\mathcal{C}) \end{aligned}$$

Abbiamo usato il corollario 3.5 delle classi monotone. □

#### 4.1.1 Caratterizzazione

**Teorema 4.3** (*J-P Th 7.2*)

$F : \mathbb{R} \rightarrow [0, 1]$  è funzione di ripartizione di una (e una sola) probabilità  $\mathbb{P} : \mathcal{B} \rightarrow [0, 1]$  se e solo se:

1.  $F$  è non-decrescente;
2.  $F$  è continua da destra;
3.  $\lim_{x \rightarrow -\infty} F(x) = 0$ ,  $\lim_{x \rightarrow +\infty} F(x) = 1$ .

#### Dimostrazione

Dimostriamo ( $\implies$ ):

1.  $F$  è non-decrescente:  $x_1 < x_2 \implies F(x_1) \leq F(x_2)$   
 Infatti, poiché  $(-\infty, x_1] \subseteq (-\infty, x_2]$ , per il corollario 1.16 si ha:

$$F(x_1) = \mathbb{P}((-\infty, x_1]) \leq \mathbb{P}((-\infty, x_2]) = F(x_2)$$

2.  $F$  continua da destra:  $\forall x, \quad \forall x_n \downarrow x \quad F(x) = \lim_n F(x_n)$

Infatti:

$$(-\infty, x_n] \supseteq (-\infty, x_{n+1}] \implies (-\infty, x_n) \downarrow \bigcap_n (-\infty, x_n] = (-\infty, x]$$

Per dimostrare l'ultima uguaglianza basti osservare che

$$y \in \bigcap_n (-\infty, x_n] \implies y \in (-\infty, x_n] \quad \forall n$$

Allora,  $y \leq x_n \quad \forall n$  e dunque, per il teorema dei due carabinieri,  $y \leq x \leq x_n$ .

Infine:

$$F(x_n) = \mathbb{P}((-\infty, x_n]) \downarrow \mathbb{P}((-\infty, x]) = F(x)$$

$$3. \lim_{x \rightarrow -\infty} F(x) = 0, \quad \lim_{x \rightarrow +\infty} F(x) = 1:$$

$$(a) x_n \downarrow -\infty \implies (-\infty, x_n] \downarrow \emptyset \implies F(x_n) = \mathbb{P}((-\infty, x_n]) \downarrow \mathbb{P}(\emptyset) = 0$$

$$(b) x_n \uparrow +\infty \implies (-\infty, x_n] \uparrow \mathbb{R} \implies F(x_n) = \mathbb{P}((-\infty, x_n]) \uparrow \mathbb{P}(\mathbb{R}) = 1$$

Dimostriamo ( $\Leftarrow$ )<sup>8</sup>:

Sia  $F : \mathbb{R} \rightarrow [0, 1]$  che soddisfa (1), (2), (3).

Definiti inoltre:

$$\mathcal{B}_0 = \left\{ B = \bigcup_{k=1}^n (x_k, y_k] \mid n \in \mathbb{N}, -\infty \leq x_1 \leq y_1 \leq x_2 \leq y_2 \leq \dots \leq +\infty \right\}$$

e  $\mathbb{P}_0$  tale che:

$$\mathbb{P}_0 : \mathcal{B}_0 \rightarrow [0, 1], \quad \mathbb{P}_0 \left( \bigcup_{k=1}^n (x_k, y_k] \right) = \sum_{k=1}^n [F(y_k) - F(x_k)]$$

dove  $F(-\infty) = 0$  e  $F(+\infty) = 1$

a)  $\mathbb{P}_0$  è ben definita, perché  $\mathbb{P}_0((a, a]) = 0$  e  $\mathbb{P}_0((a, b] \cup (b, c]) = F(c) - F(a)$

b)  $\mathbb{P}_0(\mathbb{R}) = 1$  perché

$$\mathbb{P}_0(\mathbb{R}) = \mathbb{P}_0((-\infty, +\infty)) = \mathbb{P}_0((-\infty, +\infty]) = F(+\infty) - F(-\infty) = 1 - 0 = 1$$

c)  $\mathbb{P}_0$  è  $\sigma$ -additiva su  $\mathcal{B}_0$ .

Poiché  $\mathbb{P}_0$  è additiva,  $\mathbb{P}_0$  risulta  $\sigma$ -additiva se e solo se  $\mathbb{P}_0(A_n) \downarrow 0 \forall$  successione tale che:

$$A_n \in \mathcal{B}_0, A_n \downarrow \emptyset, A_l = \bigcup_{k=1}^{n_l} (x_k^l, y_k^l]$$

dove  $-\infty \leq x_1^l < y_1^l < x_2^l < \dots < y_{n_l}^l \leq +\infty$  e  $A_1 \supseteq A_2 \supseteq A_3 \dots$

Vorremmo passare a dei nuovi insiemi  $\bar{B}_l$  compatti, approssimando  $A_l$ , dove  $\bar{B}_l \downarrow \emptyset \implies \bar{B}_l = \emptyset$  definitivamente  $\implies \mathbb{P}_0(\bar{B}_l) = 0$  definitivamente.

Fissato  $\varepsilon > 0$  si mostra che  $\exists \mathcal{L} = \mathcal{L}(\varepsilon)$  tale che  $\mathbb{P}_0(A_l) \leq 3\varepsilon \forall l > \mathcal{L}$ .

1. Si trova  $z$  tale che  $F(-z) \leq \varepsilon$  e  $1 - F(z) \leq \varepsilon$  ovvero  $\mathbb{P}_0((-z, z]^C) \leq 2\varepsilon$ .

Allora si può mostrare che:

$$\lim_{x \rightarrow -\infty} F(x) = 0 \implies \exists z_1 \text{ t.c. } F(x) \leq \varepsilon \forall x \leq z_1$$

$$\lim_{x \rightarrow +\infty} F(x) = 1 \implies \exists z_2 \text{ t.c. } F(x) \geq 1 - \varepsilon \forall x \geq z_2$$

<sup>8</sup>Fuggite sciocchi, finché siete in tempo

Prendiamo quindi il massimo<sup>9</sup> dei due valori:

$$z = \max\{|z_1|, |z_2|\} = |z_1| \vee |z_2| \implies F(-z) \leq \varepsilon \quad F(z) \geq 1 - \varepsilon$$

Allora:

$$\begin{aligned} \mathbb{P}_0(A_l) &= \mathbb{P}_0(A_l \wedge (-z, z]^C) + \mathbb{P}_0(A_l \wedge (-z, +z]) \\ &\leq \mathbb{P}_0((-z, z]^C) + \mathbb{P}_0(A_l, (-z, z]) \\ &\leq 2\varepsilon + \mathbb{P}_0(A_l, (-z, z]) \end{aligned}$$

2.  $\forall l \in \mathbb{N}, \forall k = 1, \dots, n_l \exists a_k^l \in (x_k^l, y_k^l)$  tale che  $F(a_k^l) - F(x_k^l) \leq \frac{\varepsilon}{2^{l+k}}$  per la proprietà 2.
3. Possiamo quindi trovare dei  $B_l'$  che:

$$B_l' = \left( \bigcup_{k=1}^{n_l} (a_k^l, y_k^l) \right) \cap (-z, z] \quad \text{con } B_l' \in \mathcal{B}_0, B_l' \in A_l, \bar{B}_l' \subseteq A_l, \bar{B}_l' \subseteq [-z, z]$$

Allora:

$$\begin{aligned} \mathbb{P}_0(A_l \setminus \bar{B}_l') &\leq 2\varepsilon + \mathbb{P}_0((A_l \setminus B_l') \cap [-z, z]) \\ &\leq 2\varepsilon + \mathbb{P}_0\left(\bigcup_{k=1}^{n_l} [x_k^l, y_k^l] \setminus \bigcup_{k=1}^{n_l} [a_k^l, y_k^l]\right) \quad (\text{per definizione}) \\ &\leq 2\varepsilon + \mathbb{P}_0\left(\bigcup_{k=1}^{n_l} [x_k^l, a_k^l]\right) \\ &= 2\varepsilon + \sum_{k=1}^{n_l} \mathbb{P}_0([x_k^l, a_k^l]) \quad (\text{per } \sigma\text{-additività}) \\ &= 2\varepsilon + \sum_{k=1}^{n_l} (F(a_k^l) - F(x_k^l)) \quad (\text{per le prop. di } F) \end{aligned}$$

Passando ai  $B_l'$ :

$$\begin{aligned} \mathbb{P}_0(A_l \setminus B_l') &\leq 2\varepsilon + \sum_{k=1}^{n_l} \frac{\varepsilon}{2^{l+k}} \quad (\text{sfruttando il punto (2)}) \\ &\leq 2\varepsilon + \frac{\varepsilon}{2^l} \sum_{n=1}^{+\infty} \frac{1}{2^n} \quad (\text{serie a termini positivi}) \\ &\leq \left(2 + \frac{1}{2^l}\right) \varepsilon \quad (\text{la sommatoria vale 1}) \end{aligned}$$

Tuttavia i  $B_l'$  non sono inscatolati.

---

<sup>9</sup>Si noti la notazione alternativa per il massimo tra due elementi. Similmente, per il minimo tra due elementi  $x$  e  $y$  si scriverà  $\max\{x, y\} = x \wedge y$ .



4. Sia:

$$B_l = \bigcap_{m=1}^l B'_m \implies B_l \in \mathcal{B}_0, \quad B_l \in B'_l \subseteq A_l, \quad \bar{B}_l \subseteq \bigcap_{m=1}^{+\infty} \bar{B}'_m \subseteq \bar{B}'_l \subseteq A_l$$

Allora possiamo mostrare che  $B_l \supseteq B_{l+1}$  e  $\bigcap_l B_l \subseteq \bigcap_l A_l = \emptyset \implies B_l \downarrow \emptyset$ , quindi analogamente  $\bar{B}_l \downarrow \emptyset$

A questo punto, la differenza tra  $A_l$  e  $B_l$ :

$$\begin{aligned} A_l \setminus B_l &= A_l \cap B_l^C = A_l \cap \left( \bigcap_{m=1}^l B'_m \right)^C \\ &= A_l \cap \left( \bigcup_{m=1}^l B_m'^C \right) = \bigcup_{m=1}^l (A_l \cap B_l'^C) \\ &\subseteq \bigcup_{m=1}^l (A_m \cap B_m'^C) = \bigcup_{m=1}^l (A_m \setminus B'_m) \end{aligned}$$

dove per l'inclusione si è usato il fatto che gli  $A_l$  sono inscatolati. Possiamo quindi concludere che:

$$\begin{aligned} \mathbb{P}_0(A_l \setminus B_l) &\leq \mathbb{P}_0 \left( \bigcup_{m=1}^l (A_m \setminus B'_m) \right) \\ &\leq \sum_{m=1}^l \mathbb{P}_0 (A_m \setminus B'_m) && \text{(per subadditività)} \\ &\leq 2\varepsilon + \mathbb{P}_0 \left( \bigcup_{m=1}^l (A_m \setminus B'_m) \cap (-z, z] \right) && \text{(analogamente a (3))} \\ &\leq 2\varepsilon + \sum_{m=1}^l \mathbb{P}_0 ((A_m \setminus B'_m) \cap (-z, z]) && \text{(per } \sigma\text{-additività)} \\ &= 2\varepsilon + \frac{\varepsilon}{2^l} \leq 3\varepsilon && \text{(analogamente a (3))} \end{aligned}$$

5.  $\exists \mathcal{L}$  tale che  $B_l = \emptyset \forall l \geq \mathcal{L}$ . Infatti, i  $\bar{B}_l$  sono compatti e  $\bar{B}_l \downarrow \emptyset$ : se per assurdo esistesse  $l_j$  tale che  $\bar{B}_{l_j} \neq \emptyset$ , si avrebbe  $\forall j \ l_j \rightarrow +\infty$  allora  $\exists b_j \in \bar{B}_{l_j} \forall j \rightarrow b_j \in [-z, z] \forall j$ .

Quindi, per le proprietà dei compatti,  $\exists b_{j_i} \xrightarrow{i} b \in [-z, z]$ , ma  $b \in \bar{B}_l \forall l$  essendo  $b_j \in \bar{B}_{l_j}$  definitivamente,  $\bar{B}_{l_j}$  compatti allora  $\bigcap_l \bar{B}_l \ni b$ . Quest'ultima affermazione è un assurdo, quindi  $\bar{B}_l = \emptyset$  e  $B_l = \emptyset$  entrambe definitivamente.

6. Visti i risultati dei passaggi precedenti si può affermare che:

$$\begin{aligned}\mathbb{P}_0(A_l) &= \mathbb{P}_0(A_l) - \mathbb{P}_0(B_l) + \mathbb{P}_0(B_l) \quad (\text{aggiungendo e togliendo } \mathbb{P}_0(B_l)) \\ &= \mathbb{P}_0(A_l \setminus B_l) + \mathbb{P}_0(B_l) \\ &= \mathbb{P}_0(A_l \setminus B_l) \leq 3\varepsilon \quad (\text{se } l \geq \mathcal{L} \text{ allora } \mathbb{P}_0(B_l) = 0)\end{aligned}$$

Questo conclude la dimostrazione.  $\square$

#### 4.1.2 Proprietà

Introduciamo la notazione  $F(x^-) := \lim_{y \uparrow x} F(y)$ . Sicuramente  $F(x^-) \leq F(x)$ .

##### Corollario 4.4

Sia  $F$  funzione di ripartizione di  $\mathbb{P}$  su  $(\mathbb{R}, \mathcal{B})$ . Allora,  $\forall x < y$ :

1.  $\mathbb{P}((x, y]) = F(y) - F(x)$
2.  $\mathbb{P}([x, y]) = F(y) - F(x^-)$
3.  $\mathbb{P}((x, y)) = F(y^-) - F(x)$
4.  $\mathbb{P}([x, y)) = F(y^-) - F(x^-)$
5.  $\mathbb{P}(\{x\}) = F(x) - F(x^-)$

##### Dimostrazione

1. Per costruzione.
2.  $[x, y] = \bigcap_{n=1}^{+\infty} \left(x - \frac{1}{n}, y\right]$ . Dunque:

$$\begin{aligned}\mathbb{P}([x, y]) &= \lim_{n \rightarrow +\infty} \mathbb{P}\left(\left(x - \frac{1}{n}, y\right]\right) \\ &= \lim_{n \rightarrow +\infty} \left(F(y) - F\left(x - \frac{1}{n}\right)\right) = F(y) - F(x^-)\end{aligned}$$

Analogamente si verificano i punti (3), (4), (5).  $\square$

##### Definizione 4.5

Si definisce massa o **delta di Dirac** in  $\alpha$  (indicata con  $\delta_\alpha$ ) la probabilità così definita:

$$\mathbb{P}(A) = \begin{cases} 1 & \text{se } \alpha \in A \\ 0 & \text{se } \alpha \notin A \end{cases} \quad \text{con } \alpha \in \mathbb{R}, A \in 2^{\mathbb{R}}$$

La sua funzione di ripartizione è:

$$F(x) = \mathbb{P}((-\infty, x]) = \begin{cases} 1 & \text{se } x \geq \alpha \\ 0 & \text{se } x < \alpha \end{cases}$$

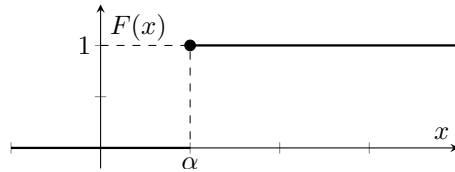


Figura 4.1: funzione di ripartizione della delta di Dirac

**NB:** Questa distribuzione è un caso degenere e verrà spesso utilizzata per i limiti di variabili aleatorie, quando queste tendono ad avere varianza nulla.

*Esempio.* Sia data  $\mathbb{P}$  probabilità discreta sui punti  $x_k$  ( $k \in \mathbb{N}$ ) con la seguente densità:

$$p_j \geq 0 \text{ tale che } \sum_{k=1}^{+\infty} p_k = 1$$

Per  $A \in 2^{\mathbb{R}}$ , ovvero non restringendo il campo ai soli boreliani:

$$\mathbb{P}(A) := \sum_{k: x_k \in A} p_k, \text{ da cui } F(x) := \sum_{k: x_k \leq x} p_k$$

Il grafico della  $F(x)$  sarà *a scala*, presentando discontinuità di tipo salto in corrispondenza delle  $x_k$ .

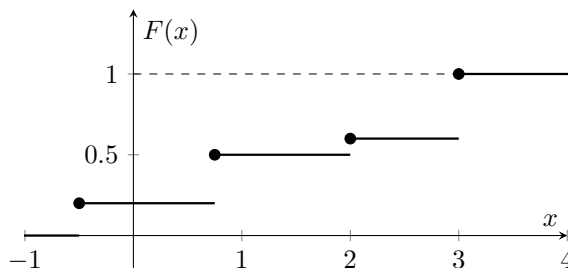


Figura 4.2: funzione di ripartizione di una variabile aleatoria discreta

*Esempi di probabilità discrete:*

- binomiale ( $\mathbb{P} = B(n, p)$ , con  $n \in \mathbb{N}$ ,  $0 \leq p \leq 1$ );
- Poisson ( $\mathbb{P} = Po(\lambda)$ , con  $\lambda \in \mathbb{R}$ ).

## 4.2 Densità continua di probabilità

### Definizione 4.6

Si dice **densità continua di probabilità** una  $f : \mathbb{R} \rightarrow [0, +\infty)$ , non negativa, Riemann-integrabile e tale che:

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

Si dice inoltre che  $F$  *ammette densità*  $f$  se vale:

$$F(x) = \int_{-\infty}^x f(t) dt$$

Osserviamo che, se  $F$  ammette densità  $f$ ,  $F'(x) = f(x) \forall x$  e  $F \in C^0$ , quindi ogni punto ha probabilità zero:

$$\mathbb{P}(\{x\}) = 0 \quad \forall x \in \mathbb{R}$$

Vedremo più avanti che i punti sono *insiemi di misura trascurabile* rispetto agli insiemi non numerabili.

Se  $f \in C^0$  a tratti, allora  $F \in C^1$  a tratti. Inoltre, se  $F \in C^0$ ,  $C^1$  a tratti,  $F$  ammette densità e  $f(x) = F'(x)$  per ogni  $x$  in cui  $F$  è derivabile (i punti di non derivabilità non influenzano  $f$ ).

Sia poi  $\mathbb{P}$  la probabilità su  $(\mathbb{R}, \mathcal{B})$  associata a  $F$ , allora:

$$\mathbb{P}((a, b)) = \mathbb{P}([a, b]) = \int_a^b f(t) dt = \int_{-\infty}^b f(t) dt - \int_{-\infty}^a f(t) dt$$

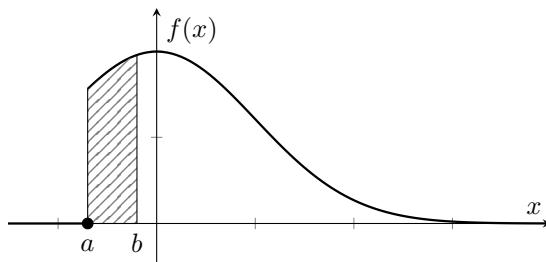


Figura 4.3: probabilità di un evento caratterizzato da  $f \in C^0$  a tratti. L'area tratteggiata rappresenta  $\mathbb{P}(a, b)$

### 4.2.1 Densità notevoli

Si rimanda alle appendici B e C a pagina 260 e 268 per una lista più completa delle densità notevoli.

- **Distribuzione uniforme:**

$$\mathbb{P} = U([a, b]), \quad f(s) = \frac{1}{b-a} \mathbb{1}_{[a,b]}(s)$$

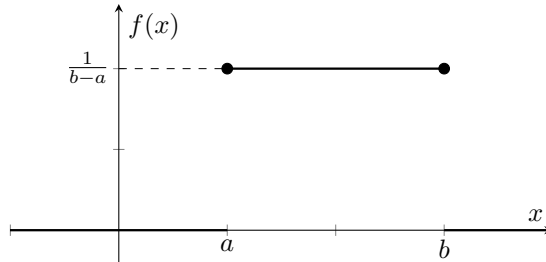


Figura 4.4: densità uniforme

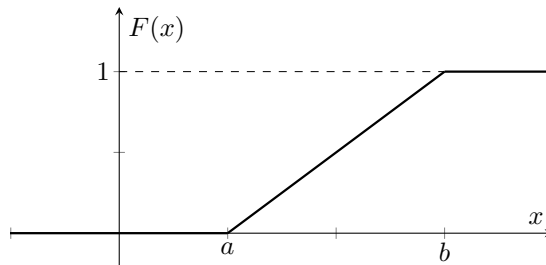


Figura 4.5: funzione di ripartizione uniforme

- **Distribuzione esponenziale:**

$$\mathbb{P} = \mathcal{E}(\lambda), \quad f(s) = \lambda e^{-\lambda s} \mathbb{1}_{[0,+\infty)}(s), \quad F(s) = (1 - e^{-\lambda s}) \mathbb{1}_{[0,+\infty)}(s)$$

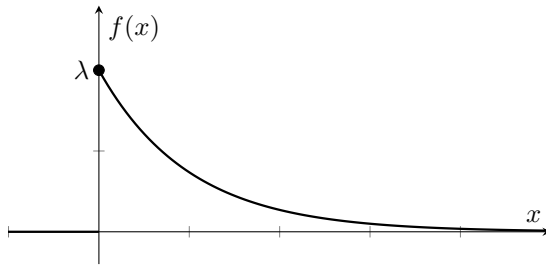


Figura 4.6: densità esponenziale

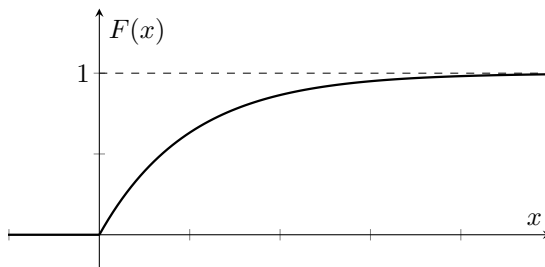


Figura 4.7: funzione di ripartizione esponenziale

- **Distribuzione gaussiana o normale:**

$$\mathbb{P} = N(\mu, \sigma^2), \quad f(s) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(s - \mu)^2}{2\sigma^2}\right\}$$

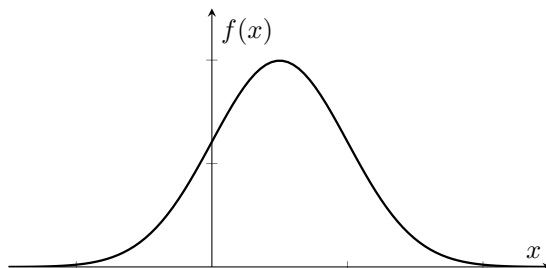


Figura 4.8: densità gaussiana

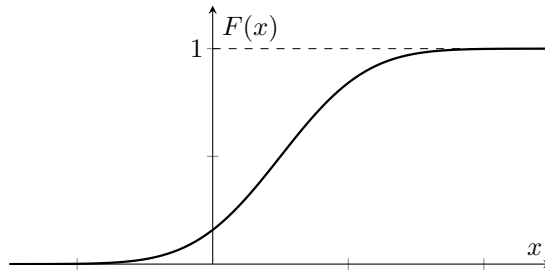


Figura 4.9: funzione di ripartizione gaussiana

**NB:** Per ammettere densità la funzione di ripartizione deve essere continua, essendo il risultato di un'integrazione.

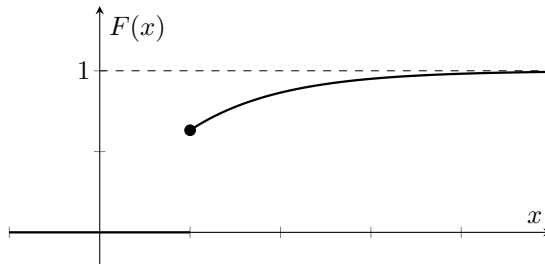


Figura 4.10: funzione di ripartizione che non ammette densità





---

## 5 Variabili aleatorie

Introduciamo ora uno strumento cardine dello studio della probabilità: la *variabile aleatoria*, nome che denota una quantità ignota a priori, il cui valore dipende dall'esito di un esperimento aleatorio. Essa associa gli esiti elementari su un dominio  $\Omega$ , anche non numerico, a valori reali su un codominio  $F$ , trasformando gli eventi sul dominio in eventi sul codominio. L'enorme utilità di questo strumento risiede proprio nel permettere l'applicazione della teoria su  $\mathbb{R}$  precedentemente introdotta, spostando l'attenzione da un dominio generico a un codominio numerico, molto più facile da analizzare. Agli eventi sul codominio, similmente a quanto visto per il dominio, è associata una probabilità, detta *legge* della variabile aleatoria, da cui discende la nozione di *uguaglianza quasi certa* tra variabili aleatorie. Sarà inoltre analizzato il concetto di *misurabilità*, proprio delle variabili aleatorie, e le relative proprietà.

### 5.1 Definizione

#### Definizione 5.1

Data una funzione  $X : \Omega \rightarrow F$  e un insieme  $B \subseteq F$ , si dice **controimmagine** di  $B$  l'insieme  $X^{-1}(B) = (X \in B) := \{\omega \in \Omega : X(\omega) \in B\} \subseteq \Omega$ .

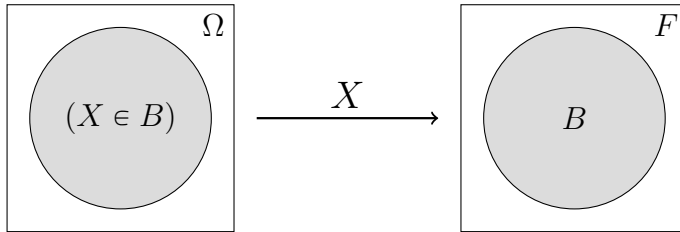


Figura 5.1: immagine e controimmagine di una VA

#### Definizione 5.2

Dati gli spazi misurabili  $(\Omega, \mathcal{A})$  e  $(F, \mathcal{F})$ , una funzione  $X : \Omega \rightarrow F$  si dice **funzione misurabile** o **variabile aleatoria** (abbreviato in **VA**) se:

$$(X \in B) \in \mathcal{A} \quad \forall B \in \mathcal{F}$$

Talvolta useremo la notazione sintetica  $X : (\Omega, \mathcal{A}, \mathbb{P}) \rightarrow (F, \mathcal{F})$  per indicare che  $X$  è una variabile aleatoria (o una funzione) che ha dominio  $\Omega$  e codominio  $F$ , che a  $\Omega$  è associato lo spazio di probabilità  $(\Omega, \mathcal{A}, \mathbb{P})$  e che a  $F$  è associato lo spazio misurabile  $(F, \mathcal{F})$ .

**Definizione 5.3**

L'insieme  $F$  che contiene tutti i possibili valori della VA si dice **spazio degli stati**.

*Proprietà della controimmagine:*

- $(X \in B^C) = X^{-1}(B^C) = (X^{-1}(B))^C = (X \in B)^C = (X \notin B)$
- $(X \in \bigcup_n B_n) = \bigcup_n (X \in B_n)$
- $(X \in \bigcap_n B_n) = \bigcap_n (X \in B_n)$

*Esempio.* Definiti  $\Omega = \{1, \dots, 6\}^2$ ,  $\mathcal{A} = 2^\Omega$ ,  $F = \mathbb{R}$ ,  $\mathcal{F} = \mathcal{B}$ ,  $X : \Omega \rightarrow \mathbb{R}$  e  $X(\omega) = \omega_1 + \omega_2$  la variabile aleatoria associa ad un evento nella  $\mathcal{F}$  (es. *la somma dei due dadi dà 3*) un evento nella  $\mathcal{A}$ .

In questo caso si consideri l'evento  $B = \{3\}$ : allora  $(X \in B) = \{(1, 2), (2, 1)\} \in \mathcal{A}$ . Inoltre  $\text{im}(X) = (2, \dots, 12)$ ,  $\#\text{im}(X) = 11 \leq \#\text{im}(\Omega) = 36$ , infatti ogni evento elementare può avere al massimo un esito.

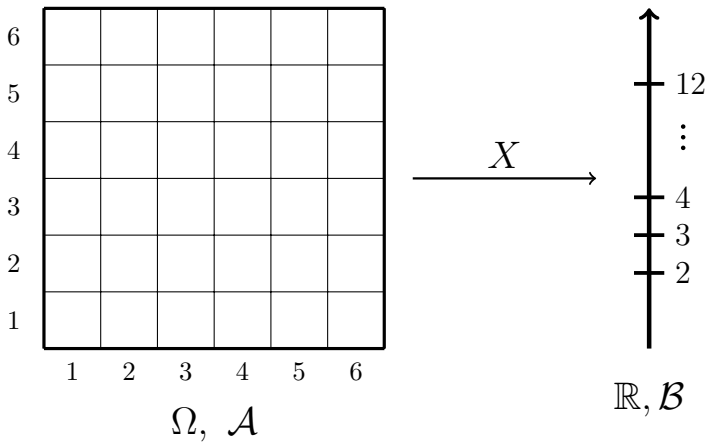
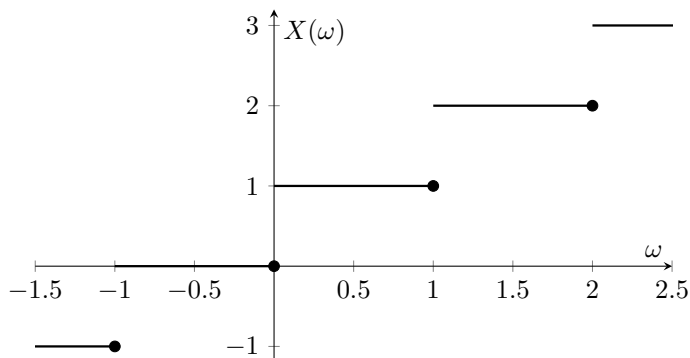


Figura 5.2: somma del lancio di due dadi

*Esempio.* Definita la funzione  $X : (\mathbb{R}, \mathcal{B}, \mathbb{P}) \rightarrow (\mathbb{R}, \mathcal{B})$ , con  $X(\omega) = [\omega]$  e  $\mathbb{P} = \mathcal{E}(\lambda)$ , si può affermare che  $\text{im}(X) = \mathbb{Z}$  e  $(X \in B) \in \mathcal{A} = \mathcal{B} \quad \forall B \in \mathcal{F}$ .

$X$  è effettivamente una VA poiché  $(X \in B)$  è un'unione al più numerabile di intervalli del tipo  $(k, k + 1]$  ( $k \in \mathbb{Z}$ ), cioè  $(X \in B) \in \mathcal{B}$ .

Figura 5.3: rappresentazione di  $X(\omega) = [\omega]$ 

*Esempio.* Definita la funzione  $X : (\mathbb{R}, \mathcal{B}, \mathbb{P}) \rightarrow (\mathbb{R}, \mathcal{B})$ , con  $X(\omega) = \omega$  e  $\mathbb{P} = \mathcal{E}(\lambda)$ , allora si può affermare che  $\text{im}(X) = \mathbb{R}$ . Le controimmagini coincidono con l'insieme di partenza, dunque  $X$  è una variabile aleatoria.

#### Teorema 5.4

Siano  $(\Omega, \mathcal{A})$  e  $(F, \mathcal{F})$  spazi misurabili generici,  $X : \Omega \rightarrow F$  una funzione, e  $\mathcal{C} \subseteq \mathcal{F}$  tale che  $\sigma(\mathcal{C}) = \mathcal{F}$ . Allora:

$$(X \in B) \in \mathcal{A} \quad \forall B \in \mathcal{F} \quad \iff \quad (X \in B) \in \mathcal{A} \quad \forall B \in \mathcal{C}$$

In altre parole, la misurabilità su un insieme coincide con la misurabilità su un insieme più piccolo che genera l'altro insieme.

#### Dimostrazione

( $\implies$ ): ovvia.

( $\impliedby$ ): ponendo  $\mathcal{D} = \{B \in \mathcal{F} : (X \in B) \in \mathcal{A}\}$  si ha che  $\mathcal{C} \subseteq \mathcal{D} \subseteq \mathcal{F}$  e che  $\mathcal{D}$  è una  $\sigma$ -algebra poiché unioni, intersezioni, e complementazioni commutano con le controimmagini, infatti:

$$\begin{aligned} B \in \mathcal{D} &\implies (X \in B) \in \mathcal{A} \implies (X \in B)^C \in \mathcal{A} \\ &\implies (X \in B^C) \in \mathcal{A} \implies B^C \in \mathcal{D} \end{aligned}$$

Dalla prima relazione ( $\mathcal{C} \subseteq \mathcal{D} \subseteq \mathcal{F}$ ) otteniamo  $\sigma(\mathcal{C}) \subseteq \sigma(\mathcal{D}) \subseteq \sigma(\mathcal{F})$ , da cui  $\mathcal{F} \subseteq \mathcal{D} \subseteq \mathcal{F}$ . Allora  $\mathcal{D} = \mathcal{F}$ , cioè tutti gli elementi  $B$  di  $\mathcal{F}$  sono tali per cui  $(X \in B) \in \mathcal{A}$ , da cui la tesi.  $\square$

## 5.2 Criteri di misurabilità

### Definizione 5.5

Siano  $(E, \mathcal{E}), (F, \mathcal{F})$  spazi misurabili.

Una funzione  $X : E \rightarrow F$  è detta **misurabile** se  $X^{-1}(B) \in \mathcal{E} \quad \forall B \in \mathcal{F}$ , oppure  $X^{-1}(\mathcal{F}) \subseteq \mathcal{E}$ .

**NB:** Quando  $(E, \mathcal{E}) = (\Omega, \mathcal{A})$ , la funzione misurabile  $X$  è una variabile aleatoria.

### Teorema 5.6 (J-P Th 8.1)

Siano  $(E, \mathcal{E}), (F, \mathcal{F})$  spazi misurabili,  $X : E \rightarrow F$ , e  $\mathcal{C} \in \mathcal{F}$  tale che  $\sigma(\mathcal{C}) = \mathcal{F}$ .

Allora:

$$X \text{ misurabile} \iff X^{-1}(\mathcal{C}) \subseteq \mathcal{E}$$

### Corollario 5.7

$(\Omega, \mathcal{A}), (\mathbb{R}, \mathcal{B})$  spazi misurabili,  $X : \Omega \rightarrow \mathbb{R}$ ,  $X_n : \Omega \rightarrow \mathbb{R}$ ,  $n \in \mathbb{N}$

1.  $X$  è VA misurabile:

$$(a) \iff (X \leq t) \in \mathcal{A} \quad \forall t \in \mathbb{R}$$

$$(b) \iff (X < t) \in \mathcal{A} \quad \forall t \in \mathbb{R}$$

2.  $X_n$  misurabile  $\forall n \implies$  sono VA:

$$(a) \sup_n X_n$$

$$(b) \inf_n X_n$$

$$(c) \limsup_n X_n$$

$$(d) \liminf_n X_n$$

3.  $X_n$  VA misurabile  $\forall n$ ,  $X_n \xrightarrow[n]{} X$  puntualmente  $\implies X$  VA misurabile.

Questo ci consente di lavorare con le variabili aleatorie comodamente su  $\mathbb{R}$  perché se  $X_n(\omega) \rightarrow X(\omega) \quad \forall \omega \in \mathbb{R}$  è sufficiente che il codominio sia  $\mathbb{R}$ , senza considerare il dominio  $\Omega$ .

### Dimostrazione

Viene dimostrato solo il corollario, non il teorema.

1. ( $\implies$ ) è ovvia per entrambe, dimostriamo ( $\impliedby$ ) per (a) e (b):

(a) Definiamo la semiretta  $(-\infty, t]$  come  $\mathcal{C}$ :

$$F = \mathbb{R}, \mathcal{F} = \mathcal{B}, \mathcal{C} = \{(-\infty, t], t \in \mathbb{R}\}, \sigma(\mathcal{C}) = \mathcal{B}$$

Quindi se  $(X \leq t) \in \mathcal{A} \quad \forall t \in \mathbb{R} \implies X$  è misurabile.

(b) Come sopra, definiamo la semiretta  $(-\infty, t)$  come  $\mathcal{C}_0$ :

$$F = \mathbb{R}, \mathcal{F} = \mathcal{B}, \mathcal{C}_0 = \{(-\infty, t), t \in \mathbb{R}\}$$

Possiamo affermare che  $\sigma(\mathcal{C}_0) = \mathcal{B}$  costruendo una serie  $(-\infty, t + \frac{1}{n})$ :

$$\bigcap_{n=1}^{+\infty} \left(-\infty, t + \frac{1}{n}\right) = (-\infty, t]$$

In questo modo si ha che  $\mathcal{C} \subseteq \mathcal{C}_0$  e di conseguenza:

$$\mathcal{B} = \sigma(\mathcal{C}) \subseteq \sigma(\mathcal{C}_0) \subseteq \mathcal{B}$$

2. (a) Essendo  $X$  misurabile  $\{X_n \leq t\} \in \mathcal{A}$ , quindi:

$$\left\{ \sup_n (X_n) \leq t \right\} = \left\{ \bigcap_n (X_n \leq t) \right\}$$

Essendo  $X_n$  VA, si sfrutta il punto (1), ovvero  $(X_n \leq t) \in \mathcal{A}$ . Intersecando si può affermare che:

$$\bigcap_n (X_n \leq t) \in \mathcal{A}$$

Servendosi di nuovo del punto (1) si mostra che  $\sup_n (X_n)$  è una VA.

(b) La dimostrazione di  $\inf_n X_n$  sarà affrontata più nel dettaglio:

Si può definire un  $\omega$  nel codominio per mostrare che è presente anche nel dominio:

$$\omega \in \{\inf_n (X_n < t)\} = \bigcup_n (X_n < t), (X_n < t) \in \mathcal{A}$$

Per dimostrare la prima uguaglianza si può passare per le esistenze:

$$\omega \in \{\inf_n (X_n \leq t)\} \iff \{\inf_n X_n\} \leq t$$

Essendo  $\inf_n X_n < t$ , allora:

$$\begin{aligned} \exists n_0 \text{ tale che } X_{n_0}(\omega) < t &\iff \exists n_0 \text{ tale che } \omega \in (X_{n_0} < t) \\ &\iff \omega \in \bigcup_n (X_n < t) \end{aligned}$$

Questo ragionamento è valido per ogni  $t$ .

- (c)  $\limsup_n X_n = \inf_n \left( \sup_{k>n} X_k \right)$ : sup è misurabile per (a) e inf per (b)
- (d)  $\liminf_n X_n = \sup_n \left( \inf_{k>n} X_k \right)$ : sup è misurabile per (a) e inf per (b)
3.  $\lim_n X_n = \limsup_n X_n$  se  $X_n$  converge puntualmente. Per la proprietà (2c) questo è misurabile.  $\square$

*Esempio:* controimmagine di una parabola. Dati  $\Omega = F = \mathbb{R}$ ,  $\mathcal{A} = \mathcal{F} = \mathcal{B}$ ,  $\mathbb{P} = \mathcal{E}(\lambda)$ ,  $X(\omega) = \omega^2$  possiamo affermare che  $X$  è misurabile<sup>10</sup> se e solo se  $(X \leq t) \in \mathcal{B} \quad \forall t$ , ovvero per una parabola del tipo:

$$(X \leq t) = \begin{cases} \emptyset & t < 0 \\ [-\sqrt{t}, \sqrt{t}] & t \geq 0 \end{cases} \in \mathcal{B} \quad \forall t$$

In entrambi i casi la controimmagine appartiene a  $\mathcal{B}$ :

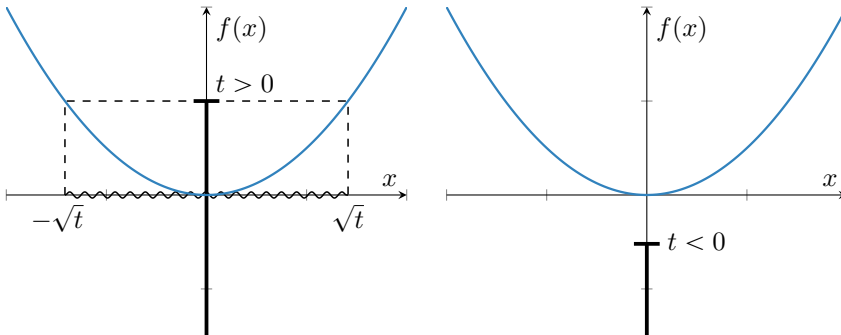


Figura 5.4: controimmagine di una parabola

### 5.3 Proprietà delle funzioni misurabili

**Teorema 5.8: composizione di funzioni misurabili** (*J-P Th 8.2*)  
 $(\Omega, \mathcal{A}), (F, \mathcal{F}), (G, \mathcal{G})$  spazi misurabili,  $X : \Omega \rightarrow F, h : F \rightarrow G$ . Allora:

$$X \text{ e } h \text{ misurabili} \implies Y = h \circ X = h(X) : \Omega \rightarrow G \text{ misurabile}$$

#### Dimostrazione

Si sceglie  $B$  nella  $\sigma$ -algebra finale  $\mathcal{G}$ , ovvero  $B \in \mathcal{G}$ . La controimmagine è:

$$(Y \in B) = Y^{-1}(B) = (h \circ X)^{-1}(B) = X^{-1}(h^{-1}(B))$$

<sup>10</sup>Spoilerone: dimostreremo che tutte le funzioni continue tra boreliani sono misurabili

Essendo  $h$  misurabile  $h^{-1}(B) \in \mathcal{F}$  ed essendo  $X$  misurabile  $X^{-1}(h^{-1}(B)) \in \mathcal{A}$ , che conclude la dimostrazione.  $\square$

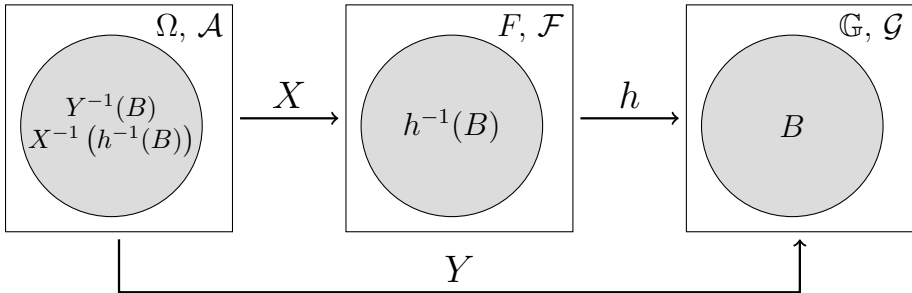


Figura 5.5: composizione di VA misurabili

**Definizione 5.9**

Per generare  $\mathcal{B}^n$  in  $\mathbb{R}^n$ , si definiscono gli **iper-quadranti di sud-ovest**:

$$\mathcal{B}^n = \sigma \left( \bigtimes_{k=1}^n (-\infty, t_k] : t_k \in \mathbb{Q} \right) = \sigma(\mathcal{C})$$

Questi non sono altro che l'equivalente multidimensionale delle semirette impiegate per definire  $\mathcal{B}$ . In due dimensioni, possono essere rappresentati come dei *quadranti mobili*.

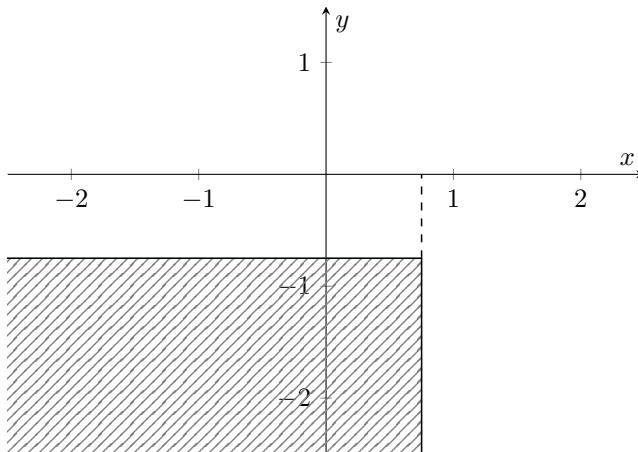


Figura 5.6: quadranti  $\tau_n$  generatori di  $\mathcal{B}^2$

**Definizione 5.10**

Se  $(F, \mathcal{F}) = (\mathbb{R}^n, \mathcal{B}^n)$  e  $(G, \mathcal{G}) = (\mathbb{R}^k, \mathcal{B}^k)$ , con  $\mathcal{B}^n = \sigma(\tau_n)$  e  $\mathcal{B}^k = \sigma(\tau_k)$ , allora una funzione misurabile  $h : \mathbb{R}^n \rightarrow \mathbb{R}^k$  è detta **funzione boreliana**.

**Proposizione 5.11** (*J-P Th 8.3*)

Se  $h : \mathbb{R}^n \rightarrow \mathbb{R}^k$  è continua, allora  $h$  è boreliana.<sup>11</sup>

**Dimostrazione**

Se  $B$  è aperto, allora  $B \in \mathcal{B}^k \implies h^{-1}(B)$  aperto: infatti per ogni funzione continua le controimmagini di aperti sono aperti:

$$\begin{aligned} h^{-1}(B) \in \tau_n \subseteq \sigma(\tau_n) = \mathcal{B}^n &\implies h^{-1}(B) \in \mathcal{B}^n \quad \forall B \in \tau_k \subseteq \mathcal{B}^k \\ &\implies h^{-1}(B) \in \mathcal{B}^n \quad \forall B \in \mathcal{B}^k \quad \square \end{aligned}$$

**5.3.1 Criteri di misurabilità per vettori aleatori**

**Teorema 5.12** (*J-P Th 8.4*)

Sia  $(\Omega, \mathcal{A})$  spazio misurabile, allora:

1.  $\mathbb{1}_A$  misurabile  $\iff A \in \mathcal{A}$  (legame biunivoco tra funzione e insieme);
2.  $X_k : \Omega \rightarrow \mathbb{R}$  misurabili  $\forall k = 1, \dots, n \iff X = (X_1, \dots, X_n) : \Omega \rightarrow \mathbb{R}^n$  misurabile (trasferibilità della misurabilità da e sulle componenti);
3.  $X_k : \Omega \rightarrow \mathbb{R}$  misurabili  $\forall k = 1, \dots, n$  e  $h : \mathbb{R}^n \rightarrow \mathbb{R}$  boreliana  $\implies h(X_1, \dots, X_n)$  è misurabile;
4.  $X : \Omega \rightarrow \mathbb{R}, Y : \Omega \rightarrow \mathbb{R}$  misurabili  $\implies X + Y, X \cdot Y, \frac{X}{Y}$  (se  $Y \neq 0$ ),  $\max(X, Y), \min(X, Y)$  sono misurabili.

**Dimostrazione**

1.  $(\implies)$ :  $\mathbb{1}_A$  misurabile  $\implies A = (\mathbb{1}_A = 1) \in \mathcal{A}$  perché  $B = \{1\} \in \mathcal{B}$ .

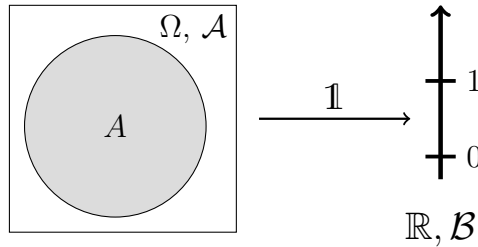
$(\impliedby)$ :

$$A \text{ misurabile} \implies \mathbb{1}_A \in \mathcal{B} = \begin{cases} \Omega & 0 \in B, 1 \in B \\ A^C & 0 \in B, 1 \notin B \\ A & 0 \notin B, 1 \in B \\ \emptyset & 0 \notin B, 1 \notin B \end{cases}$$

---

<sup>11</sup>Si vedrà in corsi di analisi più avanzati che questa proprietà regge anche per strumenti matematici più astratti e generali, non solo su spazi misurabili del tipo  $(\mathbb{R}^n, \mathcal{B}^n)$ . Più precisamente, dati  $(F, \mathcal{U})$  e  $(G, \mathcal{V})$  *spazi topologici*, cioè coppie composte da uno spazio astratto e da una collezione di suoi sottoinsiemi aperti (detta *topologia* dello spazio), allora  $F \rightarrow G$  continua  $\implies h$  misurabile.





Per qualsiasi  $A$  misurabile la controimmagine è sempre nella  $\sigma$ -algebra.

2. ( $\Leftarrow$ ): Sia  $B \in \tau$  (ovvero aperto):

$$\begin{aligned} (X_1 \in B) &= X_1^{-1}(B) = (X_1 \in B, X_2 \in \mathbb{R} \dots, X_n \in \mathbb{R}) \\ &= (X \in B \times \mathbb{R}^{n-1}) = X^{-1}(B \times \mathbb{R}^{n-1}) \in \mathcal{A} \end{aligned}$$

Essendo  $(B \times \mathbb{R}^{n-1})$  un aperto di  $\mathbb{R}^n$  contenuto in  $\mathcal{B}^n$  si può affermare che  $X$  è misurabile.

( $\Rightarrow$ ): Si impiegano gli iper-quadranti di sud ovest per trovare la controimmagine, individuando così tutto  $\mathbb{R}^n$ :

$$\begin{aligned} X^{-1} \left( \bigtimes_{k=1}^n (-\infty, t_k] \right) &= \bigcap_{k=1}^n X_k^{-1} ((-\infty, t_k]) \\ &= (X_1 \leq t_1 \dots, X_n \leq t_n) \in \mathcal{A}, \forall t_1 \dots, t_n \in \mathbb{R} \end{aligned}$$

Trovare la controimmagine in un iper-quadrante significa individuare un punto in nel rettangolo, quindi il vettore  $X$  è misurabile.

3.  $X_k$  misurabile  $\forall k = 1, \dots, n$ ,  $h$  misurabile, quindi  $h(X_1, \dots, X_n)$  è misurabile per il punto (2).
4.  $X, Y$  misurabili (e reali), si definisce una funzione  $h : \mathbb{R}^2 \rightarrow \mathbb{R}$  tale che  $h(X, Y) = X + Y$  e grazie al punto (3) si può affermare che è misurabile. Questo ragionamento è estendibile a tutte le operazioni elencate, e con un *atto di fede* anche al rapporto, a patto che  $Y \neq 0$ .  $\square$

*Esempio.* Dati  $\Omega = \mathbb{R}, \mathcal{A} = \mathcal{B}, \mathbb{P} = \mathcal{E}, F = \mathbb{R}^2, \mathcal{F} = \mathcal{B}^2$  si può affermare che:

$$X : \mathbb{R} \rightarrow \mathbb{R}^2, X(\omega) = \begin{bmatrix} |\omega| \\ \omega^2 \end{bmatrix} \implies X \text{ misurabile}$$

*Esempio.* Dati  $\Omega = \mathbb{R}, \mathcal{A} = \mathcal{B}, \mathbb{P} = \mathcal{E}, F = \mathbb{R}^2, \mathcal{F} = \mathcal{B}, X : \mathbb{R} \rightarrow \mathbb{R}$  così definita:

$$X(\omega) = \omega^2 + \mathbf{1}_{[-1,1]}(\omega) + \frac{\log(1 + |\omega|)}{1 + \omega^2}$$

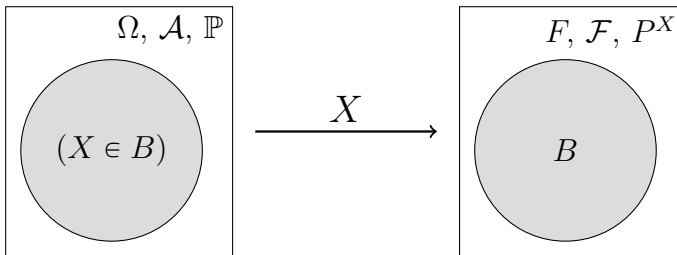
Allora  $X$  è misurabile in quanto composizione di misurabili.

### 5.4 Legge di una variabile aleatoria

**Definizione 5.13**

Dati  $(\Omega, \mathcal{A}, \mathbb{P})$  spazio di probabilità,  $(F, \mathcal{F})$  spazio misurabile e  $X : \Omega \rightarrow F$  VA, è detta **legge** o **distribuzione** di  $X$  la probabilità:

$$P^X : \mathcal{F} \rightarrow [0, 1] \quad P^X(B) = \mathbb{P}(X \in B)$$



In generale,  $\mathbb{P}$  è in funzione di un elemento della  $\sigma$ -algebra, oggetto che può essere spesso astratto. Per lo sviluppo della teoria, eviteremo di studiare direttamente le  $\sigma$ -algebre, ma lavoreremo con le variabili aleatorie. Questo motiva la definizione di una funzione  $P^X$ , che riceve direttamente in ingresso un elemento del codominio  $F$  della variabile aleatoria.

**Teorema 5.14** (*J-P Th 8.5*)

$P^X$  è una probabilità.

**Dimostrazione**

Per dimostrare che  $P^X$  sia una probabilità dobbiamo verificare le due proprietà fondamentali delle probabilità:

1.  $P^X(F) = \mathbb{P}(X \in F) = \mathbb{P}(\Omega) = 1$
2. La  $\sigma$ -additività.

Dati  $B_n \in \mathcal{F}$   $n \in \mathbb{N}$ ,  $B_l \cap B_k = \emptyset \quad \forall k \neq l$  disgiunti, allora per le proprietà delle controimmagini:

$$\begin{aligned} P^X \left( \bigcup_{n=1}^{+\infty} B_n \right) &= \mathbb{P} \left( X \in \bigcup_{n=1}^{+\infty} B_n \right) = \mathbb{P} \left( \bigcup_{n=1}^{+\infty} (X \in B_n) \right) \\ &= \sum_{n=1}^{+\infty} \mathbb{P}(x \in B_n) = \sum_{n=1}^{+\infty} P^X(B_n) \end{aligned}$$

Si presti attenzione al fatto che il ragionamento è valido solo se gli  $A_n$  sono disgiunti. Questo è garantito dall'incompatibilità dei  $B_n$ : se un punto è in  $A_l \cap A_k$  allora dovrebbe essere anche in  $B_l \cap B_k$ , ma questo non è possibile in quanto i  $B_n$  sono disgiunti.  $\square$

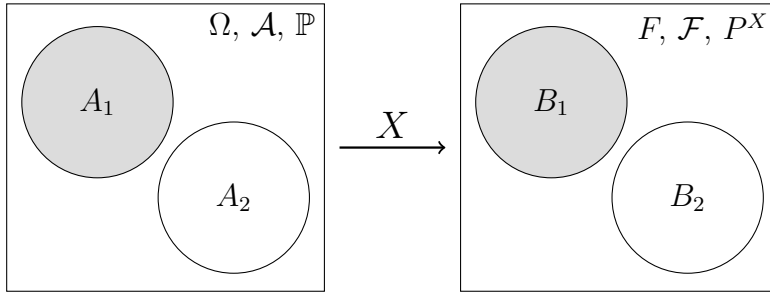


Figura 5.7:  $A_n$  disgiunti comportano  $B_n$  disgiunti, e viceversa.

**Osservazione 5.15**

Se  $(F, \mathcal{F}) = (\mathbb{R}, \mathcal{B})$  allora  $P^X$  è caratterizzata dalla sua funzione di ripartizione:

$$F_X : \mathbb{R} \rightarrow [0, 1], \quad F_X(t) = P^X((-\infty, t]) = \mathbb{P}(X \leq t)$$

1. Se  $F_X$  ammette densità  $f_X$ , allora  $f_X$  caratterizza  $F_X$  e quindi la legge  $P^X$ ;
2. Se  $P^X$  probabilità discreta su  $\{x_k\}_{k \in \mathbb{N}}$  con densità discreta  $\{p_k\}_{k \in \mathbb{N}}$  allora  $(\{x_k\}_k; \{p_k\}_k)$  caratterizza  $P^X$ .

Questa è la base della definizione delle VA su  $\mathbb{R}$ .

**5.4.1 Esempi di leggi**

1. Lancio di due dadi:  $\Omega = \{1, \dots, 6\}^2$ ,  $\mathcal{A} = 2^\Omega$ ,  $\mathbb{P}$  uniforme,  $X : \Omega \rightarrow \mathbb{R}$ ,  $X(\omega) = \omega_1 + \omega_2$ .  $P^X$  è una probabilità discreta:

$\Omega$	2	3	4	5	6	7	8	9	10	11	12
Densità $p_i$	$\frac{1}{36}$	$\frac{1}{18}$	$\frac{1}{12}$	$\frac{1}{9}$	$\frac{5}{36}$	$\frac{1}{6}$	$\frac{5}{36}$	$\frac{1}{9}$	$\frac{1}{12}$	$\frac{1}{18}$	$\frac{1}{36}$

È importante controllare sempre che la somma delle densità  $p_i$  sia 1.

2. Dati  $\Omega = \mathbb{R}$ ,  $\mathcal{A} = \mathcal{B}$ ,  $\mathbb{P} = \mathcal{E}(\lambda)$ ,  $F = \mathbb{R}$ ,  $\mathcal{F} = \mathcal{B}$ ,  $X : \mathbb{R} \rightarrow \mathbb{R}$ ,  $X(\omega) = [\omega]$  allora  $\text{im}(X) = \mathbb{Z}$ .

In tal caso  $P^X$  è una probabilità discreta su  $\mathbb{Z}$  con densità discreta  $\{p_k\}_{k \in \mathbb{Z}}$ .

$$\begin{aligned}
 p_k = P^X(\{k\}) &= \mathbb{P}(X = k) = \mathbb{P}((k-1, k]) = \int_{k-1}^k f(t) dt = \\
 &= \int_{k-1}^k \lambda e^{-\lambda t} \mathbb{1}_{[0, +\infty)}(t) dt = \begin{cases} 0 & k \leq 0 \\ e^{-\lambda(k-1)} - e^{-\lambda k} & k \in \mathbb{N}^+ \end{cases}
 \end{aligned}$$

ovvero  $X \sim G(1 - e^{-\lambda})$ : densità geometrica

3. Dati  $\Omega = \mathbb{R}$ ,  $\mathcal{A} = \mathcal{B}$ ,  $\mathbb{P} = \mathcal{E}(\lambda)$ ,  $F = \mathbb{R}$ ,  $\mathcal{F} = \mathcal{B}$ , allora

$$X(\omega) = \omega \implies P^X = \mathbb{P} = \mathcal{E}(\lambda)$$

Attenzione: stiamo solo definendo una singola VA, non sono ancora validi i ragionamenti che ne considerano una moltitudine.

4. Dati  $\Omega = \mathbb{R}$ ,  $\mathcal{A} = \mathcal{B}$ ,  $\mathbb{P} = \mathcal{E}(\lambda)$ ,  $F = \mathbb{R}$ ,  $\mathcal{F} = \mathcal{B}$  e  $X(\omega) = \omega^2$  possiamo scoprire  $P^X$  tramite la  $F$ :

$$F_X(t) = \mathbb{P}(X \leq t) = \dots = \left(1 - e^{-\lambda\sqrt{t}}\right) \mathbb{1}_{[0, +\infty)}(t)$$

Questa funzione è  $C^0 \cap C^1$  (a tratti), quindi possiamo anche definire  $f_X$ :

$$f_X(t) = F'_X(t) = \frac{\lambda e^{-\lambda\sqrt{t}}}{2\sqrt{t}} \mathbb{1}_{[0, +\infty)}(t)$$

## 5.5 Uguaglianza quasi certa

### Definizione 5.16

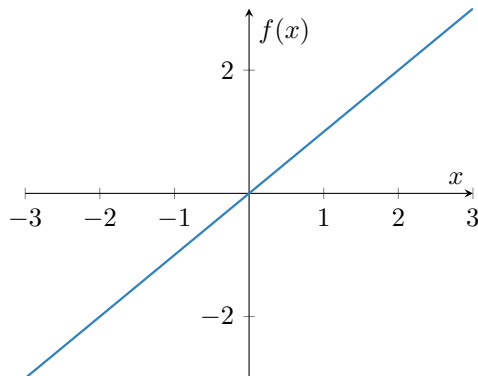
Siano  $(\Omega, \mathcal{A}, \mathbb{P})$  uno spazio di probabilità,  $(F, \mathcal{F})$  uno spazio misurabile e  $X, Y : \Omega \rightarrow F$  variabili aleatorie.

Si dice che  $X = Y$  **quasi certamente** ( $X \stackrel{\text{qc}}{=} Y$ ) se  $\mathbb{P}(X = Y) = 1$ .

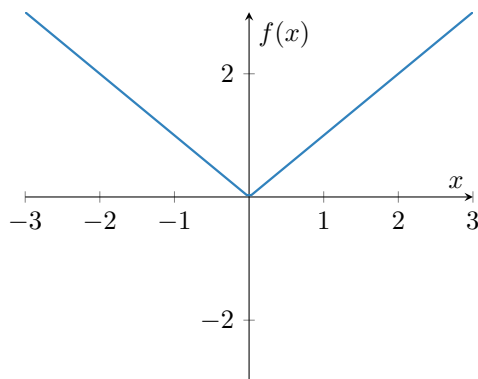
**NB:** Si noti che  $(X = Y) = (X - Y = 0) \in \mathcal{A}$ . Le due VA sono cioè modellisticamente indistinguibili:  $X$  e  $Y$  sono la descrizione del medesimo fenomeno aleatorio.

*Esempio.* Dati  $\Omega = \mathbb{R}$ ,  $\mathcal{A} = \mathcal{B}$ ,  $\mathbb{P} = \mathcal{E}(\lambda)$ ,  $F = \mathbb{R}$ ,  $\mathcal{F} = \mathcal{B}$  definiamo le seguenti VA:

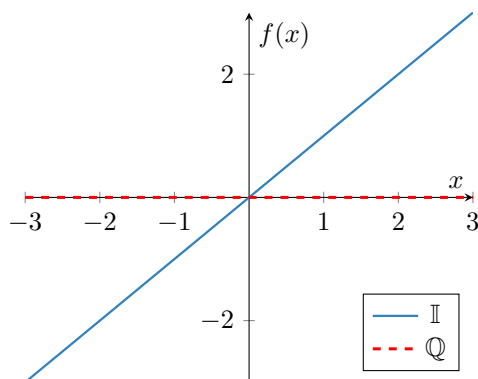
- $X_1(\omega) = \omega$



- $X_2(\omega) = |\omega|$



- $X_3(\omega) = \begin{cases} \omega & \omega \in \mathbb{R} \setminus \mathbb{Q} \\ 0 & \omega \in \mathbb{Q} \end{cases}$



Mostriamo che queste tre VA sono quasi certamente uguali.

- La prima e la seconda sono equivalenti:

$$(X_1 = X_2) = [0, +\infty) \quad (\text{exp definito solo per } x \geq 0)$$

$$\mathbb{P}(X_1 = X_2) = \mathbb{P}([0, +\infty)) = \int_0^{+\infty} \lambda e^{-\lambda t} dt = 1$$

- La prima e la terza sono equivalenti:

$$(X_1 = X_3) = \mathbb{R} \setminus \mathbb{Q} \cup \{0\}$$

$$\begin{aligned} \mathbb{P}(X_1 = X_3) &= \mathbb{P}(\mathbb{R} \setminus \mathbb{Q} \cup \{0\}) \\ &= \mathbb{P}(\mathbb{R} \setminus \mathbb{Q}) + \mathbb{P}(\{0\}) && \text{(eventi disgiunti)} \\ &= \mathbb{P}(\mathbb{R}) - \mathbb{P}(\mathbb{Q}) && \text{(la prob. di un singolo punto è 0)} \\ &= \mathbb{P}(\mathbb{R}) \end{aligned}$$

Nell'ultimo passaggio, come nel precedente, si usa il fatto che la probabilità di un'infinità numerabile di punti sia nulla.

**Proposizione 5.17**

Siano  $(\Omega, \mathcal{A}, \mathbb{P})$  spazio di probabilità,  $(F, \mathcal{F})$  spazio misurabile,  $X, Y : \Omega \rightarrow F$  VA. Allora:

$$X \stackrel{\text{qc}}{=} Y \implies P^X = P^Y$$

**Dimostrazione**

Sia  $B \in \mathcal{F}$ . Dunque  $P^X(B) = \mathbb{P}(X \in B) \stackrel{\text{qc}}{=} \mathbb{P}(X \in B, X = Y)$ , poiché intersecando un evento con un altro quasi certo, la probabilità non si modifica. Allora:

$$P^X(B) = \mathbb{P}(Y \in B, X = Y) \stackrel{\text{qc}}{=} \mathbb{P}(Y \in B) = P^Y(B) \quad \forall B \quad \square$$

**NB:** Questa dimostrazione non prova che date due VA iid<sup>12</sup> (es. lancio di due dadi) queste hanno lo stesso risultato. Il simbolo  $\stackrel{\text{qc}}{=}$  indica una situazione del tipo *tiro un dado e copio due volte lo stesso risultato*.

**Definizione 5.18**

Una proprietà vale **quasi certamente** se:

$$\exists A \in \mathcal{A} \text{ tale che } \begin{cases} \mathbb{P}(A) = 1 \\ \text{la proprietà vale } \forall \omega \in A \end{cases}$$

---

<sup>12</sup>Indipendenti e Identicamente Distribuite: le variabili hanno tutte la stessa distribuzione di probabilità e sono tutte statisticamente indipendenti.

---

## 6 Integrazione rispetto alla probabilità

Come osservato nei capitoli precedenti, una variabile aleatoria può assumere diversi valori con determinate probabilità. Il *valore atteso* è, in buona sostanza, la media di tutti i valori che una variabile aleatoria reale può assumere, ciascuno pesato per la sua probabilità di comparire. A partire da questo capitolo costruiremo la *regola del valore atteso* cominciando da casi elementari, le variabili *semplici*, per arrivare a casi via via più complessi e generici, ma sempre tenendo come riferimento il concetto di valore atteso sopra illustrato. Saranno inoltre sviluppate le utilissime proprietà del valore atteso, che assumeranno crescente importanza nel proseguimento del corso, e di un altro indicatore strettamente collegato al valore atteso, la *varianza*, che descrive l'ampiezza della distribuzione intorno alla media.

### Definizione 6.1

Una variabile aleatoria a valori reali  $X : \Omega \rightarrow \mathbb{R}$  si dice **variabile aleatoria reale** (d'ora in avanti **VAR**).

Ovviamente solo una variabile aleatoria a valori numerici (reali nel caso più generale) può ammettere una media sui valori del codominio.

### Definizione 6.2

Siano dati uno spazio di probabilità  $(\Omega, \mathcal{A}, \mathbb{P})$ , una VA  $X : \Omega \rightarrow \mathbb{R}$  e una legge  $P^X$  su  $\mathcal{B}$ . Si definisca  $\mathbb{E}[X]$  (talvolta scritto  $\mu$ ), **valore atteso** di  $X$ , o **integrale** di  $X$ , come il valore centrale rispetto ai possibili esiti  $X(\omega)$  di  $X$ , ovvero una media pesata dei possibili valori di  $X(\omega) \forall \omega \in \Omega$ . Per il valore atteso saranno utilizzate le seguenti scritte equivalenti:

$$\mathbb{E}[X] = \int_{\Omega} X(\omega) \mathbb{P}(d\omega) = \int_{\Omega} X(\omega) d\mathbb{P}(\omega) = \int_{\Omega} X d\mathbb{P}$$

## 6.1 Variabili aleatorie semplici

### Definizione 6.3

Una VAR  $X$  è detta **semplice** se può assumere solo un numero finito di valori, ovvero se  $\#\text{im}(X) = n < +\infty$  (con  $n \in \mathbb{N}$ ).

### Proposizione 6.4: criterio per VAR semplici

$X : \Omega \rightarrow \mathbb{R}$  è una VAR semplice se e solo se:

$$X = \sum_{k=1}^n x_k \mathbb{1}_{A_k}, \quad \text{con } A_k = (X = x_k) \in \mathcal{A} \quad \forall k = 1, \dots, n$$

Il criterio dà una formula di “scomposizione” universale per tutte le VA semplici. Come è possibile evincere dalla figura 6.1, gli  $A_k$  sono disgiunti (ovvero tali che

$A_k \cap A_l = \emptyset \forall k \neq l$ ) e coprono l'intero dominio  $\left(\bigcup_{k=1}^n A_k = \Omega\right)$ , ovvero la collezione  $\{A_k\}$  è una partizione<sup>13</sup>. Si ha dunque che:

$$X(\omega) = \sum_{k=1}^n x_k \mathbb{1}_{A_k}(\omega) = \begin{cases} x_1 & \omega \in A_1 \\ x_2 & \omega \in A_2 \\ \dots & \dots \\ x_n & \omega \in A_n \end{cases}$$

Questa è una “*finta sommatoria*”, nel senso che per ogni  $\omega$  una sola delle indicatrici vale 1, e quindi un solo elemento della sommatoria alla volta è non nullo. In questo caso, inoltre, tutte le controimmagini sono misurabili.

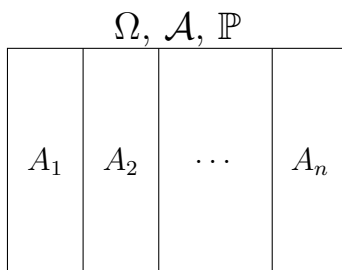


Figura 6.1: dominio di una VAR semplice

**Dimostrazione**

( $\implies$ ):  $X = \sum_{k=1}^n x_k \mathbb{1}_{(X=x_k)}$  perché  $\text{im}(X) = \{x_1, \dots, x_n\}$ .

( $\impliedby$ ):

1.  $X$  è reale perché  $x_k \in \mathbb{R} \forall k$ ;
2.  $X$  è semplice perché  $\#\text{im}(X) = n \in \mathbb{N}$ ;
3.  $X$  è misurabile perché le indicatrici sono misurabili e il prodotto di quantità misurabili è misurabile. □

**NB:** Sia  $\{\tilde{A}_i\}$  una collezione di cardinalità  $n$  di insiemi generici appartenenti ad  $\mathcal{A}$  e siano  $a_i \in \mathbb{R} \forall i$ . A differenza dell'esempio precedente, non chiediamo che gli  $\tilde{A}_i$  siano disgiunti né che  $\bigcup_i \tilde{A}_i = \Omega$ .

---

<sup>13</sup>In realtà nella definizione di partizione è richiesto che tutti gli insiemi devono avere probabilità non nulla, mentre qui non necessariamente  $\mathbb{P}(A_k) > 0 \forall k$ ; tuttavia questa precisazione non ha particolare importanza nel caso in esame.



Definiamo ora  $X = \sum_{i=1}^v a_i \mathbb{1}_{\tilde{A}_i}$ , dove  $v < +\infty$  è il numero (arbitrario e non necessariamente uguale a  $n$ ) di elementi sommati. Allora  $X$  è una VAR semplice, perché  $\#\text{im}(X) < +\infty$ : essendo  $v$  un numero finito, la cardinalità dell'immagine non può essere infinita.

*Esempio.* Presa la VAR sopra definita, e dato un dominio di cardinalità  $v = 2$ , i possibili valori della VAR sono 4. In generale, l'immagine ha cardinalità  $2^v < +\infty$  casi.<sup>14</sup>

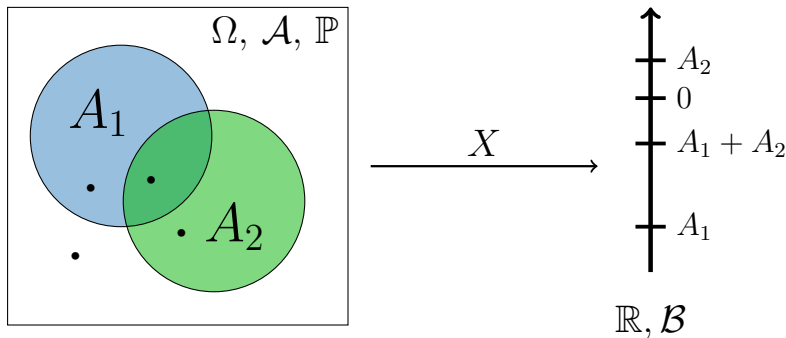


Figura 6.2: VAR semplice con cardinalità 2

## 6.2 Integrazione di variabili aleatorie reali

### 6.2.1 Variabili semplici

#### Definizione 6.5

Data  $X = \sum_{k=1}^n x_k \mathbb{1}_{(X=x_k)}$  VAR semplice definita sul dominio  $\Omega$ , si dice **valore atteso** di  $X$  la seguente quantità:

$$\mathbb{E}[X] = \int_{\Omega} X \, d\mathbb{P} := \sum_{k=1}^n x_k \mathbb{P}(X = x_k)$$

Nel caso delle VA semplici, l'integrale è formato solo da un numero finito di casi, quindi si trasforma in una sommatoria. Si osservi la figura 6.3. Ogni evento  $A_k$  ha la sua probabilità di verificarsi, rappresentata dalle aree dei rispettivi rettangoli. La VA  $X$  assegna un valore reale ad ognuno di questi eventi. Il valore atteso di  $X$  è semplicemente la media di questi valori, pesati rispetto alle probabilità dei corrispondenti eventi.

<sup>14</sup>Questo è vero in quanto “esistono due casi per ogni insieme: appartenenza e non appartenenza” -Capitan Findus

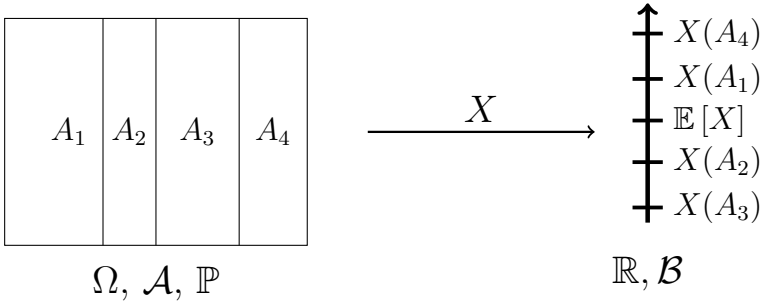


Figura 6.3: valore atteso per una VAR semplice

**Proposizione 6.6**

Siano  $X, Y$  VAR semplici su  $(\Omega, \mathcal{A}, \mathbb{P})$  e sia  $a \in \mathbb{R}$ . Allora:

1.  $aX$  è una VAR semplice
2.  $\mathbb{E}[aX] = a\mathbb{E}[X]$
3.  $X = \sum_{i=1}^v a_i \mathbb{1}_{\tilde{A}_i}, \tilde{A}_i \in \mathcal{A} \forall i \implies \mathbb{E}[X] = \sum_{i=1}^v a_i \mathbb{P}(\tilde{A}_i)$
4.  $X + Y$  è una VAR semplice
5.  $\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$
6.  $X \leq Y \forall \omega \in \Omega \implies \mathbb{E}[X] \leq \mathbb{E}[Y]$

Le proprietà (2) e (5) dimostrano la linearità del valore atteso, e la (6) ne dimostra la positività.

**Dimostrazione**

1. È immediata, definendo la nuova VA  $Y = aX = \sum_{k=1}^n (aX_k) \mathbb{1}_{A_k}$ .

Si noti che  $A_k = (X = x_k) = (Y = ax_k) \forall a \neq 0$ .

$\Omega$  in  $\mathbb{P}(\Omega)$  è l'unica controimmagine possibile, infatti tutti gli elementi  $\omega$  sono mappati in zero perché è l'unico valore che  $Y$  può assumere.

2. Nel caso di  $a = 0$  allora  $Y = 0 \cdot X = 0$ , quindi  $\mathbb{E}[Y] = 0 \cdot \mathbb{P}(\Omega) = 0 = 0 \cdot \mathbb{E}[X]$ . Nel caso generale di  $a \neq 0$  si ha invece che:

$$\mathbb{E}[Y] = \mathbb{E}[aX] = \sum_{k=1}^n ax_k \mathbb{P}(aX = ax_k) = a \sum_{k=1}^n x_k \mathbb{P}(x = x_k) = a\mathbb{E}[X]$$

La tesi è dunque valida  $\forall a \in \mathbb{R}$ .

3. Si ponga  $v = 2$  per semplicità di notazione (la dimostrazione del caso generale si svolge in modo simile).

È possibile definire esplicitamente  $X$  VAR semplice bidimensionale nel seguente modo:

$$X = \sum_{i=1}^2 a_i \mathbb{1}_{\tilde{A}_i} = a_1 \mathbb{1}_{\tilde{A}_1 \setminus \tilde{A}_2} + a_2 \mathbb{1}_{\tilde{A}_2 \setminus \tilde{A}_1} + (a_1 + a_2) \mathbb{1}_{\tilde{A}_1 \cap \tilde{A}_2} + 0 \cdot \mathbb{1}_{(\tilde{A}_1 \cup \tilde{A}_2)^c}$$

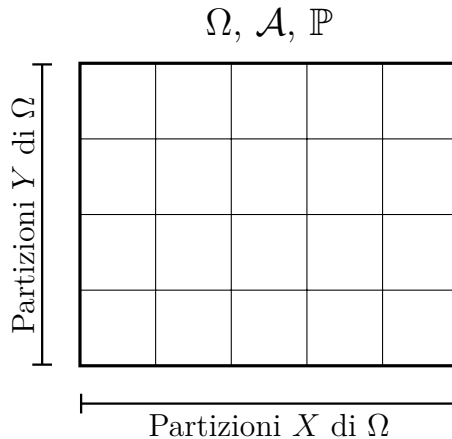
Infatti i quattro valori combinati linearmente sono gli unici valori che  $X$  può assumere, e i quattro insiemi delle indicatrici formano una partizione per  $\Omega$ . Si delineano insomma quattro casi, come illustrato nell'esempio precedente con la figura 6.2:

$$\begin{aligned} \mathbb{E}[X] &= a_1 \mathbb{P}(\tilde{A}_1 \setminus \tilde{A}_2) + a_2 \mathbb{P}(\tilde{A}_2 \setminus \tilde{A}_1) + (a_1 + a_2) \mathbb{P}(\tilde{A}_1, \tilde{A}_2) \\ &= a_1 (\mathbb{P}(\tilde{A}_1 \setminus \tilde{A}_2) + \mathbb{P}(\tilde{A}_1, \tilde{A}_2)) + a_2 (\mathbb{P}(\tilde{A}_2 \setminus \tilde{A}_1) + \mathbb{P}(\tilde{A}_1, \tilde{A}_2)) \\ &= a_1 \mathbb{P}(\tilde{A}_1) + a_2 \mathbb{P}(\tilde{A}_2) \end{aligned}$$

4. Essendo  $X$  e  $Y$  VAR, la loro somma è misurabile e semplice (vedi teorema 5.12).
5. Scomponendo  $X$  e  $Y$  come “finte sommatorie” si ottiene:

$$\mathbb{E}[X + Y] = \mathbb{E} \left[ \sum_{k=1}^n x_k \mathbb{1}_{(X=x_k)} + \sum_{l=1}^m y_l \mathbb{1}_{(Y=y_l)} \right]$$

Partizioniamo ora  $\Omega$  regolarmente lungo  $X$  e  $Y$ :



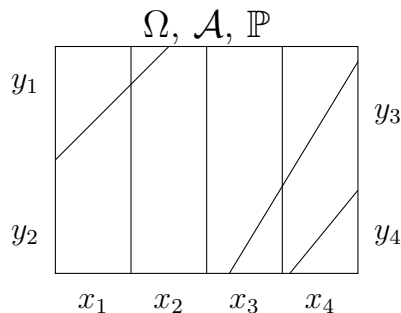
Si verifica facilmente che il risultato ottenuto ha la seguente formulazione equivalente:

$$\begin{aligned}
 \mathbb{E}[X + Y] &= \mathbb{E} \left[ \sum_{k=1}^n \sum_{l=1}^m (x_k + y_l) \mathbb{1}_{(X=x_k, Y=y_l)} \right] = \\
 &\quad \text{(applicando il punto (3))} \\
 &= \sum_{k=1}^n \sum_{l=1}^m (x_k + y_l) \mathbb{P}(X = x_k, Y = y_l) = \\
 &= \sum_{k=1}^n \left[ x_k \sum_{l=1}^m \mathbb{P}(X = x_k, Y = y_l) \right] + \sum_{l=1}^m \left[ y_l \sum_{k=1}^n \mathbb{P}(X = x_k, Y = y_l) \right] = \\
 &\quad \text{(per le probabilità totali)} \\
 &= \sum_{k=1}^n [x_k \mathbb{P}(X = x_k)] + \sum_{l=1}^m [y_l \mathbb{P}(Y = y_l)] = \\
 &= \mathbb{E}[X] + \mathbb{E}[Y]
 \end{aligned}$$

6. Siano  $X$  e  $Y$  due VAR semplici così espresse:

$$X = \sum_{k=1}^n x_k \mathbb{1}_{(X=x_k)}, \quad Y = \sum_{l=1}^m y_l \mathbb{1}_{(Y=y_l)}$$

Si noti che non è detto che  $x_k \leq y_l \forall k, l$ , perché i valori  $x_k$  e  $y_l$  non sono necessariamente definiti sulla stessa parte del dominio. Ad esempio nel caso raffigurato la partizione che ha immagine  $x_3$  non deve essere minore di tutte le altre, bensì è sufficiente che valga  $x_3 \leq y_2, y_3$ . Questo è simile a quanto succede per una normale coppia di funzioni: il fatto che una funzione sia maggiore di un'altra non vuol dire necessariamente che, preso un qualsiasi valore della prima, esso sarà maggiore di *tutti* i valori della seconda; il confronto avviene solo sulle coppie di valori che sono mappati dalla stessa parte di dominio.



Bisogna quindi creare una *partizione comune*  $\{\tilde{A}_j\}$  per  $X$  e  $Y$  composta da tutte le intersezioni delle due partizioni precedenti. In altre parole, prendiamo la collezione  $\{\tilde{A}_j\}$  tale che:

- $\tilde{A}_j \in \mathcal{A} \forall j$
- $\bigcup_j \tilde{A}_j = \Omega$
- $\tilde{A}_j \cap \tilde{A}_i = \emptyset \forall j \neq i$ .

Ora si può dunque ripartizionare le VA  $X$  e  $Y$  sui nuovi  $\tilde{A}$ :

$$X = \sum_j a_j \mathbb{1}_{\tilde{A}_j}, \quad Y = \sum_j b_j \mathbb{1}_{\tilde{A}_j}$$

Poiché con l'introduzione della partizione comune gli insiemi delle indicatrici sono gli stessi per entrambe le VA, è finalmente possibile confrontare i valori delle VA su ogni pezzo di dominio. In particolare, possiamo affermare che  $a_j \leq b_j \forall j$ . Passando ai valori attesi:

$$\mathbb{E}[X] = \sum_j a_j \mathbb{P}(\tilde{A}_j) \leq \sum_j b_j \mathbb{P}(\tilde{A}_j) = \mathbb{E}[Y] \quad \square$$

### 6.2.2 Variabili positive

Si vuole ora estendere il concetto di valore atteso dalle VA semplici a quelle che possono assumere ogni possibile valore su  $\mathbb{R}$ , o comunque su insiemi di cardinalità infinita.

Partiamo dalla trattazione per VA a valori positivi<sup>15</sup> sul codominio.

#### Definizione 6.7

Data  $X : \Omega \rightarrow [0, +\infty]$ <sup>16</sup> VA positiva, si definisce **valore atteso** la seguente quantità:

$$\mathbb{E}[X] = \int_{\Omega} X d\mathbb{P} := \sup \left\{ \mathbb{E}[Y] : Y \text{ VAR semplice e } 0 \leq Y \leq X \right\}$$

Si può dire che  $Y$  approssima  $X$  “da sotto”: rimpicciolendo gli intervalli su cui sono definiti i valori di  $Y$ , si rende l'approssimazione sempre più precisa. Vedremo come costruire una successione di VAR semplici che approssima una VA generica a pagina 74.

Si noti che anche in questo caso il valore atteso è *lineare*, grazie alla linearità dello stesso su VA semplici (dimostrata nella proposizione 6.6).

<sup>15</sup>La nomenclatura tecnicamente corretta sarebbe “non negativi”, in quanto lo zero può essere incluso nel codominio della VA, ma in tutta la letteratura di probabilità si usa il termine “positivo” per immediatezza di scrittura.

<sup>16</sup>Si noti che in questa definizione è consentito che  $X(\omega) = +\infty$  per qualche  $\omega$ . Opereremo non di rado con VA infinite, o dal valore atteso infinito (e quindi non strettamente VA “reali”).

**NB:** Notazioni alternative per indicare che  $X$  è una VA positiva sono  $X(\omega) \geq 0 \forall \omega \in \Omega$  o, più sinteticamente,  $X \geq 0$ .

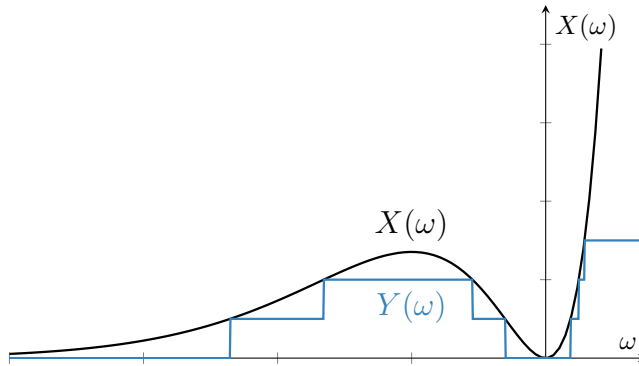


Figura 6.4:  $X$  VA e  $Y$  VAR semplice, approssimazione di  $X$

**Proposizione 6.8**

Data una VAR positiva  $X$ :

1.  $\mathbb{E}[X]$  esiste sempre, ovvero si dice che è *ben definito*;
2.  $\mathbb{E}[X] \in [0, +\infty]$  (non negativo e, al massimo, infinito);
3.  $\mathbb{E}[X]$  coincide con la definizione precedente (la 6.5) per VAR semplici se  $X$  è semplice, ovvero, questa definizione è un'estensione della precedente;
4. Se  $\mathbb{P}(X = +\infty) > 0$ , allora  $\mathbb{E}[X] = +\infty$ .

L'implicazione inversa della proprietà (4) non è vera: non è detto che se la media sia infinita, la probabilità che la VAR assuma valore infinito sia non nulla.

*Controesempio.* Sia  $X : (\Omega, \mathcal{A}, \mathbb{P}) \rightarrow (\mathbb{R}, \mathcal{B})$  con:

$$\Omega = \mathbb{N} \cup \{0\}, \mathcal{A} = 2^\Omega \quad \text{e} \quad \mathbb{P}(\{\omega\}) = e^{-\lambda} \frac{\lambda^\omega}{\omega!}, \text{ con } \lambda > 1$$

Ovvero una distribuzione di Poisson. Per definizione  $X(\omega) = \omega! < +\infty \forall \omega \in \Omega$ . Dunque non esiste  $\omega$  per cui  $X$  sia infinita, quindi l'evento  $X = +\infty$  ha controimmagine vuota, ovvero la sua probabilità è nulla:  $\mathbb{P}(X = +\infty) = \mathbb{P}(\emptyset) = 0$ .

Per calcolare il valore atteso è necessario definire la variabile ausiliaria  $Y_n$ :

$$Y_n(\omega) := \begin{cases} \omega! & \text{per } \omega \leq n \\ 0 & \text{per } \omega > n \end{cases} = X \cdot \mathbb{1}_{[0,n]}$$

Applicando la definizione 6.5 per le VAR semplici:

$$\mathbb{E}[Y_n] = \sum_{\omega=0}^n e^{-\lambda} \frac{\lambda^\omega}{\omega!} \omega! = \sum_{\omega=0}^n e^{-\lambda} \lambda^\omega$$

Applichiamo la definizione di valore atteso per le VA positive:

$$\mathbb{E}[X] = \sup \left\{ \mathbb{E}[Y] : Y \text{ VAR semplice e } 0 \leq Y \leq X \right\} \geq \sup_n \left\{ \mathbb{E}[Y_n] \right\}$$

Il segno di maggiore o uguale è giustificato dal fatto che le  $Y_n$  definite come sopra rappresentano solo una parte delle possibili  $Y$  che approssimano  $X$ : l'estremo superiore di un insieme non aumenta (ovvero, o diminuisce o rimane invariato) se si prende un insieme più piccolo in esso contenuto.

Ma essendo che  $\lambda > 1$ , la somma delle sue potenze crescenti è illimitata al crescere di  $n$ , ovvero la serie diverge:

$$\mathbb{E}[X] \geq \sup_n \left\{ e^{-\lambda} \sum_{\omega=0}^n \lambda^\omega \right\} = e^{-\lambda} \lim_{n \rightarrow +\infty} \sum_{k=0}^n \lambda^k = +\infty$$

Necessariamente quindi  $\mathbb{E}[X] = +\infty$ . Questo accade nonostante la probabilità che  $X = +\infty$  sia nulla!

### 6.2.3 Variabili su $\mathbb{R}$

Conclusa la trattazione per VAR positive, è possibile espandere ulteriormente la teoria del valore atteso alla classe più generale di VAR, cioè quelle generiche con valori su tutto  $\mathbb{R}$ .

#### Definizione 6.9

Data  $X : \Omega \rightarrow \mathbb{R}$  VA generica, si definisce:

- **parte positiva** di  $X$  la VA  $X_+ := \max\{X, 0\}$ ;
- **parte negativa** di  $X$  la VA  $X_- := -\min\{X, 0\}$ .

La VAR generica si può scomporre come  $X = X_+ - X_-$ . Inoltre,  $|X| = X_+ + X_-$ .

#### Definizione 6.10

Sia  $X : \Omega \rightarrow \mathbb{R}$  una VA. Si dice che  $X$  **ammette valore atteso generalizzato** se  $\mathbb{E}[X_+]$  e  $\mathbb{E}[X_-]$  non sono entrambi infiniti. In tal caso, il **valore atteso** di  $X$  si definisce come:

$$\mathbb{E}[X] = \int_{\Omega} X d\mathbb{P} := \mathbb{E}[X_+] - \mathbb{E}[X_-], \quad \text{con } \mathbb{E}[X] \in [-\infty, +\infty]$$

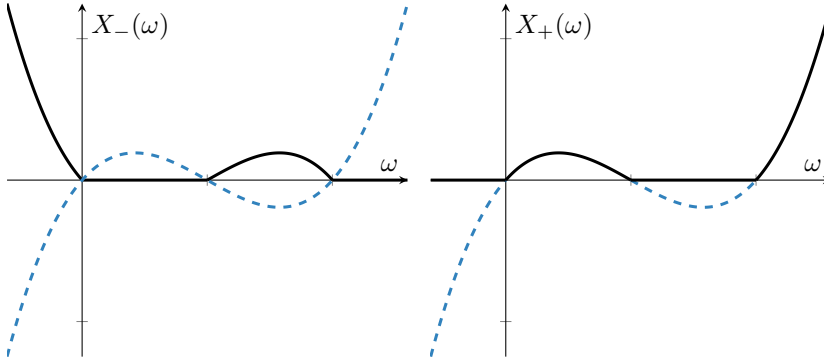


Figura 6.5: parte negativa e positiva di  $X$

**Definizione 6.11**

Chiamiamo  $\mathcal{L}^1(\Omega, \mathcal{A}, \mathbb{P})$  l'insieme delle VAR integrabili:

$$X : \Omega \rightarrow \mathbb{R} \text{ VAR}, X \in \mathcal{L}^1 \iff \mathbb{E}[X] \in \mathbb{R}$$

$X$  si dice **integrabile**, ovvero ha valore atteso finito, se  $\mathbb{E}[X_-] < +\infty$  e  $\mathbb{E}[X_+] < +\infty$ . In questo caso, il **valore atteso** di  $X$  è:

$$\mathbb{E}[X] = \int_{\Omega} X d\mathbb{P} := \mathbb{E}[X_+] - \mathbb{E}[X_-]$$

Stiamo quindi esplicitamente escludendo le  $X$  VAR con valore atteso infinito, sia esso  $+\infty$  o  $-\infty$ .

**NB:** La nuova definizione è compatibile con la 6.7 quando  $X \geq 0$ : le ultime definizioni sono un'estensione delle precedenti.

In particolare, il valore atteso per VA reali è ancora *lineare*.

**6.3 Proprietà del valore atteso e dello spazio  $\mathcal{L}^1$**

D'ora in poi, quando si parlerà di valore atteso senza ulteriori specificazioni, si intenderà sempre il valore atteso generalizzato, che è la sua definizione più ampia e che include tutte le altre.

**Teorema 6.12 (J-P Th 9.1)**

Siano  $X, Y, X_n \forall n \in \mathbb{N}$  VA reali.

1.  $\mathcal{L}^1$  è uno spazio vettoriale.  
Inoltre  $\mathbb{E} : \mathcal{L}^1 \rightarrow \mathbb{R}$  è una mappa lineare positiva, ovvero  $X \geq 0 \implies \mathbb{E}[X] \geq 0$  e, come precisato precedentemente:

$$\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y] \quad \forall a, b \in \mathbb{R}, \forall X, Y \in \mathcal{L}^1$$



2.  $X \in \mathcal{L}^1 \iff |X| \in \mathcal{L}^1$ .

Inoltre  $|\mathbb{E}[X]| \leq \mathbb{E}[|X|]$  e, se  $X$  è limitata (ovvero  $\exists c : |X(\omega)| \leq c \forall \omega$ ), allora  $X \in \mathcal{L}^1$ .

3. Siano  $X, Y$  VA tali che  $X \stackrel{\text{qc}}{=} Y$ . Allora  $\mathbb{E}[X] = \mathbb{E}[Y]$  (anche se infinito).

4. **Convergenza monotona:**

Siano  $X_n$  successione di VA e  $X$  VA tali che  $0 \leq X_n \leq +\infty$  e che  $X_n \uparrow X$  qc. Allora:

$$\mathbb{E}[X_n] \xrightarrow{n} \mathbb{E}[X], \quad \text{ovvero} \quad \mathbb{E}[X] = \mathbb{E}\left[\lim_n X_n\right] = \lim_n \mathbb{E}[X_n]$$

5. **Lemma di Fatou:**

Siano  $X_n$  successione di VAR e  $Y$  VAR tali che  $X_n \geq Y$  qc e  $Y \in \mathcal{L}^1$ .<sup>17</sup> Allora:

$$\mathbb{E}\left[\liminf_n X_n\right] \leq \liminf_n \mathbb{E}[X_n]$$

6. **Convergenza dominata:**

Siano  $X_n$  successione di VAR e  $X, Y$  VAR tali che  $X_n \xrightarrow{n} X$  qc,  $|X_n| \leq Y$  qc  $\forall n$ , e  $Y \in \mathcal{L}^1$ . Allora:

$$X_n \in \mathcal{L}^1, X \in \mathcal{L}^1 \quad \text{e} \quad \mathbb{E}[X_n] \xrightarrow{n} \mathbb{E}[X]$$

7.  $X \geq 0$  e  $\mathbb{E}[X] = 0 \implies X \stackrel{\text{qc}}{=} 0$ .

Convergenza monotona e convergenza dominata forniscono due importanti casi di *scambio di limiti* tra successione e integrale. Questo sarà chiarificato più avanti, quando calcoleremo il valore atteso come un integrale a tutti gli effetti.

Prima di dimostrare il teorema sarà necessario introdurre nozioni aggiuntive. Anzitutto, evidenziamo alcuni risultati notevoli sul valore atteso.

**Proposizione 6.13**

1.  $X \stackrel{\text{qc}}{=} 0 \implies \mathbb{E}[X] = 0$ ;
2.  $\mathbb{E}[X] > 0 \implies \mathbb{P}(X > 0) > 0$ ;
3.  $\mathbb{E}[|X|] = \mathbb{E}[X_+] + \mathbb{E}[X_-]$ ;
4.  $a \leq X \leq b \implies a \leq \mathbb{E}[X] \leq b$ .

---

<sup>17</sup>Si noti che  $Y$  non compare nella tesi, in quanto il suo ruolo è semplicemente quello di fornire un “limite inferiore” a  $X$  per garantirle un certo grado di regolarità. Molto spesso viene usata  $Y = 0$ , la variabile aleatoria identicamente nulla.

**Osservazione 6.14**

È sempre possibile approssimare una VA positiva, anche infinita, mediante il limite crescente di una successione di VAR semplici. Più precisamente, data una generica VA  $X$  tale che  $0 \leq X \leq +\infty$ , esiste una successione  $X_n$  di VAR semplici tale che  $0 \leq X_n \leq +\infty$  e  $X_n \uparrow X$ . Infatti, si può costruire  $X_n(\omega)$  in questo modo:

$$X_n(\omega) = \begin{cases} n & \text{per } X(\omega) \geq n \\ \frac{k}{2^n} & \text{per } \frac{k}{2^n} \leq X(\omega) \leq \frac{k+1}{2^n} \quad \text{con } k = 0, \dots, (n \cdot 2^n - 1) \end{cases}$$

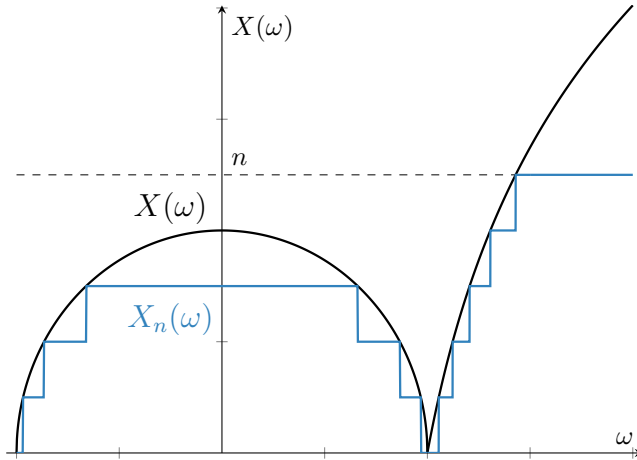


Figura 6.6: costruzione di una VAR semplice

Stiamo cioè approssimando  $X$  con la funzione a intervalli  $X_n$ , VAR semplice. Esprimiamo le controimmagini di  $X_n$  in altra forma:

$$\left( X_n = \frac{k}{2^n} \right) = \left( \frac{k}{2^n} \leq X \leq \frac{k+1}{2^n} \right) \in \mathcal{A}$$

Notiamo che  $|X_n(\omega) - X(\omega)| < \frac{1}{2^n} \forall \omega$ , e dunque per  $n \rightarrow +\infty$  l'errore tende a 0, ovvero l'approssimazione diventa sempre migliore. Pertanto  $X_n$  converge puntualmente a  $X$  ed, essendo che  $X_{n+1} \geq X_n \forall n, \forall \omega$ ,  $X_n$  si avvicinerà a  $X$  dal basso ( $X_n \uparrow X$ ).

**Lemma 6.15**

Siano  $X$  VAR e  $X_n$  successione di VAR semplici tali che  $0 \leq X \leq +\infty$ ,  $0 \leq X_n < +\infty$  e  $X_n \uparrow X$ . Allora:

$$\mathbb{E}[X_n] \xrightarrow{n} \mathbb{E}[X]$$

In questa costruzione c'è il cosiddetto “scambio di limiti”, anche se sotto ipotesi più restrittive della convergenza monotona o della convergenza dominata.

**Dimostrazione**

Per ipotesi  $X_n \uparrow X$ , ovvero  $X_{n+1} \geq X_n$ , dunque per la proposizione 6.6  $\mathbb{E}[X_{n+1}] \geq \mathbb{E}[X_n]$ ; si definisce allora il limite della successione come  $\mathbb{E}[X_n] \uparrow a := \lim_n \mathbb{E}[X_n]$ .

Si dimostra innanzitutto che  $a \leq \mathbb{E}[X]$ ; infatti:

$$\begin{aligned} \mathbb{E}[X] &= \sup \{ \mathbb{E}[Y] : 0 \leq Y \leq X, Y \text{ sempl.} \} && \text{(per definizione)} \\ &\geq \sup \{ \mathbb{E}[X_n] \} && \text{(sottoinsieme delle possibili } Y) \\ &= \lim_n \mathbb{E}[X_n] = a && (\mathbb{E}[X_n] \text{ cresce: ammette limite}) \end{aligned}$$

Dimostrando che  $a \geq \mathbb{E}[X]$  si ha l'uguaglianza della tesi.

Per farlo, si prenda  $Y$  VAR semplice tale che  $0 \leq Y \leq X$ , ovvero:

$$Y = \sum_{l=1}^m y_l \mathbf{1}_{(Y=y_l)}$$

Si prenda inoltre un  $\varepsilon \in (0, 1)$  arbitrario ma fissato, e si definisca la seguente successione:

$$\begin{aligned} Y_{n,\varepsilon} &:= (1 - \varepsilon) \cdot Y \cdot \mathbf{1}_{((1-\varepsilon)Y \leq X_n)} \\ &= (1 - \varepsilon) \cdot \left( \sum_{l=1}^n y_l \mathbf{1}_{(Y=y_l)} \right) \cdot \mathbf{1}_{((1-\varepsilon)Y \leq X_n)} && \text{(per definizione di } Y) \\ &= (1 - \varepsilon) \cdot \sum_{l=1}^n y_l \mathbf{1}_{(Y=y_l, (1-\varepsilon)Y \leq X_n)} && \text{(per le prop. di } \mathbf{1}) \end{aligned}$$

Ogni  $Y_{n,\varepsilon}$  è misurabile perché è data da prodotti e somme di VA.

Per costruzione  $Y_{n,\varepsilon} \leq X_n \leq X$ , perché si annulla quando  $(1 - \varepsilon)Y > X_n$ . Inoltre  $Y_{n,\varepsilon}$  è semplice perché assume i valori di  $(1 - \varepsilon)Y$ , che hanno cardinalità finita, oppure 0.

Il suo valore atteso è quindi:

$$\mathbb{E}[Y_{n,\varepsilon}] = (1 - \varepsilon) \sum_{l=1}^n y_l \cdot \mathbb{P}(Y = y_l, (1 - \varepsilon)Y \leq X_n)$$

La successione di controimmagini  $((1 - \varepsilon)Y \leq X_n)_n$  è crescente (ovvero gli insiemi sono “in scatolati”), perché per ipotesi  $X_n$  è crescente e quindi l'insieme si allarga, o al più rimane invariato.

Notando che  $(1 - \varepsilon)Y(\omega) < Y(\omega) \leq X(\omega) \quad \forall \omega$  si ottiene che  $(1 - \varepsilon)Y(\omega) < X(\omega)$ .

Per  $n$  sufficientemente grande si può affermare che  $(1 - \varepsilon)Y(\omega) < X_n(\omega) \leq X(\omega)$ , poiché al crescere di  $n$  il termine  $X_n$ , che deve tendere al limite  $X$ , si inserisce tra gli altri due, che invece restano costanti al variare di  $n$ . In altre parole,  $(1 - \varepsilon)Y(\omega) < X_n(\omega) \leq X(\omega)$  definitivamente (cioè per tutti gli  $n$  maggiori di un certo  $\bar{n}$ ) e per ogni  $\omega \in \Omega$ . Ciò vuol dire che ogni elemento dell'insieme  $\Omega$ , da un certo  $\bar{n}$  in poi, appartiene a una delle controimmagini  $((1 - \varepsilon)Y \leq X_n)$ , e quindi anche a tutte le successive:

$$\bigcup_{n=1}^{+\infty} ((1 - \varepsilon)Y \leq X_n) = \Omega \quad \implies \quad ((1 - \varepsilon)Y \leq X) = \Omega$$

Pertanto:

$$\begin{aligned} \mathbb{E}[Y_{n,\varepsilon}] &\xrightarrow{n \rightarrow +\infty} (1 - \varepsilon) \sum_{l=1}^{+\infty} y_l \cdot \mathbb{P}(Y = y_l, \underbrace{(1 - \varepsilon)Y \leq X}_{=\Omega}) \\ &= (1 - \varepsilon) \sum_{l=1}^{+\infty} y_l \cdot \mathbb{P}(Y = y_l) \\ &= (1 - \varepsilon) \mathbb{E}[Y] \end{aligned}$$

Per la positività del valore atteso,  $\mathbb{E}[Y_{n,\varepsilon}] \leq \mathbb{E}[X_n] \leq a \quad \forall n$ . Applicando i limiti per  $n \rightarrow +\infty$  al primo e al terzo membro ricaviamo  $(1 - \varepsilon) \mathbb{E}[Y] \leq a$ , da cui segue che  $\mathbb{E}[Y] \leq a$ , essendo  $0 < \varepsilon < 1$ .

Infine, ricordiamo che per definizione  $\mathbb{E}[X] = \sup \{ \mathbb{E}[Y] : 0 \leq Y \leq X, \text{ con } Y \text{ semplice} \}$ . Ma allora, per l'arbitrarietà di  $Y$ , dalle considerazioni precedenti ricaviamo che necessariamente  $\mathbb{E}[X] \leq a$ . Avendo prima anche dimostrato che  $\mathbb{E}[X] \geq a$ , si ha la tesi.  $\square$

*Esempio.* Siano dati  $\Omega$  discreto,  $\mathcal{A} = 2^\Omega$  e  $\mathbb{P} = p_\omega$ . Allora, per ogni VA  $X : \Omega \rightarrow [0, +\infty]$ , è vera la seguente formula:

$$\mathbb{E}[X] = \sum_{\omega} X(\omega) p_\omega$$

Questa è una prima formula esplicita per valori attesi di VA positive, anche più che semplici.

### Dimostrazione

Cominciamo mostrando che il risultato vale per le  $X$  semplici. Applicando la definizione:

$$X = \sum_{k=1}^n x_k \mathbb{1}_{A_k} \quad \text{con } A_k = (X = x_k)$$

Di conseguenza:

$$\begin{aligned} \sum_{\omega \in \Omega} X(\omega) p_\omega &= \sum_{\omega \in \Omega} \left( \sum_{k=1}^n x_k \mathbb{1}_{A_k}(\omega) p_\omega \right) = \sum_{k=1}^n x_k \left( \sum_{\omega \in \Omega} \mathbb{1}_{A_k}(\omega) p_\omega \right) \\ &= \sum_{k=1}^n \left( x_k \sum_{\omega \in A_k} p_\omega \right) = \sum_{k=1}^n x_k \mathbb{P}(A_k) = \mathbb{E}[X] \end{aligned}$$

Estendiamo il risultato alle VA della forma  $X : \Omega \rightarrow [0, +\infty)$ .

Cominciamo ordinando arbitrariamente gli elementi del dominio:  $\Omega = \{\Omega_i\}_{i \in \mathbb{N}}$ . Definiamo poi la successione  $X_n := X \cdot \mathbb{1}_{\{\omega_1, \dots, \omega_n\}}$ , che è tale che  $X_n \leq X_{n+1} \forall n$ , ovvero è crescente. Ciascuna delle  $X_n$  è una VA semplice in quanto la sua immagine ha cardinalità massima  $n + 1$ :

$$X_n(\omega_i) = \begin{cases} 0 & \text{per } i > n \\ X(\omega_i) & \text{per } i \leq n \end{cases}$$

Per  $n \rightarrow +\infty$ ,  $X_n(\omega_i) \uparrow X(\omega_i) \forall i$  in quanto la parte  $i > n$  non viene più scelta. Ma per il lemma 6.15  $\mathbb{E}[X] = \lim_n \mathbb{E}[X_n]$ , quindi:

$$\lim_n \mathbb{E}[X_n] = \lim_n \sum_{\omega \in \Omega} X_n(\omega) p_\omega = \lim_n \sum_{i=1}^n X_n(\omega_i) p_{\omega_i} = \sum_{\omega \in \Omega} X(\omega) p_\omega = \mathbb{E}[X]$$

Il caso  $X : \Omega \rightarrow [0, +\infty]$  è lasciato allo studente più motivato.  $\square$

### Dimostrazione del teorema 6.12

Saranno dimostrati solo i punti (1) e (2) del teorema. In particolare, il punto (3) sarà rimandato a un teorema successivo.

1. La tesi afferma che  $\mathcal{L}^1$  è uno spazio vettoriale.

Si comincia provando il teorema per le VA positive. Siano  $X, Y : \Omega \rightarrow [0, +\infty)$  VA tali che  $X, Y \in \mathcal{L}^1$  e  $\alpha > 0$ . Siano inoltre  $X_n$  e  $Y_n$  VA semplici tali che  $0 \leq X_n \leq X$ ,  $X_n \uparrow X$  e  $0 \leq Y_n \leq Y$ ,  $Y_n \uparrow Y$ .

Per la proposizione 6.6 le VA  $\alpha X_n$ , e inoltre  $\alpha X_n \uparrow \alpha X$ . Dunque, grazie allo scambio di limiti del lemma 6.15:

$$\mathbb{E}[\alpha X] = \mathbb{E} \left[ \lim_n \alpha X_n \right] = \lim_n \mathbb{E} [\alpha X_n]$$

Ciò dimostra che  $\mathcal{L}^1$  è chiuso rispetto al prodotto per scalare.

Occupiamoci ora della somma. Le VA  $X_n + Y_n$  sono semplici, maggiori o uguali a 0, e tali che  $X_n + Y_n \uparrow X + Y$ . Con un ragionamento simile al precedente e sfruttando ancora la proposizione 6.6 si dimostra che  $\mathcal{L}^1$  è

chiuso rispetto alla somma. La tesi è dunque dimostrata per tutte le VA positive.

Passiamo ora alle VA generiche. Siano  $X, Y : \Omega \rightarrow \mathbb{R}$  VA in  $\mathcal{L}^1$  e  $\alpha \in \mathbb{R}$ .

Allora si può scomporre  $\alpha X$  in parte positiva e parte negativa. In particolare:

$$(\alpha X)_+ + (\alpha X)_- = |\alpha X| = |\alpha| \cdot |X| = |\alpha| \cdot (X_+ + X_-)$$

Essendo tutti i termini positivi, si ottiene che:

$$\begin{cases} (\alpha X)_+ \leq |\alpha| \cdot (X_+ + X_-) \\ (\alpha X)_- \leq |\alpha| \cdot (X_+ + X_-) \end{cases}$$

Da qui, per la positività del valore atteso:

$$\begin{cases} \mathbb{E}[(\alpha X)_+] \leq \mathbb{E}[|\alpha| \cdot (X_+ + X_-)] = |\alpha| \cdot (\mathbb{E}[X_+] + \mathbb{E}[X_-]) < +\infty \\ \mathbb{E}[(\alpha X)_-] \leq \mathbb{E}[|\alpha| \cdot (X_+ + X_-)] = |\alpha| \cdot (\mathbb{E}[X_+] + \mathbb{E}[X_-]) < +\infty \end{cases}$$

Pertanto  $\alpha X \in \mathcal{L}^1$ , ovvero  $\mathcal{L}^1$  è chiuso rispetto al prodotto per scalare. Dimostriamo ora che effettivamente valga  $\mathbb{E}[\alpha X] = \alpha \mathbb{E}[X]$  distinguendo tra i seguenti tre casi e sfruttando la linearità del valore atteso:

- $\alpha > 0$ :

$$\begin{aligned} \mathbb{E}[\alpha X] &= \mathbb{E}[(\alpha X)_+] - \mathbb{E}[(\alpha X)_-] = \alpha \mathbb{E}[X_+] - \alpha \mathbb{E}[X_-] \\ &= \alpha (\mathbb{E}[X_+] - \mathbb{E}[X_-]) = \alpha \mathbb{E}[X] \end{aligned}$$

- $\alpha = 0$ :

$$0 = \alpha \mathbb{E}[X] = \mathbb{E}[(\alpha X)_+] - \mathbb{E}[(\alpha X)_-] = \mathbb{E}[0] - \mathbb{E}[0] = 0 = \alpha \mathbb{E}[X]$$

- $\alpha < 0$ :

$$\begin{aligned} \mathbb{E}[\alpha X] &= \mathbb{E}[(\alpha X)_+] - \mathbb{E}[(\alpha X)_-] = \mathbb{E}[|\alpha|X_-] - \mathbb{E}[|\alpha|X_+] \\ &= |\alpha| \mathbb{E}[X_-] - |\alpha| \mathbb{E}[X_+] = -|\alpha| (\mathbb{E}[X_+] - \mathbb{E}[X_-]) \\ &= -|\alpha| \mathbb{E}[X] = \alpha \mathbb{E}[X] \end{aligned}$$

Non resta che dimostrare che  $\mathcal{L}^1$  è chiuso rispetto alla somma. Applicando la disuguaglianza triangolare:

$$(X + Y)_+ + (X + Y)_- = |X + Y| \leq |X| + |Y| = X_+ + X_- + Y_+ + Y_-$$

Poiché le VA della disuguaglianza sono positive, si può applicare il valore atteso:  $\mathbb{E}[(X + Y)_+] \leq \mathbb{E}[X_+ + X_- + Y_+ + Y_-] < +\infty$ , e lo stesso vale per  $\mathbb{E}[(X + Y)_-]$ . Ne consegue che  $X + Y \in \mathcal{L}^1$ .

Definiamo ora la VA  $Z := X + Y$ , scomponibile come  $Z = Z_+ - Z_-$ .  
Si ha dunque:

$$\begin{aligned} Z &= X + Y \\ Z_+ - Z_- &= X_+ - X_- + Y_+ - Y_- \\ Z_+ + X_- + Y_- &= Z_- + X_+ + Y_+ \\ \mathbb{E}[Z_+ + X_- + Y_-] &= \mathbb{E}[Z_- + X_+ + Y_+] \\ \mathbb{E}[Z_+] + \mathbb{E}[X_-] + \mathbb{E}[Y_-] &= \mathbb{E}[Z_-] + \mathbb{E}[X_+] + \mathbb{E}[Y_+] \\ \mathbb{E}[Z_+] - \mathbb{E}[Z_-] &= \mathbb{E}[X_+] + \mathbb{E}[Y_+] - \mathbb{E}[X_-] - \mathbb{E}[Y_-] \end{aligned}$$

Si separano i termini positivi e negativi per avere entrambe i membri positivi e poter applicare la regola del valore atteso. Grazie alla linearità di  $\mathbb{E}$  si ricompono  $\mathbb{E}[X + Y] = \mathbb{E}[Z] = \mathbb{E}[Z_+] - \mathbb{E}[Z_-] = \mathbb{E}[X] + \mathbb{E}[Y]$ .

$\mathcal{L}^1$  è dunque uno spazio vettoriale anche per VA generiche.

2. La tesi afferma che  $X \in \mathcal{L}^1 \iff |X| \in \mathcal{L}^1$  e che  $|\mathbb{E}[X]| \leq \mathbb{E}[|X|]$ .

- ( $\implies$ ): per ipotesi,  $X \in \mathcal{L}^1$ , ovvero  $\mathbb{E}[X_+] < +\infty$  e  $\mathbb{E}[X_-] < +\infty$ .  
Perciò  $\mathbb{E}[|X|] = \mathbb{E}[X_+ + X_-] = \mathbb{E}[X_+] + \mathbb{E}[X_-] < +\infty$ , dunque  $|X| \in \mathcal{L}^1$ .
- ( $\impliedby$ ): per ipotesi,  $|X| \in \mathcal{L}^1$ , ovvero  $\mathbb{E}[|X|] = \mathbb{E}[X_+] + \mathbb{E}[X_-] < +\infty$ .  
Ma  $\mathbb{E}[X_+], \mathbb{E}[X_-] > 0$ , quindi la loro somma è finita se e solo se sono entrambi minori di  $+\infty$ .  
Dunque  $\mathbb{E}[X] = \mathbb{E}[X_+ - X_-] = \mathbb{E}[X_+] - \mathbb{E}[X_-] < +\infty$ .

Grazie alla disuguaglianza triangolare, in pochi passaggi si verifica anche la seconda richiesta:

$$|\mathbb{E}[X]| = |\mathbb{E}[X_+] - \mathbb{E}[X_-]| \leq \mathbb{E}[X_+] + \mathbb{E}[X_-] = \mathbb{E}[|X|] \quad \square$$

*Controesempio* per i punti (4), (5) e (6) del teorema 6.12.

Si definisca lo spazio di probabilità  $\Omega = \mathbb{N} \cup \{0\}$ ,  $\mathcal{A} = 2^\Omega$ ,  $\mathbb{P} = Po(\lambda)$  e in esso la successione di VA  $\{X_n\}_{n \in \mathbb{N}}$  così definita:

$$X_n(k) := \frac{n!}{\lambda^n} \mathbb{1}_{\{n\}}(k)$$

Notiamo che  $X_n \geq 0 \forall n$ , e che  $X_n \xrightarrow{n} 0$  *puntualmente*  $\forall k \in \Omega$  (in quanto per tutti i  $k > n$ ,  $X_n$  è definitivamente nullo); quindi  $\liminf_n X_n = \lim_n X_n = 0$  che ha valore atteso 0.

Inoltre, le  $X_n$  sono semplici perché possono assumere solo i valori 0 e  $\frac{n!}{\lambda^n}$ . Applicando la formula del valore atteso per le VA semplici:

$$\mathbb{E}[X_n] = \frac{n!}{\lambda^n} \cdot \mathbb{P}(\{n\}) + 0 \cdot \mathbb{P}(\{n\}^C) = \frac{n!}{\lambda^n} \cdot e^{-\lambda} \frac{\lambda^n}{n!} + 0 = e^{-\lambda}$$

Ovviamente  $\liminf_n \mathbb{E}[X_n] = \lim_n \mathbb{E}[X_n] = e^{-\lambda}$ .

Abbiamo dunque trovato un esempio in cui *non* vale l'uguaglianza  $\mathbb{E} \left[ \liminf_n X_n \right] = \liminf_n \left\{ \mathbb{E}[X_n] \right\}$ . Questo perché in generale *non* è vero che l'integrale del limite è uguale al limite dell'integrale; in particolare, in questo caso, la sola convergenza puntuale non è sufficiente per garantire uno scambio di limiti. Esso infatti può avvenire solo sotto ipotesi più restrittive, quali per esempio la convergenza monotona e la convergenza dominata, nessuna delle quali è presente in questo esempio ( $X_n$  può essere dominata solo da una variabile con valore atteso infinito). Si noti, infine, che la disuguaglianza del lemma di Fatou è comunque verificata perché  $X_n \geq 0$ .

*Esempio.* Siano  $\Omega$  un generico dominio *discreto*,  $\mathcal{A} = 2^\Omega$ ,  $\mathbb{P} \leftrightarrow \{p_\omega\}_\omega$  e  $X \in \mathcal{L}^1$ . Dimostriamo che:

$$\mathbb{E}[X] = \int_\Omega X d\mathbb{P} = \sum_{\omega \in \Omega} X(\omega)p_\omega$$

**Dimostrazione**

Ordiniamo i punti di  $\Omega$  nella successione  $\{\omega_i\}_{i \in \mathbb{N}}$  e definiamo la seguente VA:

$$X_n(\omega) := X(\omega)\mathbb{1}_{\{\omega_1, \omega_2, \dots, \omega_n\}}(\omega) = \begin{cases} X(\omega) & \text{da } \omega_1 \text{ a } \omega_n \\ 0 & \text{da } \omega_{n+1} \text{ in poi.} \end{cases}$$

Per costruzione questo garantisce che  $X_n(\omega) \xrightarrow{n} X(\omega) \quad \forall \omega \in \Omega$ .

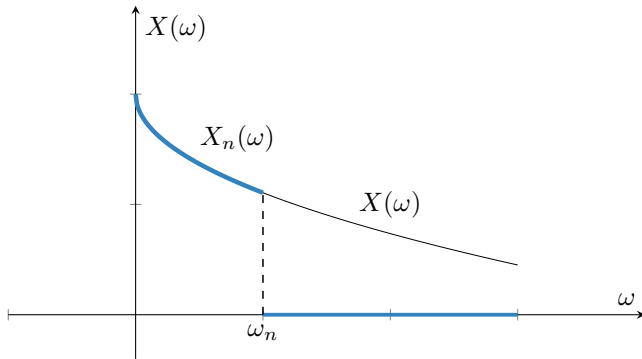


Figura 6.7: rappresentazione di  $X_n(\omega)$  e  $X(\omega)$



Inoltre  $|X_n(\omega)| \leq |X(\omega)| \quad \forall \omega \in \Omega$ , e  $X \in \mathcal{L}^1$ . Possiamo dunque applicare la convergenza dominata:

$$\mathbb{E}[X_n] \xrightarrow{n} \mathbb{E}[X] \implies \lim_n \mathbb{E}[X_n] = \lim_n \sum_{i=1}^n X_n(\omega_i) p_{\omega_i} = \sum_{\omega \in \Omega} X(\omega) p_{\omega}$$

L'ultimo passaggio è giustificato dal fatto che  $X_n$  può essere scritto come “finta sommatoria” con al massimo 1 addendo non nullo:

$$X_n(\omega) = \sum_{i=1}^n X(\omega) \mathbf{1}_{\{\omega_i\}}(\omega) \quad \square$$

## 6.4 Spazi $L^p$

### Osservazione 6.16

Date due VA  $X$  e  $Y$ , sappiamo già che:

$$X \stackrel{\text{qc}}{=} Y \implies \mathbb{E}[X] = \mathbb{E}[Y] \quad \text{e} \quad P^X = P^Y$$

Ovvero  $\mathbb{E}[X]$  e  $P^X$  non cambiano:

- modificando  $X$  su un evento improbabile, perché un'uguaglianza quasi certa non è influenzata da tali eventi;
- conoscendo  $X$  solo su un evento quasi certo;
- se  $X$  esiste solo su un evento quasi certo, perché l'uguaglianza quasi certa è determinata da tali eventi.

Quindi, per studiare una certa VA (per esempio per calcolarne valore atteso e legge) è sufficiente conoscerla/studiarla solo su un evento quasi certo, che di fatto rappresenta l'intera VA eccetto uno o più eventi improbabili che, appunto, non influenzano l'andamento della variabile.

Infine, è anche possibile osservare che l'uguaglianza quasi certa è una *relazione di equivalenza*: è infatti riflessiva ( $X \stackrel{\text{qc}}{=} X$ ), simmetrica ( $X \stackrel{\text{qc}}{=} Y \iff Y \stackrel{\text{qc}}{=} X$ ), e transitiva ( $X \stackrel{\text{qc}}{=} Y$  e  $Y \stackrel{\text{qc}}{=} Z \iff X \stackrel{\text{qc}}{=} Z$ ).

La transitività necessita di una breve dimostrazione: si definiscano gli eventi  $B_1 = (X = Y)$ ,  $B_2 = (Y = Z)$  e  $B_3 = (X = Z) \supseteq B_1 \cap B_2$ . Quest'ultima relazione, unitamente al fatto che  $\mathbb{P}(B_1) = \mathbb{P}(B_2) = 1$  per definizione di  $B_1$  e  $B_2$ , fornisce che:

$$\mathbb{P}(B_3) \geq \mathbb{P}(B_1 \cap B_2) = \mathbb{P}(B_1) = 1 \implies \mathbb{P}(B_3) = 1$$

Il fatto che l'uguaglianza quasi certa sia una relazione di equivalenza ci porta a formulare le seguenti definizioni:

**Definizione 6.17**

L'insieme  $\mathcal{L}^1$  quozientato rispetto all'uguaglianza quasi certa è l'insieme chiamato  $L^1$ . Non ci si riferisce più alla variabile aleatoria  $X$ , ma alla sua **classe di equivalenza**  $[X]$ , ovvero l'insieme così definito:

$$[X] := \{Y \in \mathcal{L}^1 : X \stackrel{\text{qc}}{=} Y\}$$

$L^1$  ed  $\mathcal{L}^1$  sono ben distinti perché, come fatto notare poc'anzi,  $\mathbb{E}[X]$  e  $P^X$  sono invarianti dell'intera classe  $[X]$  (e non della sola variabile  $X$ ) e la caratterizzano. Tuttavia, in questo testo  $L^1$  ed  $\mathcal{L}^1$  saranno usati in maniera interscambiabile come se fossero sinonimi<sup>18</sup>.

**Definizione 6.18**

Dato  $p \in [1, +\infty]$ , si definisce  $\mathcal{L}^p$  l'insieme delle VA reali  $X$  tali che  $|X|^p \in \mathcal{L}^1$  e  $L^p$  l'insieme delle classi di equivalenza di VA reali  $X$  tali che  $|X|^p \in L^1$ .

Da queste definizioni discende che:

$$X, Y \in \mathcal{L}^p \text{ e } X \stackrel{\text{qc}}{=} Y \iff [X] = [Y] \in L^p$$

**6.4.1 Spazio  $L^2$** 

**Teorema 6.19:**  $L^2$  vs  $L^1$  (*J-P Th 9.3*)

1. Siano  $X, Y \in L^2$ . Allora  $XY \in L^1$  e inoltre vale la **disuguaglianza di Cauchy-Schwarz**:

$$|\mathbb{E}[XY]| \leq \sqrt{\mathbb{E}[X^2] \mathbb{E}[Y^2]}$$

2.  $L^2 \subseteq L^1$ , e, se  $X \in L^2$ , allora  $(\mathbb{E}[X])^2 \leq \mathbb{E}[X^2]$ .
3.  $L^2$  è uno spazio vettoriale.

**Dimostrazione**

Dalla formula del quadrato di binomio è noto che:

$$|ab| \leq \frac{a^2}{2} + \frac{b^2}{2} \quad \forall a, b \in \mathbb{R}$$

Quindi, prese  $X, Y$  VAR:

$$|X(\omega)Y(\omega)| \leq \frac{X^2(\omega)}{2} + \frac{Y^2(\omega)}{2} \quad \forall \omega \in \mathbb{R}$$

<sup>18</sup>Si ricorre quindi a un lieve abuso di *scrittura*.

In particolare, se  $X, Y \in L^2$ , per definizione  $X^2, Y^2 \in L^1$ . Pertanto:

$$\frac{X^2}{2}, \frac{Y^2}{2} \in L^1 \implies \frac{X^2}{2} + \frac{Y^2}{2} \in L^1 \implies |XY| \in L^1 \implies XY \in L^1$$

$XY, X^2$  e  $Y^2$  sono quindi tutte VA e ammettono valore atteso. Ora, preso un generico  $a \in \mathbb{R}$ ,  $(aX + Y)^2$  è una VA per la dimostrazione precedente e perché la somma di variabili integrabili è integrabile. Quindi, si può scrivere che:

$$\begin{aligned} 0 \leq \mathbb{E}[(aX + Y)^2] &= \mathbb{E}[a^2X^2 + 2aXY + Y^2] \\ &= a^2\mathbb{E}[X^2] + 2a\mathbb{E}[XY] + \mathbb{E}[Y^2] \end{aligned}$$

Quest'ultima scomposizione è possibile perché tutti gli addendi sono integrabili e per la linearità del valore atteso. Vedendo la disuguaglianza come una disequazione di secondo grado in  $a$ , deduciamo che il polinomio in  $a$  in quanto non negativo deve avere discriminante<sup>19</sup> non positivo:

$$\frac{\Delta}{4} \leq 0 \implies [\mathbb{E}[XY]]^2 - \mathbb{E}[X^2]\mathbb{E}[Y^2] \leq 0$$

Segue immediatamente la disuguaglianza di Cauchy-Schwarz.

La tesi (2) segue altrettanto immediatamente prendendo  $Y = 1$ .

Rimane da dimostrare che  $L^2$  è uno spazio vettoriale. Siano  $X, Y \in L^2$  e  $a, b \in \mathbb{R}$ :

$$(aX + bY)^2 = a^2X^2 + b^2Y^2 + 2abXY$$

Tutti gli addendi sono elementi di  $L^1$  (l'ultimo grazie alla dimostrazione del punto (1)), dunque anche la loro somma lo è, e anche la VA  $aX + bY$ . Pertanto, per definizione,  $(aX + bY) \in L^2$ , il quale è quindi chiuso rispetto a combinazione lineare, ovvero è uno spazio vettoriale.  $\square$

**NB:**  $L^2 \subseteq L^1$  perché  $\mathbb{P}(\Omega) < +\infty$ . Infatti vedremo che questa proprietà non potrà essere generalizzata alle *measure*, estensioni delle probabilità, in quanto potranno anche avere valore massimo infinito.

## 6.5 Varianza

### Definizione 6.20

Sia  $X \in L^2(\Omega, \mathcal{A}, \mathbb{P})$ . Definiamo la **varianza** di  $X$  come:

$$\text{Var}(X) = \sigma^2 := \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[(X - \mu)^2]$$

$\sigma^2$  è ben definito ed è reale perché:

- $\mu$  è ben definito in quanto  $X \in L^2$ ;

---

<sup>19</sup>Penso che siano 5 anni che non dico o scrivo “discriminante”

- $(X - \mu) \in L^2 \implies \mathbb{E}[(X - \mu)^2] < +\infty$ .

La varianza rappresenta il “grado di sparpagliamento” dei possibili valori che  $X$  può assumere, rispetto al valore centrale, cioè il valore atteso  $\mu$ .

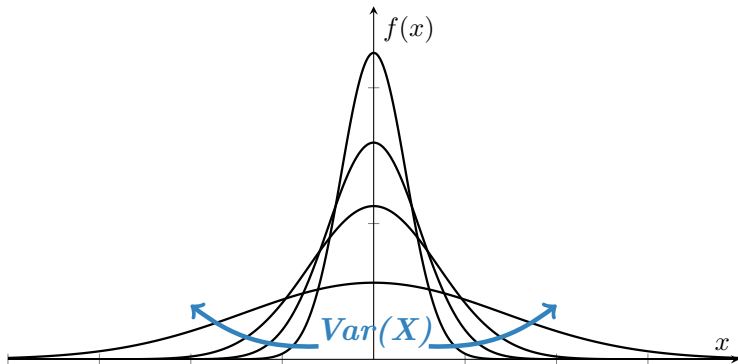


Figura 6.8: gaussiane con  $\mu = 0$  e diversa varianza

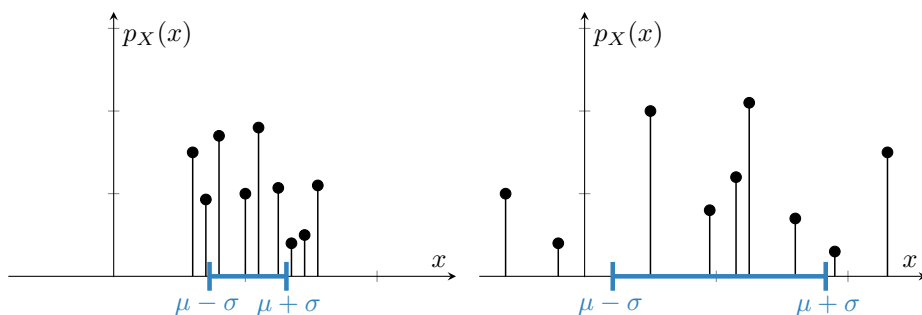


Figura 6.9: varianza per due VA discrete

### Osservazione 6.21

Dalla definizione sappiamo che  $\sigma^2 \geq 0$ ; dunque:

$$\mathbb{E}[(X - \mu)^2] = \mathbb{E}[X^2] - 2\mu\mathbb{E}[X] + \mathbb{E}[\mu^2] = \mathbb{E}[X^2] - 2\mathbb{E}[X]\mathbb{E}[X] + \mu^2$$

Poiché  $\mathbb{E}[X] = \mu$ , abbiamo ottenuto un'utilissima formula di calcolo per la varianza:

$$\sigma^2 = \mathbb{E}[X^2] - \mathbb{E}^2[X]$$

Il caso limite  $\sigma^2 = 0$  significa che la probabilità si concentra su un solo punto, portando così ad una distribuzione degenera, denominata delta di Dirac (definizione 4.5). In questo caso si può affermare che  $\mathbb{P}(X = \mu) = 1$  e che non esiste più

l'incertezza. Questo aspetto è di fondamentale importanza per i vettori gaussiani e verrà approfondito nel capitolo 12.

### 6.5.1 Proprietà

**Teorema 6.22: disuguaglianza di Markov** (*J-P Coro 5.1*)

Siano  $X$  VAR e  $a > 0$  un generico valore di soglia. Allora:

$$\mathbb{P}(|X| \geq a) \leq \frac{1}{a} \mathbb{E}[|X|]$$

#### Dimostrazione

Definiamo la VAR  $Y$ :

$$Y := a \cdot \mathbf{1}_{(|X| \geq a)} = \begin{cases} a & \text{per } |X| \geq a \\ 0 & \text{per } |X| < a \end{cases}$$

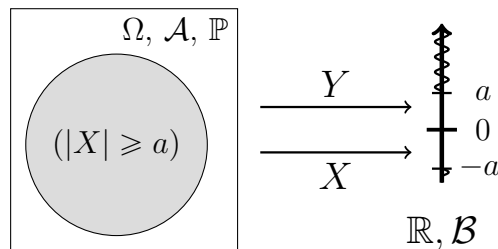


Figura 6.10: rappresentazione della VA  $Y$

Notiamo che  $|X(\omega)| \geq Y(\omega) \forall \omega$ . Abbiamo dunque:

$$\mathbb{E}[|X|] \geq \mathbb{E}[Y] = a \cdot \mathbb{E}[\mathbf{1}_{(|X| \geq a)}] = a \cdot \mathbb{P}(|X| \geq a)$$

Dividendo per  $a$  si ottiene la tesi. □

**Corollario 6.23: disuguaglianza di Chebyshev** <sup>20</sup> (*J-P Th 9.4*)

Siano  $X$  VAR e un generico valore di soglia  $a > 0$ . Allora:

$$\mathbb{P}(|X - \mu| \geq a) \leq \frac{\text{Var}(X)}{a^2}$$

La dimostrazione è immediata<sup>21</sup> e segue da Markov prendendo  $(X - \mu)^2$  come VA e  $a^2$  come valore di soglia.

<sup>20</sup>In realtà si scrive Čebyšev e si legge *Ciebicìoff* (Micheletti approves)

<sup>21</sup>La saprebbe fare anche un gestionale

Questa disuguaglianza rivela che la varianza controlla la probabilità che  $X$  si allontani dal valore centrale  $\mu$  (come peraltro era già stato fatto notare nella sua definizione): più piccola è la varianza, minore è la probabilità (massima) che l'“errore” di  $X$  rispetto a  $\mu$  sia superiore a una certa soglia manipolabile arbitrariamente.

**Corollario 6.24**

Sia  $X$  VAR positiva e tale che  $\mathbb{E}[X] = 0$ . Allora  $X \stackrel{\text{qc}}{=} 0$ .

**Dimostrazione**

Sia un arbitrario  $a > 0$ . Per la disuguaglianza di Markov  $\mathbb{P}(|X| \geq a) \leq \mathbb{E}[|X|]$ , con  $|X| = X$  per ipotesi. Allora  $\mathbb{P}(X \geq a) = 0$ . La successione di eventi  $(X \geq \frac{1}{n})$  è crescente, e in particolare  $(X \geq \frac{1}{n}) \uparrow (X > 0)$ ; pertanto:

$$\mathbb{P}\left(X \geq \frac{1}{n}\right) \rightarrow \mathbb{P}(X > 0) = 0 \implies \mathbb{P}(X = 0) = 1 \text{ ovvero } X \stackrel{\text{qc}}{=} 0 \quad \square$$

**Corollario 6.25 al corollario<sup>22</sup>**

Sia  $X \in L^1$  VAR. Allora:

$$\begin{aligned} \sigma^2 = \mathbb{E}[(X - \mu)^2] = 0 &\iff \exists c \in \mathbb{R} : X \stackrel{\text{qc}}{=} c \\ &\iff \exists c : X \sim \delta_c \text{ (delta di Dirac in } c) \end{aligned}$$

**Dimostrazione**

( $\iff$ ): Grazie alla regola del valore atteso per le VAR semplici, è noto che  $\mu = c$ . Si ha quindi che  $X - \mu \stackrel{\text{qc}}{=} 0$ , ovvero che  $(X - \mu)^2 \stackrel{\text{qc}}{=} 0$ . Similmente a prima, si conclude che  $\sigma^2 = \mathbb{E}[(X - \mu)^2] = 0$ .

( $\implies$ ): Poiché  $(X - \mu)^2$  è una VAR positiva, per il corollario precedente  $(X - \mu)^2 \stackrel{\text{qc}}{=} 0$ , ovvero  $X \stackrel{\text{qc}}{=} \mu$ , dove  $\mu$  è il  $c$  cercato. □

**6.6 Valore atteso di serie di variabili aleatorie**

**Teorema 6.26** (*J-P Th 9.2*)

Sia  $\{X_n\}_n$  una successione di VA sullo spazio di probabilità  $(\Omega, \mathcal{A}, \mathbb{P})$ .

1. Se le  $X_n$  sono positive, ovvero se  $X_n : \Omega \rightarrow [0, +\infty]$ , allora:

$$\mathbb{E}\left[\sum_{n=1}^{+\infty} X_n\right] = \sum_{n=1}^{+\infty} \mathbb{E}[X_n] \in [0, +\infty]$$

2. Se  $\sum_{n=1}^{+\infty} \mathbb{E}[|X_n|] < +\infty$ , allora:

---

<sup>22</sup>We heard you like corollari so we put a corollario into your corollario

- (i)  $\sum_{n=1}^{+\infty} X_n$  converge quasi certamente;
- (ii)  $\sum_{n=1}^{+\infty} X_n \in L^1$ ;
- (iii)  $\mathbb{E} \left[ \sum_{n=1}^{+\infty} X_n \right] = \sum_{n=1}^{+\infty} \mathbb{E}[X_n]$ .<sup>23</sup>

**Dimostrazione**

Definiamo le seguenti somme parziali:

$$S_n := \sum_{k=1}^n |X_k| \quad \text{e} \quad T_n := \sum_{k=1}^n X_k$$

Ne segue, per la linearità (su somme finite) del valore atteso, che:

$$\mathbb{E}[S_n] = \mathbb{E} \left[ \sum_{k=1}^n |X_k| \right] = \sum_{k=1}^n \mathbb{E}[|X_k|]$$

Essendo  $S_n$  una somma di termini positivi, è crescente e ammette pertanto limite, una VA che chiameremo  $S$ :

$$S_n \uparrow S \quad \forall \omega, \quad \text{con} \quad S : \Omega \rightarrow [0, +\infty], \quad S = \sum_{k=1}^{+\infty} |X_k|$$

1. Se  $X_n \geq 0$ , la tesi è immediata in quanto  $|X_n| = X_n$ .
2. Per ipotesi  $\sum_{n=1}^{+\infty} \mathbb{E}[|X_n|] < +\infty$ . Quindi, per dimostrazione precedente, vale anche:

$$\mathbb{E} \left[ \sum_{k=1}^{+\infty} |X_k| \right] = \mathbb{E}[S] < +\infty \implies \mathbb{P}(S = +\infty) = 0 \implies \mathbb{P}(S < +\infty) = 1$$

In altre parole  $S < +\infty$  quasi certamente, che è la tesi (i).

Ora, per la disuguaglianza triangolare si ha che:

$$|T_n| = \left| \sum_{k=1}^n X_k \right| \leq \sum_{k=1}^n |X_k| = S_n \quad \forall \omega, \forall n$$

Prendendo i limiti su  $n$ :

$$|T_n| \xrightarrow[n \rightarrow +\infty]{\text{qc}} |T| = \left| \sum_{n=1}^{+\infty} X_n \right| \quad \text{e} \quad S_n \xrightarrow[n \rightarrow +\infty]{\forall \omega} S = \sum_{n=1}^{+\infty} |X_n|$$

---

<sup>23</sup>“Ovvero, siamo nel migliore dei mondi possibili” -Gottfried von Gregoratti

Ne deduciamo che  $0 \leq |T| \leq S$  quasi certamente; per le proprietà del valore atteso con VA positive, vale:

$$0 \leq \mathbb{E} \left[ \sum_{n=1}^{+\infty} X_n \right] \leq \mathbb{E} \left[ \sum_{n=1}^{+\infty} |X_n| \right] < +\infty \quad \text{per ipotesi}$$

Ne segue che:

$$\left| \sum_{n=1}^{+\infty} X_n \right| \in L^1 \implies \sum_{n=1}^{+\infty} X_n \in L^1$$

Questo risultato è la tesi (ii). Da qui segue anche che:

$$T_n = \sum_{k=1}^n X_k \xrightarrow[n \rightarrow +\infty]{\text{qc}} T = \sum_{k=1}^{+\infty} X_k$$

Poiché  $|T_n| \leq S_n \leq S \in L^1$ , per convergenza dominata vale la tesi (iii):

$$\mathbb{E} \left[ \sum_{n=1}^{+\infty} X_n \right] = \lim_{n \rightarrow +\infty} \mathbb{E}[T_n] = \sum_{n=1}^{+\infty} \mathbb{E}[X_n] \quad \square$$

## 6.7 Integrazione di funzioni misurabili

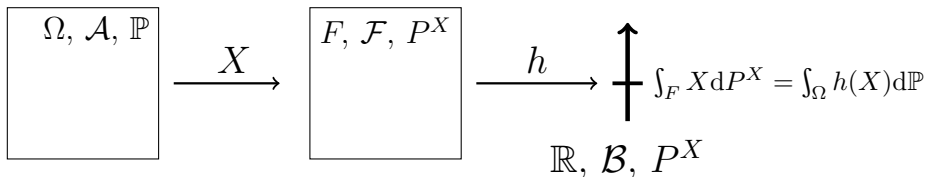
**Teorema 6.27: regola del valore atteso** (*J-P Th 9.5*)

Siano  $(\Omega, \mathcal{A}, \mathbb{P})$  spazio di probabilità,  $(F, \mathcal{F})$  spazio misurabile,  $X : \Omega \rightarrow F$  VA, e  $h : F \rightarrow \mathbb{R}$  oppure  $[0, +\infty]$  funzione misurabile. Allora:

1.  $h(X) \in L^1(\Omega, \mathcal{A}, \mathbb{P}) \iff h \in L^1(F, \mathcal{F}, P^X)$
2. Sia nel caso  $h : F \rightarrow [0, +\infty]$ , sia nel caso  $h \in L^1(\mathbb{R}, \mathcal{B}, P^X)$ , si ha:

$$\int_{\Omega} h(X) d\mathbb{P} = \int_F h dP^X$$

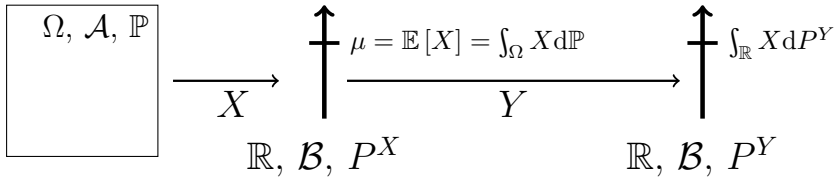
Questo teorema è molto importante nel calcolo del valore atteso, perché afferma che svolgere l'integrale su  $\Omega$  corrisponde a calcolarlo su  $F$ . Infatti, in molti casi il dominio  $\Omega$  può essere un insieme astratto composto da oggetti qualsiasi, su cui non è definita l'integrabilità.





Imponendo  $F = \mathbb{R}$  il calcolo si trasforma in un integrale su  $\mathbb{R}$ , rendendolo possibile o più facile. Questo è il principale metodo di risoluzione degli esercizi per semplificare i conti.

Osserviamo che  $\mathbb{E}$  e  $\int_{\Omega}$  sono la stessa cosa, tuttavia a seconda dei casi una notazione può risultare più comoda rispetto all'altra.



**Dimostrazione**

- (i) Inizialmente, si prenda per semplicità il caso particolare in cui  $h$  è una funzione indicatrice:  $h = \mathbb{1}_B$ , con  $B \in \mathcal{F}$ . Dunque, sfruttando la definizione di controimmagine:

$$\mathbb{1}_B(X) = \begin{cases} 1 & \text{per } X(\omega) \in B \\ 0 & \text{per } X(\omega) \notin B \end{cases} = \begin{cases} 1 & \text{per } \omega \in (X \in B) \\ 0 & \text{per } \omega \in (X \notin B) \end{cases} = \mathbb{1}_{(X \in B)}(\omega)$$

Da ciò segue che:

$$\int_{\Omega} \mathbb{1}_B(X) d\mathbb{P} = \int_{\Omega} \mathbb{1}_{(X \in B)} d\mathbb{P} = 1 \cdot \mathbb{P}(X \in B) = 1 \cdot P^X(B) = \int_F \mathbb{1}_B dP^X$$

Il risultato è la tesi per  $h = \mathbb{1}_B$ .

- (ii) Prendendo una funzione  $h$  semplice, notiamo che è semplice anche la sua composizione:

$$h = \sum_{k=1}^n h_k \mathbb{1}_{B_k} \quad \text{con } h_k \in \mathbb{R}, B_k \in \mathcal{F}$$

La dimostrazione si svolge, come nel caso precedente, sfruttando la linearità degli integrali.

- (iii) Dimostriamo ora la regola nel caso in cui il codominio sia positivo:  $h : F \rightarrow [0, +\infty]$  misurabile. È possibile costruire una successione di funzioni semplici  $h_n \geq 0$  tali che  $h_n \uparrow h$ . Questo implica che le  $Y_n := h_n(X)$  sono VA semplici, che  $Y_n \geq 0 \forall n$  e che  $Y_n \uparrow h(X)$ .

Grazie alla convergenza monotona possiamo scambiare integrale e limite e

sfruttare la proprietà delle funzioni semplici dimostrata al punto (ii):

$$\begin{aligned} \int_{\Omega} h(x) d\mathbb{P} &= \int_{\Omega} \lim_n h_n(X) d\mathbb{P} \stackrel{\text{cm}}{=} \lim_n \int_{\Omega} h_n(X) d\mathbb{P} \\ &= \lim_n \int_F h_n dP^X \stackrel{\text{cm}}{=} \int_F \lim_n h_n dP^X = \int_F h dP^X \end{aligned}$$

Ma visto che  $h \in L^1(\mathbb{P}) \iff h \in L^1(P^X)$ , è evidente che:

$$\int_{\Omega} h(x) d\mathbb{P} = \int_F h dP^X \quad \forall h : F \rightarrow [0, +\infty] \text{ misurabili}$$

(iv) Si dimostra ora il caso generale:  $h \in L^1(P^X)$  con  $h : F \rightarrow \mathbb{R}$  misurabile.

Si può separare  $h$  nella sua parte positiva  $h_+$  e negativa  $h_-$ :

$$\begin{aligned} h(X)_+(\omega) &= \begin{cases} h(X)(\omega) & \text{per } h(X)(\omega) \geq 0 \\ 0 & \text{per } h(X)(\omega) < 0 \end{cases} \\ &= \begin{cases} h(X(\omega)) & \text{per } h(X(\omega)) \geq 0 \\ 0 & \text{per } h(X(\omega)) < 0 \end{cases} = h_+(X(\omega)) \end{aligned}$$

Analogamente per la parte negativa  $h(X)_-(\omega) = h_-(X(\omega))$ .

Grazie alla linearità degli integrali possiamo dividere anche questo in parte positiva e negativa:

$$\int_{\Omega} h(x) d\mathbb{P} = \int_{\Omega} h(x)_+ d\mathbb{P} - \int_{\Omega} h(x)_- d\mathbb{P} = \int_{\Omega} h_+(X) d\mathbb{P} - \int_{\Omega} h_-(X) d\mathbb{P}$$

Essendo entrambe le funzioni positive, per la dimostrazione al punto (iii):

$$\int_{\Omega} h_+(X) d\mathbb{P} - \int_{\Omega} h_-(X) d\mathbb{P} = \int_F h_+ dP^X - \int_F h_- dP^X = \int_F h dP^X$$

Sfruttando la linearità dell'integrale al contrario otteniamo così la tesi per VA generiche.  $\square$

**Corollario 6.28**

Siano  $X$  e  $Y$  VA semplici. Allora:

$$X \stackrel{\text{qc}}{=} Y \implies P^X = P^Y \implies \mathbb{E}[X] = \mathbb{E}[Y]$$

Il corollario amplia il simile risultato che era stato ottenuto per le VA reali: qui  $X$  e  $Y$  non sono necessariamente reali.

**Definizione 6.29**

Dati  $X$  VAR e  $A \in \mathcal{A}$  misurabile, si definisce **integrale su un insieme** misurabile della VAR la seguente quantità:

$$\int_A X \, d\mathbb{P} := \int_{\Omega} X \cdot \mathbb{1}_A \, d\mathbb{P} = \mathbb{E}[X \mathbb{1}_A]$$

Si noti come nella definizione sia importante la misurabilità di  $A$ , senza la quale non è possibile definire un'indicatrice misurabile. Inoltre si può verificare che questa definizione ha senso in quanto la composizione di misurabili è misurabile.

**NB:** La probabilità di un evento  $A$  è anche calcolabile con il valore della sua indicatrice:

$$\mathbb{P}(A) = \int_A 1 \, d\mathbb{P} = \mathbb{E}[\mathbb{1}_A]$$



---

## 7 Variabili aleatorie reali discrete

In questo capitolo svilupperemo la trattazione di un'importante famiglia di variabili aleatorie reali, quelle *discrete*. Esse sono un'estensione delle variabili semplici, nel senso che la loro probabilità è concentrata su singoli punti, i quali, però, possono anche essere in quantità *numerabile* (e quindi non finita).

### Teorema 7.1

Dati  $(\Omega, \mathcal{A}, \mathbb{P})$  spazio di probabilità e  $X : \Omega \rightarrow \mathbb{R}$  variabile aleatoria reale, allora  $X$  è una variabile aleatoria reale discreta se e solo se:

- $P^X$  è una probabilità discreta, cioè al più concentrata su un insieme numerabile di punti; ovvero, esiste un insieme al più numerabile  $\{x_k\}_k$  di numeri reali per cui esistono coefficienti  $p_k$ , con  $p_k \geq 0$  e  $\sum_k p_k = 1$ , tali che  $P^X(B) = \sum_{x_k \in B} p_k$ ;
- $F_X$  è costante a tratti, quindi troviamo in ogni  $x_k$  un salto di valore pari a  $p_k$ ;
- $\mathbb{P}(X \in \{x_k\}_k) = P^X(\{x_k\}) = \sum_k p_k = 1$ ;
- $\text{im}(X)$  è l'insieme degli  $x_k$  a meno degli insiemi di misura nulla;
- $X \stackrel{\text{qc}}{=} \tilde{X}$ , dove  $\text{im}(\tilde{X}) = \{x_k\}_k$  e  $\tilde{X}(\omega) = \begin{cases} X(\omega) & \text{se } X(\omega) \in \{x_k\}_k \\ x_1 & \text{altrimenti} \end{cases}$   
 $x_1$  può essere arbitrario in quanto non influisce sull'immagine.

### Proposizione 7.2: la regola del valore atteso colpisce ancora

Data  $X$  VAR discreta con legge  $P^X \leftrightarrow (\{x_k\}_k, \{p_k\})$ , allora  $\forall h \geq 0$  e  $\forall h \in L^1(P^X)$  si ha che:

$$\int_{\Omega} h(X) d\mathbb{P} = \int_{\mathbb{R}} h d\mathbb{P}^X = \sum_k h(x_k) p_k$$

Si lascia la dimostrazione per esercizio, suggerendo di usare convergenza monotona e dominata.

### Osservazione 7.3

Dalla regola del valore atteso si deducono le formule dell'attesa e della varianza per VAR discrete:

- $\mu = \mathbb{E}[X] = \sum_k x_k p_k$  se  $X \in L^1(\mathbb{P})$  oppure  $X \geq 0$
- $\text{Var}(X) = \sum_k (x_k - \mu)^2 p_k$  se  $X \in L^2(\mathbb{P})$

Si presti attenzione al fatto che per la varianza è richiesta l'appartenenza a  $L^2$ , non solo a  $L^1$ .



---

## 8 Variabili aleatorie reali continue

Gran parte delle grandezze aleatorie del mondo reale non possono essere descritte con una probabilità concentrata in un numero discreto di punti; da qui la necessità di introdurre le *variabili aleatorie continue* che descrivono questo tipo di fenomeni. Saranno forniti criteri per stabilire se una variabile aleatoria è continua o meno (si noti fin da subito che ciò *non* significa semplicemente che la funzione rappresentante la variabile è continua). Era già stato osservato che, nel caso in cui la probabilità di una variabile sia descritta da una densità continua, il singolo punto possiede una probabilità nulla. Sebbene non sia problematico di per sé, questo fatto causa varie complicazioni nel calcolo del valore atteso con le regole introdotte nei capitoli precedenti, che darebbero inevitabilmente risultati pari a 0. Occorre dunque un nuovo strumento matematico per poter integrare funzioni continue rispetto a una probabilità continua: la *misura di Lebesgue*, concetto preso in prestito dalla teoria della misura e che estende la nozione di probabilità.

### 8.1 Misura di Lebesgue

#### Definizione 8.1

La **misura di Lebesgue** è quella funzione  $m : \mathcal{B} \rightarrow [0, +\infty)$  tale che:

1.  $\sigma$ -additività: dati  $B_n \in \mathcal{B} \ \forall n \in \mathbb{N}$ ,  $B_k \cap B_l = \emptyset \ \forall k \neq l$ , allora:

$$m \left( \bigcup_{n=1}^{+\infty} B_n \right) = \sum_{n=1}^{+\infty} m(B_n)$$

2.  $m((a, b]) = b - a$

Si noti che il punto (1) non è sufficiente a qualificare la misura di Lebesgue come misura di probabilità perché non è definita sull'intervallo  $[0, 1]$ . Il punto (2) afferma che ogni intervallo ha misura uguale alla sua lunghezza.

#### Teorema 8.2: unicità della misura di Lebesgue (J-P Th 11.1)

La misura di Lebesgue è *unica*.

#### Dimostrazione

Sia  $a < b$ , e siano  $m^1$  e  $m^2$  due misure di Lebesgue. Costruiamo ora due funzioni  $m_{a,b}^1$  e  $m_{a,b}^2$  tali che:

$$m_{a,b}^k : \mathcal{B} \rightarrow [0, 1], \quad m_{a,b}^k(A) = \frac{m(A \cap (a, b])}{b - a} \quad \text{per } k = 1, 2$$

Le  $m_{a,b}^k$  sono probabilità su  $(\mathbb{R}, \mathcal{B})$ : infatti sono  $\sigma$ -additive (in quanto lo sono le  $m^k$ ) e  $m_{a,b}^k(\mathbb{R}) = 1$ . Esse avranno dunque una funzione di ripartizione ciascuna:

$$F_{a,b}^k(x) = m_{a,b}^k((-\infty, x]) = \begin{cases} 0 & x < a \\ \frac{x-a}{b-a} & a < x < b \\ 1 & x > b \end{cases}$$

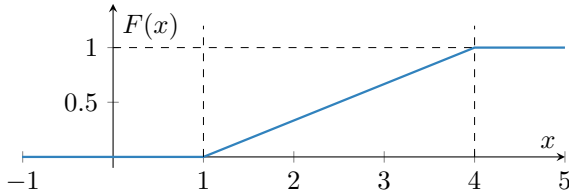


Figura 8.1: funzione di ripartizione  $F_{1,4}^k$

Osservando che nessuna delle due dipende da  $k$ , si conclude che  $F_{a,b}^1 \equiv F_{a,b}^2$  e dunque che  $m_{a,b}^1 = m_{a,b}^2$  perché la funzione di ripartizione caratterizza la probabilità. Dimostriamo ora che questa proprietà vale indipendentemente dagli estremi:

$$\begin{aligned} m^1(A) &= m^1(A \cap \mathbb{R}) \\ &= m^1\left(A \cap \bigcup_{n \in \mathbb{Z}} (n, n+1]\right) && \text{(per definizione di } \mathbb{R} \text{)} \\ &= \sum_{n \in \mathbb{Z}} m^1(A \cap (n, n+1]) && \text{(per } \sigma \text{-additività)} \\ &= \sum_{n \in \mathbb{Z}} m_{n,n+1}^1(A) && \text{(per definizione di } m_{a,b}^k \text{)} \\ &= \sum_{n \in \mathbb{Z}} m_{n,n+1}^2(A) && \text{(le probabilità sono uguali)} \\ &= m^2(A) \quad \forall A \in \mathcal{B} && \text{(ripetendo i passaggi al contrario)} \end{aligned}$$

Questo conclude la dimostrazione. □

**Teorema 8.3** (*J-P Th 11.2*)

La misura di Lebesgue *esiste*.



**Dimostrazione**

Si definisca

$$F_{n,n+1} := \begin{cases} 0 & x \leq n \\ x - n & n < x \leq n + 1 \\ 1 & x > n + 1, \end{cases}$$

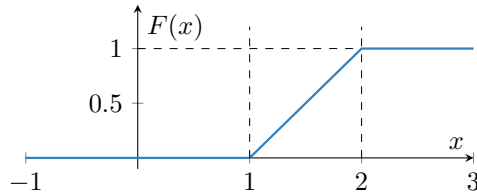


Figura 8.2: funzione di ripartizione  $F_{1,2}$

che è una funzione non decrescente, continua da destra e avente valori limite  $F(-\infty) = 0$  e  $F(+\infty) = 1$ : è dunque una funzione di ripartizione, associata alla quale esiste una probabilità  $m_{n,n+1} : \mathcal{B} \rightarrow [0, 1]$  (di cui però non sappiamo molto altro).

Si definisca ora  $m : \mathcal{B} \rightarrow [0, +\infty)$ ,  $m(A) := \sum_{n \in \mathbb{Z}} m_{n,n+1}(A)$ . Notiamo che  $m$  è  $\sigma$ -additiva, in quanto  $m_{n,n+1}$  lo è e le serie a termini positivi si scambiano; inoltre:

$$m((a, b]) = \sum_{n \in \mathbb{Z}} m_{n,n+1}((a, b]) = \sum_{n \in \mathbb{Z}} [F_{n,n+1}(b) - F_{n,n+1}(a)] = b - a$$

La tesi è dunque dimostrata.  $\square$

**8.1.1 Integrale di Lebesgue**

L'integrale di Lebesgue è compatibile con quello di Riemann, infatti è un'estensione di quest'ultimo. La teoria dell'integrazione per  $X : (\Omega, \mathcal{A}, \mathbb{P}) \rightarrow (\mathbb{R}, \mathcal{B})$  si ripete tale e quale per  $h : (\mathbb{R}, \mathcal{B}, m) \rightarrow (\mathbb{R}, \mathcal{B})$ , a meno di due variazioni:

- $h$  limitata  $\not\Rightarrow h \in L^1(m)$ : la misura di Lebesgue ammette infatti integrali propri anche illimitati;
- $L^2(m) \not\subseteq L^1(m)$ , ulteriore conseguenza della motivazione precedente.

**Definizione 8.4**

L'integrale di Lebesgue è definito come:

$$\int_{\mathbb{R}} h \, dm \quad \forall h \geq 0 \quad \forall h \in L^1(m)$$

**Definizione 8.5**

Si dice che  $h = \tilde{h}$  **quasi ovunque (qo)** se le due funzioni appartengono alla stessa classe di equivalenza  $[h]$ :

$$[h] = \{\tilde{h} : \mathbb{R} \rightarrow \mathbb{R} \text{ boreliano tale che } m(h \neq \tilde{h}) = 0\}$$

**NB:**  $\int_{\mathbb{R}} h \, dm$  dipende solo da  $[h]$ . Questo significa che il valore dell'integrale non cambia se le funzioni sono uguali a meno di un numero discreto (finito o numerabile) di punti.

**Teorema 8.6**

Sia  $h : [a, b] \rightarrow \mathbb{R}$ , limitata e Riemann-integrabile.

Allora la funzione  $h$  è Lebesgue-integrabile e i due integrali hanno lo stesso valore:

$$\int_a^b h(x) \, dx = \int_{[a,b]} h \, dm$$

*Esempio.* Alcune funzioni Lebesgue-integrabili ma non Riemann-integrabili:

- $h(x) = \frac{1}{x^2} \mathbf{1}_{[1,+\infty)}(x)$

Questa funzione ammette integrale di Riemann solo improprio, mentre ammette integrale di Lebesgue proprio, grazie al criterio di convergenza dominata di  $h(x) = \frac{1}{x^2} \mathbf{1}_{[1,k]}(x)$

- $h(x) = \mathbf{1}_{\mathbb{Q} \cap [0,1]}$  (detta *funzione di Dirichlet*)

Questa funzione ammette solo integrale di Lebesgue, perché  $h = 0$  quasi ovunque:

$$m(h \neq 0) = m(\mathbb{Q} \cap [0,1]) = \sum_{q \in \mathbb{Q} \cap [0,1]} m(\{q\}) = 0$$

Poiché l'integrale di Lebesgue e quello di Riemann sono interscambiabili qualora esistano entrambi, d'ora in avanti  $dm$  sarà indicato come  $dx$  e l'integrale di Lebesgue sarà indicato come  $\int_{\mathbb{R}} h \, dx$  per chiarezza rispetto alla variabile di integrazione.

**8.2 Densità continua di probabilità****Definizione 8.7**

Una probabilità  $\mathbb{P}$  su  $(\mathbb{R}, \mathcal{B})$  ammette **densità (continua) di probabilità** se  $\exists f : \mathbb{R} \rightarrow [0, +\infty)$  boreliana tale che:

$$F(x) = \mathbb{P}((-\infty, x]) = \int_{-\infty}^x f \, dt = \int_{\mathbb{R}} f(t) \mathbf{1}_{(-\infty, x]}(t) \, dt = \int_{(-\infty, x]} f \, dt$$

Se  $\mathbb{P} = P^X$  di una VAR  $X$  diciamo che  $X$  è una **variabile aleatoria reale (assolutamente) continua** (abbreviato in VARC) e che  $f$  è densità (continua) di  $X$ .

Quali funzioni di ripartizione ammettono densità? In questo testo saranno enunciati solo un criterio sufficiente e uno necessario, mentre il concetto di *continuità assoluta* verrà introdotto nel corso di analisi reale e funzionale.

**Proposizione 8.8: criterio necessario**

Se  $F(x) = \int_{-\infty}^x f(t) dt$  allora  $F$  è continua.

**Dimostrazione**

Si vuole capire se, per  $n \rightarrow +\infty$ ,  $x_n \rightarrow x$  implichino  $F(x_n) \rightarrow F(x)$ , allora:

$$F(x_n) = \int_{-\infty}^{x_n} f(t) dt = \int_{\mathbb{R}} f(t) \mathbb{1}_{(-\infty, x_n]} dt \xrightarrow{n \rightarrow +\infty} \int_{\mathbb{R}} f(t) \mathbb{1}_{(-\infty, x]} dt = F(x)$$

L'uguaglianza vale quasi ovunque per convergenza dominata, infatti:

$$0 \leq f(t) \mathbb{1}_{(-\infty, x_n]}(t) \leq f(t) \in L^1 \quad \square$$

**Proposizione 8.9: criterio sufficiente**

Data  $F \in C^0 \cap C^1$  a tratti, allora  $F = \int_{-\infty}^x f dt$ ,  $f = F'$  quasi ovunque.

**8.2.1 Caratterizzazione di una densità su  $\mathbb{R}$**

**Teorema 8.10**

1.  $f : \mathbb{R} \rightarrow \mathbb{R}$  densità di probabilità su  $(\mathbb{R}, \mathcal{B}) \iff f$  boreliana,  $f \geq 0$ ,  $\int_{\mathbb{R}} f dx = 1$ ;
2. Nel caso valga (1) si dice che  $f$  determina  $\mathbb{P}$  e vale  $\mathbb{P}(A) = \int_A f dx \forall A \in \mathcal{B}$ , ovvero la classe di equivalenza  $[f]$  caratterizza  $\mathbb{P}$ ;
3. Una probabilità  $\mathbb{P}$  su  $(\mathbb{R}, \mathcal{B})$  con densità  $f$  caratterizza  $[f]$ , ovvero se  $f$  e  $\tilde{f}$  sono densità per  $\mathbb{P}$  allora  $m(f \neq \tilde{f}) = 0$ .

**Dimostrazione**

1. • ( $\implies$ ):  $f$  boreliana,  $f \geq 0$  per definizione:

$$\int_{-\infty}^{+\infty} f dx = \int_{-\infty}^{+\infty} \lim_{t \rightarrow +\infty} \mathbb{1}_{(-\infty, t]}(x) f(x) dx$$

Quindi, per convergenza monotona:

$$= \lim_{t \rightarrow +\infty} \int_{-\infty}^t f(x) dx = \lim_{t \rightarrow +\infty} F(t) = 1$$

- ( $\Leftarrow$ ):  $f$  boreliana,  $f \geq 0$  per definizione,  $\int_{\mathbb{R}} f dx = 1$   
 Definisco  $F(x) = \int_{-\infty}^x f dt \implies F$  continua e decrescente.  
 Visto che l'operatore integrale è positivo e  $f \geq 0$ :

- $\lim_{t \rightarrow -\infty} F(t) = 0$  Per convergenza dominata
- $\lim_{t \rightarrow +\infty} F(t) = 1$  Per convergenza monotona/dominata

Quindi sia  $\mathbb{P}$  la probabilità su  $(\mathbb{R}, \mathcal{B})$  della funzione di ripartizione  $F$  allora  $\mathbb{P}$  ha densità  $f$  per costruzione.

2. Sia  $f$  boreliana,  $f \geq 0$ ,  $\int_{\mathbb{R}} f dx = 1$ ,  $f$  densità di  $\mathbb{P}$   
 Sia inoltre:

$$\tilde{\mathbb{P}}(A) = \int_A f dx \quad \forall A \in \mathcal{B} \implies \tilde{\mathbb{P}} : \mathcal{B} \rightarrow [0, 1], \tilde{\mathbb{P}}(\mathbb{R}) = 1$$

allora  $\tilde{\mathbb{P}}$  è  $\sigma$ -additiva. Quindi  $\tilde{\mathbb{P}}$  è una probabilità su  $(\mathbb{R}, \mathcal{B})$ , con  $\tilde{F} = F$ ,  $\tilde{\mathbb{P}} = \mathbb{P}$ , ovvero:

$$\mathbb{P}(A) = \tilde{\mathbb{P}}(A) = \int_A f dx$$

Per giustificare il passaggio in cui si afferma che  $\tilde{\mathbb{P}}$  è  $\sigma$ -additiva si può mostrare che dati  $B_n$  disgiunti, ovvero:

$$B_n \in \mathcal{B}, B_k \cap B_l = \emptyset \quad \forall k \neq l$$

allora:

$$\tilde{\mathbb{P}}\left(\bigcup_n B_n\right) = \int_{\bigcup_n B_n} f(x) dx = \int_{\mathbb{R}} f(x) \mathbf{1}_{\bigcup_n B_n}(x) dx$$

Essendo i  $B_n$  disgiunti la funzione indicatrice può essere spezzata in una sommatoria:

$$= \int_{\mathbb{R}} f(x) \sum_n \mathbf{1}_{B_n}(x) dx = \sum_n \int_{\mathbb{R}} f(x) \mathbf{1}_{B_n}(x) dx = \sum_n \tilde{\mathbb{P}}(B_n)$$

Lo scambio di integrale e sommatoria è possibile grazie al fatto che la serie è completamente a termini positivi.

Ricapitolando  $f$  determina  $\mathbb{P}$  e  $[f]$  caratterizza  $\mathbb{P}$ , ovvero se  $\mathbb{P}_1$  ha densità  $f_1$ ,  $\mathbb{P}_2$  ha densità  $f_2$  e:

$$[f_1] = [f_2] \implies \mathbb{P}_1 = \mathbb{P}_2$$

3. Sia  $\mathbb{P}$  una probabilità su  $(\mathbb{R}, \mathcal{B})$ ,  $f$  e  $\tilde{f}$  densità di  $\mathbb{P}$ ; si vuole mostrare che questo implica  $m(f \neq \tilde{f}) = 0$ . Si dimostra quindi che  $m(f < \tilde{f}) = 0$  e  $m(f > \tilde{f}) = 0$ .

Dato  $\varepsilon > 0$  sia:

$$A = \left\{ x : f(x) + \varepsilon \leq \tilde{f}(x) \right\} \subseteq \left( f < \tilde{f} \right) \implies A \uparrow \left( f < \tilde{f} \right) \text{ per } \varepsilon \downarrow 0$$

allora, passando alle probabilità:

$$\mathbb{P}(A) + \varepsilon m(A) = \int_A (f + \varepsilon) dx \leq \int_A \tilde{f}(A) = \mathbb{P}(A)$$

Essendo  $\varepsilon$  strettamente maggiore di zero:

$$m(A) = 0 \quad \forall \varepsilon \implies m(f < \tilde{f}) = 0$$

Ripetendo lo stesso procedimento per  $f > \tilde{f}$  si ottiene che:

$$m(f \neq \tilde{f}) = m(f < \tilde{f}) + m(f > \tilde{f}) = 0 \quad \square$$

### **Teorema 8.11: il ritorno della regola del valore atteso**

Dati  $(\Omega, \mathcal{A}, \mathbb{P})$  spazio di probabilità,  $X$  VAR continua di legge  $P^X$  e densità  $f$  e  $h : \mathbb{R} \rightarrow \mathbb{R}$  boreliana, allora:

1.  $h \in L^1(P^X) \iff hf \in L^1(m)$
2.  $h$  ammette integrale rispetto a  $P^X \iff hf$  ammette integrale rispetto a  $m$
3. Sotto le ipotesi di (1) e (2):

$$\mathbb{E}[h(x)] = \int_{\Omega} h(x) d\mathbb{P} = \int_{\mathbb{R}} h dP^X = \int_{\mathbb{R}} hf dx$$

L'ultima uguaglianza è la tesi fondamentale del teorema.<sup>24</sup>

### **Dimostrazione**

La dimostrazione è molto simile a quella del teorema 6.27, è sufficiente considerare che:

- $B \in \mathcal{B}$ ,  $h = \mathbf{1}_B$  per il teorema 8.10, quindi:

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

---

<sup>24</sup>Come al solito

- $h$  è semplice per linearità;
- quando  $h \geq 0$  si sfrutta convergenza monotona;
- per  $h$  qualsiasi si può impiegare la scomposizione in  $h_+$  e  $h_-$ . □

**NB:** Data una variabile continua  $X$  non è detto che  $Y = h(X)$  lo sia altrettanto. Un esempio può essere una semplice bernoulliana:  $Y = \mathbf{1}_B(X)$ .  $Y$  può infatti assumere solo i due valori 1 e 0, quindi non può essere continua a meno che non sia costante.

### 8.3 Considerazioni pratiche

Dato uno spazio di probabilità  $(\Omega, \mathcal{A}, \mathbb{P})$  e una VAR  $X : \mathbb{R} \rightarrow \mathbb{R}$ ,  $X$  è (**assolutamente**) **continua**:

$$\iff P^X \text{ ammette densità } f, \quad P^X(B) = \int_B f(x) dx \quad \forall B \in \mathcal{B}$$

$$\iff F_X(t) = \int_{-\infty}^{+\infty} f dx \quad \forall t \in \mathbb{R}$$

$$\implies F_X \in C^0, \text{ ovvero } P^X(\{x\}) = \mathbb{P}(X = x) = F_X(x) - F_X(x_-) = 0 \text{ (in quanto } F_X(x_-) = F_X(x_+) = F_X(x))$$

$$\iff F_X \in C^0 \cap C^1 \text{ a tratti}$$

$$\text{In quest'ultimo caso } f = \begin{cases} F'_X(x) & \text{dove } F_X \text{ è derivabile} \\ \text{valore arbitrario} & \text{altrimenti} \end{cases}$$

Se  $h$  è boreliana, e se  $h : \mathbb{R} \rightarrow [0, +\infty]$  oppure  $h : \mathbb{R} \rightarrow \mathbb{R}$  ( $h \in L^1(\mathbb{P})$ ), allora:

$$\mathbb{E}[h(X)] = \int_{\Omega} h(X(\omega)) p(d\omega) = \int_{\mathbb{R}} h(x) P^X(dx) = \int_{\mathbb{R}} h(x) f(x) dx$$

Inoltre:

$$f \text{ densità} \iff \begin{cases} f \text{ boreliana} \\ f \geq 0 \\ \int_{-\infty}^{+\infty} f(x) dx = 1 \end{cases}$$

---

## 9 Variabili aleatorie indipendenti

Dal concetto di indipendenza tra *eventi* discende l'*indipendenza tra variabili aleatorie*, che indica la condizione in cui gli eventi modellizzati dalle variabili non si influenzano l'un l'altro. Enunceremo alcuni criteri per stabilire l'indipendenza, alcuni dei quali inizieranno a mostrare l'utilità del valore atteso al di là dell'essere un semplice indicatore a sé stante; verrà inoltre illustrata la possibilità di costruire variabili indipendenti a partire da variabili qualsiasi. Grazie all'indipendenza si potrà iniziare a estendere la teoria precedente a *prodotti cartesiani* tra domini, a cui corrispondono gli eventi nelle  $\sigma$ -algebre prodotto e le relative probabilità, le *leggi congiunte*, costruite a partire dalle *leggi marginali* mediante l'importantissimo *teorema di Fubini-Tonelli*. Potremo presto iniziare a trattare i *vettori aleatori*, variabili aleatorie multidimensionali.

### 9.1 Indipendenza di variabili aleatorie

#### Definizione 9.1

Siano  $(\Omega, \mathcal{A}, \mathbb{P})$  spazio di probabilità,  $(E, \mathcal{E})$ ,  $(F, \mathcal{F})$  spazi misurabili, e  $X : \Omega \rightarrow E$  e  $Y : \Omega \rightarrow F$  VA.

$X$  e  $Y$  si dicono **VA indipendenti** ( $X \perp\!\!\!\perp Y$ ) se:

$$\mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A) \mathbb{P}(Y \in B) \quad \forall A \in \mathcal{E} \quad \forall B \in \mathcal{F}$$

Ciò significa che ogni evento  $A$  che riguarda  $X$  è indipendente da ogni evento  $B$  che riguarda  $Y$ . In altre parole, le *previsioni* di  $X$  e  $Y$  *non si influenzano* l'una con l'altra.

**NB:** Lo stesso  $\omega \in \Omega$  determina il valore sia di  $X$  che di  $Y$  perché  $\Omega$  è il dominio comune delle due VA.

*Esempio.* Date due monete truccate con probabilità di ottenere testa  $0 < p < 1$  siano  $\Omega = \{c, t\}^2$ ,  $\omega = (\omega_1, \omega_2)$ ,  $\mathcal{A} = 2^\Omega$ . Si considerino ora gli eventi:

- $C_1 =$  croce sulla prima moneta  $= \{(c, c), (c, t)\} = \{\omega \in \Omega : \omega_1 = c\}$
- $T_1 =$  testa sulla prima moneta  $= C_1^C = \{\omega \in \Omega : \omega_1 = t\}$
- $C_2 =$  croce sulla seconda moneta  $= \{\omega \in \Omega : \omega_2 = c\}$
- $T_2 = C_2^C$

Si sceglie  $\mathbb{P}$  tale che  $\mathbb{P}(T_1) = \mathbb{P}(T_2) = p$  e  $T_1 \perp\!\!\!\perp T_2$ . Di conseguenza,  $T_1^C \perp\!\!\!\perp T_2^C$ .

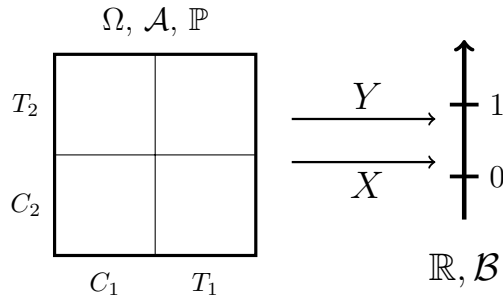


Figura 9.1: dominio e codominio della VAR

Passando dal linguaggio degli eventi al linguaggio delle VA, si introducono due VA:

$$\begin{aligned}
 X &= \begin{cases} 1 & \text{se sulla prima moneta esce testa} \\ 0 & \text{se sulla prima moneta esce croce} \end{cases} \\
 \implies X : \Omega \rightarrow \mathbb{R} & \quad (X = 1) = T_1 \quad (X = 0) = C_1 \\
 Y &= \begin{cases} 1 & \text{se sulla seconda moneta esce testa} \\ 0 & \text{se sulla seconda moneta esce croce} \end{cases} \\
 \implies Y : \Omega \rightarrow \mathbb{R} & \quad (Y = 1) = T_2 \quad (Y = 0) = C_2
 \end{aligned}$$

Si può affermare che questo implica  $X \perp\!\!\!\perp Y$ ?

Serve verificare che:

$$\mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A) \mathbb{P}(Y \in B) \quad \forall A, B \in \mathcal{B}$$

L'uguaglianza è vera perché:

- se  $A$  e  $B$  contengono un solo valore, allora sono uguali a  $T_1, T_2$  o i loro complementari ed essi sono indipendenti per costruzione;
- se  $A$  o  $B$  sono l'evento impossibile  $\emptyset$ , allora entrambi i membri sono nulli;
- se  $A$  o  $B$  sono l'evento certo  $\Omega$ , allora entrambi i membri sono uguali a 1.

Introduciamo ora le VA  $Z = X + Y$  (numero di teste su due lanci),  $W = X - Y$  e  $V = |X - Y|$  (0 se i due risultati sono uguali, 1 altrimenti).

$Z$  e  $W$  non sono indipendenti? Per dimostrarlo ci servono due eventi le cui probabilità non si fattorizzano. Ad esempio, verifichiamo che  $\mathbb{P}(Z = 2, W = 1) \neq \mathbb{P}(Z = 2)\mathbb{P}(W = 1)$ :

- $\mathbb{P}(Z = 2, W = 1) = \mathbb{P}(\emptyset)$ , in quanto non abbiamo un'intersezione in cui si verificano entrambi gli eventi. Quindi  $\mathbb{P}(\emptyset) = 0$ ;



- $\mathbb{P}(Z = 2) = \mathbb{P}(X = 1, Y = 1) = p^2$ ;
- $\mathbb{P}(W = 1) = \mathbb{P}(X = 1, Y = 0) = \mathbb{P}(X = 1)\mathbb{P}(Y = 0) = p(1 - p)$ .

Ma  $p^3(1 - p) \neq 0$  per  $p$  non degeneri. Allora  $\mathbb{P}(Z = 2, W = 1) \neq \mathbb{P}(Z = 2)\mathbb{P}(W = 1)$  e  $Z \not\perp W$ .

Si può dimostrare allo stesso modo che  $Z \not\perp V$ ; anzi,  $V = h(Z) = Z(2 - Z)$ . Si noti che  $X = f(Y)$  è l'opposto concettuale di  $X \perp Y$ .

## 9.2 Condizioni equivalenti

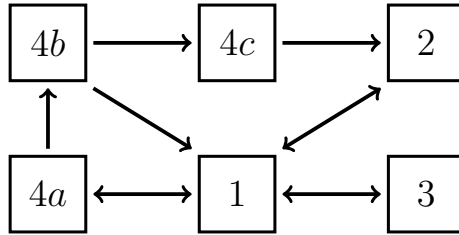
### Teorema 9.2 (J-P Th 10.1)

Siano  $(\Omega, \mathcal{A}, \mathbb{P})$  spazio di probabilità,  $(E, \mathcal{E})$  e  $(F, \mathcal{F})$  spazi misurabili. Siano inoltre le VA  $X : \Omega \rightarrow E$  e  $Y : \Omega \rightarrow F$ . Allora le seguenti affermazioni sono equivalenti:

1.  $X \perp Y$ ;
2.  $\mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A)\mathbb{P}(Y \in B) \quad \forall A \in \mathcal{C} \quad \forall B \in \mathcal{D}$   
con  $\mathcal{C}$  e  $\mathcal{D}$  chiusi per intersezioni finite,  $\mathcal{E} = \sigma(\mathcal{C})$  e  $\mathcal{F} = \sigma(\mathcal{D})$ ;
3.  $h(X) \perp g(Y) \quad \forall h : (E, \mathcal{E}) \rightarrow (E', \mathcal{E}')$  misurabile e  $\forall g : (F, \mathcal{F}) \rightarrow (F', \mathcal{F}')$  misurabile;  
(se sono state definite bene,  $h(X)$  non dà informazioni su  $Y$  e  $g(Y)$  non dà informazioni su  $X$ )
4.  $\mathbb{E}[h(X)g(Y)] = \mathbb{E}[h(X)]\mathbb{E}[g(Y)]$ 
  - (a)  $\forall h : (E, \mathcal{E}) \rightarrow (\mathbb{R}, \mathcal{B})$  misurabile e positiva,  
 $\forall g : (F, \mathcal{F}) \rightarrow (\mathbb{R}, \mathcal{B})$  misurabile e positiva;  
(gli insiemi sono ben specifici e non possono essere scelti a caso)
  - (b)  $\forall h : (E, \mathcal{E}) \rightarrow (\mathbb{R}, \mathcal{B})$  misurabile e limitata,  
 $\forall g : (F, \mathcal{F}) \rightarrow (\mathbb{R}, \mathcal{B})$  misurabile e limitata;  
(le condizioni su  $h$  e  $g$  servono per l'esistenza dei vari integrali:  $h$  dovrebbe essere  $L^1$  della legge di  $X$ , ma visto che non ne sappiamo le specifiche usiamo le condizioni per accertarci che questo sia verificato)
  - (c) se  $E, F$  sono spazi metrici e  $\mathcal{E}, \mathcal{F}$  le loro  $\sigma$ -algebre di Borel  
 $\forall h : (E, \mathcal{E}) \rightarrow (\mathbb{R}, \mathcal{B})$  continua e positiva,  
 $\forall g : (F, \mathcal{F}) \rightarrow (\mathbb{R}, \mathcal{B})$  continua e positiva;

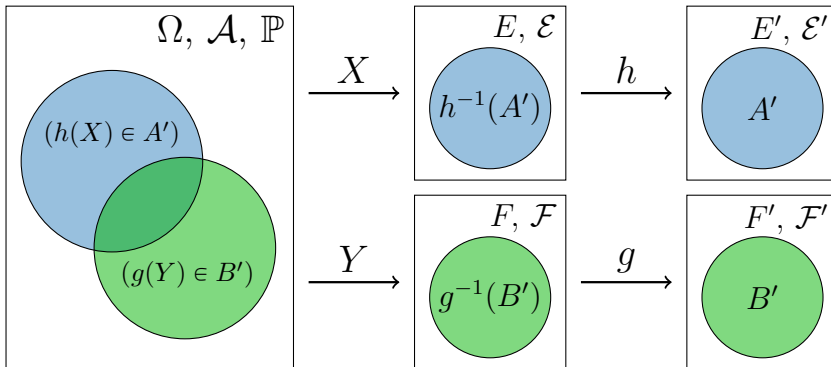
**Dimostrazione**

La dimostrazione si svolgerà così:



in modo che tutte le affermazioni si coimplichino.

- (3)  $\implies$  (1): ovvio, presi  $h(X) = X$  e  $g(Y) = Y$ .
- (1)  $\implies$  (3): Si inizia rappresentando immagini e controimmagini nei domini e codomini:



Si può osservare che grazie alle controimmagini ci si può ricondurre a  $\mathcal{E}$  e  $\mathcal{F}$  dove  $X$  e  $Y$  sono indipendenti per mostrare la tesi:

$$\begin{aligned}
 & \mathbb{P}(h(X) \in A', g(Y) \in B') \\
 &= \mathbb{P}(X^{-1}(h^{-1}(A')), Y^{-1}(g^{-1}(B'))) \\
 &= \mathbb{P}(X \in \underbrace{h^{-1}(A')}_{\in \mathcal{E}}, Y \in \underbrace{g^{-1}(B')}_{\in \mathcal{F}}) \quad (\text{per la misurabilità di } h \text{ e } g) \\
 &= \mathbb{P}(X \in h^{-1}(A')) \cdot \mathbb{P}(Y \in g^{-1}(B')) \\
 &= \mathbb{P}(h(X) \in A') \cdot \mathbb{P}(g(Y) \in B') \quad \forall A' \in \mathcal{E}' \forall B' \in \mathcal{F}'
 \end{aligned}$$

- (4a)  $\implies$  (1): scegliamo  $h = \mathbb{1}_A$  e  $g = \mathbb{1}_B$ , con  $A \in \mathcal{E}$  e  $B \in \mathcal{F}$ :

$$\begin{aligned}\mathbb{E}[h(X)g(Y)] &= \mathbb{E}[h(X)] \cdot \mathbb{E}[g(Y)] \\ \mathbb{E}[\mathbb{1}_A(X)\mathbb{1}_B(Y)] &= \mathbb{E}[\mathbb{1}_A(X)] \cdot \mathbb{E}[\mathbb{1}_B(Y)] \\ \mathbb{P}(X \in A, Y \in B) &= \mathbb{P}(X \in A) \cdot \mathbb{P}(Y \in B)\end{aligned}$$

- (1)  $\implies$  (4a):

Per ipotesi dati  $A \in \mathcal{E}$   $\forall g = \mathbb{1}_B$  e  $B \in \mathcal{F}$ :

$$\mathbb{E}[h(X)g(Y)] = \mathbb{E}[h(X)]\mathbb{E}[g(Y)] \quad \forall h = \mathbb{1}_A$$

Dunque, per linearità del valore atteso,  $\mathbb{E}[h(X)g(Y)] = \mathbb{E}[h(X)] \mathbb{E}[g(Y)] \quad \forall h, g$  semplici e positive.

Infine, per convergenza monotona, vale  $\mathbb{E}[h(X)g(Y)] = \mathbb{E}[h(X)]\mathbb{E}[g(Y)] \quad \forall h, g$  misurabili e positive.

- (4a)  $\implies$  (4b): Vale  $\mathbb{E}[h(X)g(Y)] = \mathbb{E}[h(X)]\mathbb{E}[g(Y)] \quad \forall h, g$  misurabili e limitate poiché vale  $\forall h, g$  misurabili e positive, usando la linearità del valore atteso.
- (4b)  $\implies$  (1): ragionamento analogo a (4a)  $\implies$  (1).
- (4b)  $\implies$  (4c): ovvio, le funzioni continue e limitate sono un caso particolare delle funzioni misurabili e limitate.
- (1)  $\implies$  (2): ovvio, per definizione.
- (2)  $\implies$  (1)  
Per ipotesi:

$$\mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A) \mathbb{P}(Y \in B) \quad \forall A \in \mathcal{C}, \forall B \in \mathcal{D}$$

Fissato un arbitrario  $B \in \mathcal{D}$ , si costruisca quindi la classe:

$$\tilde{\mathcal{C}}_B := \{A \in \mathcal{E} : \mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A) \mathbb{P}(Y \in B)\}$$

Osserviamo che  $\mathcal{C} \subseteq \tilde{\mathcal{C}}_B \subseteq \mathcal{E}$ ; inoltre è facilmente verificabile che  $\tilde{\mathcal{C}}_B$  è chiuso per limiti crescenti e per differenze.

Allora per le classi monotone (teorema 3.4)  $\tilde{\mathcal{C}}_B \supseteq \sigma(\mathcal{C}) = \mathcal{E}$  e dunque  $\tilde{\mathcal{C}}_B = \mathcal{E}$ . Similmente si dimostra l'estensione della proprietà da  $\mathcal{D}$  a  $\mathcal{F}$ <sup>25</sup> e la tesi è dimostrata.

---

<sup>25</sup>“Doppia invocazione delle classi monotone”, come disse Gregoratti; è una delle frasi più belle che abbia mai sentito.

- (4c)  $\implies$  (2)<sup>26</sup>

Si ha per ipotesi che, per ogni  $h, g$  continue e limitate,  $\mathbb{E}[h(X)g(Y)] = \mathbb{E}[h(X)]\mathbb{E}[g(Y)]$ . Bisogna mostrare che valga la (2) con  $\mathcal{C}$  e  $\mathcal{D}$  collezioni di insiemi chiusi (poiché gli insiemi chiusi sono complementari degli aperti, sia gli uni che gli altri fanno parte della stessa  $\sigma$ -algebra), le quali generano le rispettive  $\sigma$ -algebre (ovvero  $\mathcal{E} = \sigma(\mathcal{C})$  e  $\mathcal{F} = \sigma(\mathcal{D})$ ).

Sia  $d(x, A) := \inf\{d(x, y) : y \in A\}$  la distanza di un elemento  $x$  dall'insieme  $A$ . Mostriamo ora che  $h(x) = d(x, A)$  è una funzione continua.

Siano due punti  $x$  e  $x_0$ ; bisogna dimostrare che se  $x \rightarrow x_0$ , allora  $d(x, A) \rightarrow d(x_0, A) \forall A$ . Fissato un  $A$  arbitrario vale:

$$d(x, A) - d(x_0, A) := \inf\{d(x, y) : y \in A\} - \inf\{d(x_0, y) : y \in A\}$$

Chiameremo, rispettivamente,  $y$  e  $y_0$  i due punti la cui distanza è pari agli addendi di cui sopra. Allora:

$$\begin{aligned} d(x, A) - d(x_0, A) &= d(x, y) - d(x_0, y_0) \leq d(x, y_0) - d(x_0, y_0) \\ &\leq d(x, x_0) + d(x_0, y_0) - d(x_0, y_0) = d(x, x_0) \rightarrow 0 \end{aligned}$$

La distanza definita è dunque una funzione continua.

Si definisca ora  $h_n(x) := \min\{1, n \cdot d(x, A)\}$  (con  $A$  chiuso) che, per ogni  $n$ , è continua e tale che  $0 \leq h_n \leq 1$ , e successivamente  $\tilde{h}_n(x) := 1 - h_n(x)$ , ancora continua e limitata per ogni  $n$ . Notiamo anche che:

$$\tilde{h}_n \xrightarrow{n \rightarrow +\infty} \begin{cases} 1 & x \in A \\ 0 & x \notin A \end{cases} = \mathbb{1}_A(x) \quad \forall x \in E \text{ (ovvero il dominio di } \tilde{h}_n)$$

Analogamente si definisce  $\tilde{g}_n$  continua e limitata, e si ottiene che  $\tilde{g}_n(y) \xrightarrow{n} \mathbb{1}_B(y) \forall y \in F$ .

Entrambe le successioni di funzioni rientrano nelle ipotesi di (4c):

$$\mathbb{E}[\tilde{h}_n(x)\tilde{g}_n(x)] = \mathbb{E}[\tilde{h}_n(x)] \mathbb{E}[\tilde{g}_n(x)]$$

Prendendo i limiti per  $n \rightarrow +\infty$  di entrambi i membri, i tre valori attesi, per ogni  $x$ , convergono per convergenza monotona, infatti  $\mathbb{E}[\mathbb{1}_A(x)\mathbb{1}_B(x)] = \mathbb{E}[\mathbb{1}_A(x)] \mathbb{E}[\mathbb{1}_B(x)]$ , ovvero:

$$\mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A)\mathbb{P}(Y \in B) \quad \forall A, B \text{ chiusi}$$

Osservando che gli insiemi chiusi generano le rispettive classi, la tesi risulta quindi dimostrata.  $\square$

---

<sup>26</sup>Questa proprietà dimostra l'estensione di alcune proprietà dalle funzioni continue a quelle limitate, estensione che però è possibile solo in casi particolari come questo e "non vale sempre nella vita".

**Proposizione 9.3**

Siano le VA  $X, X', Y, Y'$ , tali che  $X \stackrel{\text{qc}}{=} X'$  e  $Y \stackrel{\text{qc}}{=} Y'$ . Allora:

$$X \perp\!\!\!\perp Y \iff X' \perp\!\!\!\perp Y'$$

**Dimostrazione**

Siano  $C = (X = X')$  e  $D = (Y = Y')$ . Allora  $C \in \mathcal{A}$  e  $\mathbb{P}(C) = 1$ , analogamente  $D \in \mathcal{A}$  e  $\mathbb{P}(D) = 1$ . Intersecando con gli eventi  $C$  e  $D$  non si modifica la probabilità, quindi:

$$\begin{aligned} \mathbb{P}(X' \in A, Y' \in B) &= \mathbb{P}(X' \in A, Y' \in B, C, D) \\ &= \mathbb{P}(X \in A, Y \in B, C, D) \\ &= \mathbb{P}(X \in A, Y \in B) \\ &= \mathbb{P}(X \in A) \cdot \mathbb{P}(Y \in B) \\ &= \mathbb{P}(X \in A, C) \cdot \mathbb{P}(Y \in B, D) \\ &= \mathbb{P}(X' \in A, C) \cdot \mathbb{P}(Y' \in B, D) \\ &= \mathbb{P}(X' \in A) \cdot \mathbb{P}(Y' \in B) \end{aligned} \quad \square$$

Mostreremo alcuni esempi per il caso  $E = F = \mathbb{R}$ .

Prima però bisogna comprendere cosa accade a misurabilità, probabilità e integrali considerando i prodotti cartesiani invece che le singole variabili:

- $X : \Omega \rightarrow E, Y : \Omega \rightarrow F \implies (X, Y) : \Omega \rightarrow E \times F$
- $X_1 : \Omega_1 \rightarrow E, Y_2 : \Omega_2 \rightarrow F \implies \begin{cases} X(\omega) = X_1(\omega_1), Y(\omega) = Y_2(\omega_2) \\ X : \Omega \rightarrow E, Y : \Omega \rightarrow F \end{cases}$

**9.3 Prodotto cartesiano**

Siano  $(\Omega_1, \mathcal{A}_1)$  e  $(\Omega_2, \mathcal{A}_2)$  spazi misurabili. Prendendo  $\Omega : \Omega_1 \times \Omega_2$ , qual è la  $\sigma$ -algebra “naturale” su questo nuovo  $\Omega$ ?

Consideriamo la seguente collezione:

$$\mathcal{A}_1 \times \mathcal{A}_2 := \{A \subseteq \Omega_1 \times \Omega_2, A = A_1 \times A_2 : A_1 \in \mathcal{A}_1, A_2 \in \mathcal{A}_2\}$$

I suoi elementi  $A$  sono rettangoli misurabili, ovvero il prodotto cartesiano dei due intervalli  $A_1$  e  $A_2$ , i quali sono dunque le “proiezioni” di  $A$  sui domini  $\Omega_1$  e  $\Omega_2$ .

Osservando la figura 9.2 si nota come già nel caso semplice  $\Omega_1 = \Omega_2 = \mathbb{R}$  unendo due elementi non si ottenga un terzo rettangolo, bensì un generico sottoinsieme di  $\Omega_1 \times \Omega_2$ : la collezione  $\mathcal{A}_1 \times \mathcal{A}_2$  non è chiusa per unioni numerabili. Per ottenere una  $\sigma$ -algebra bisogna quindi costruire uno spazio che rispetti le proprietà richieste alle  $\sigma$ -algebre, ovvero, tra le altre cose, che contenga anche le unioni numerabili di rettangoli. Ciò ci porta a formulare la definizione 9.4.

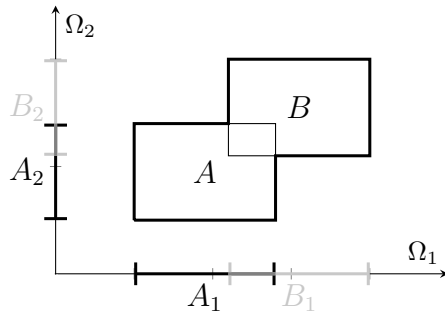


Figura 9.2: unione di due rettangoli sottoinsiemi di  $\Omega_1 \times \Omega_2$ , con  $\Omega_1 = \Omega_2 = \mathbb{R}$

**Definizione 9.4**

Si dice  **$\sigma$ -algebra prodotto** (o **prodotto cerchiato** o **prodotto tensore**) tra due  $\sigma$ -algebre  $\mathcal{A}_1$  e  $\mathcal{A}_2$  la  $\sigma$ -algebra generata dai rettangoli di  $\mathcal{A}_1$  e  $\mathcal{A}_2$ , ovvero la collezione così definita:

$$\mathcal{A}_1 \otimes \mathcal{A}_2 := \sigma(\mathcal{A}_1 \times \mathcal{A}_2) = \sigma(\{A \subseteq \Omega_1 \times \Omega_2, A = A_1 \times A_2 : A_1 \in \mathcal{A}_1, A_2 \in \mathcal{A}_2\})$$

**Osservazione 9.5**

$X : (\Omega, \mathcal{A}) \rightarrow (E, \mathcal{E})$  e  $Y : (\Omega, \mathcal{A}) \rightarrow (F, \mathcal{F})$  misurabili  $\iff (X, Y) : (\Omega, \mathcal{A}) \rightarrow (E \times F, \mathcal{E} \otimes \mathcal{F})$  misurabile.

*Esempio.* Consideriamo  $\Omega = \mathbb{R} \times \mathbb{R} = \mathbb{R}^2$ . Ci possono essere due  $\sigma$ -algebre “naturali” per questo spazio:

- $\mathcal{B}^2 = \sigma(\text{aperti in } \mathbb{R}^2)$
- $\mathcal{B} \otimes \mathcal{B}$

Si può dimostrare che le due  $\sigma$ -algebre considerate coincidono, e vale:

$$\mathcal{B}^2 = \sigma((-\infty, t_1] \times (-\infty, t_2] : t_1, t_2 \in \mathbb{Q})$$

In due dimensioni questi *quadranti mobili* possono essere rappresentati come dei rettangoli che ricoprono tutto  $\mathbb{R}^2$ .

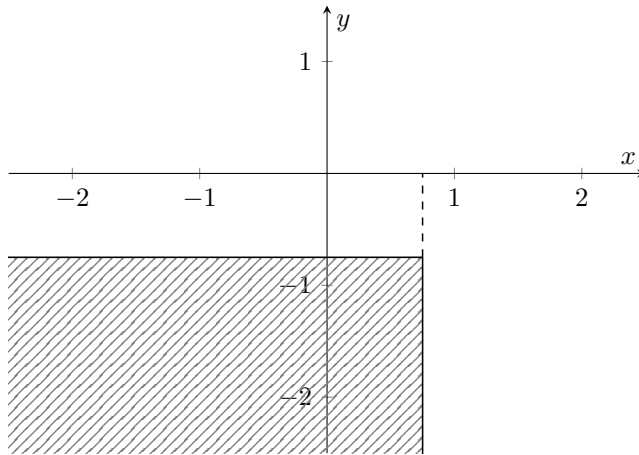


Figura 9.3: quadranti  $\tau_n$  generatori di  $\mathcal{B}^2$

Si può inoltre generalizzare al caso  $n$ -dimensionale:

$$\mathcal{B}^n = \sigma \left( \bigtimes_{k=1}^n (-\infty, t_k] : t_k \in \mathbb{Q} \ \forall k \right)$$

In questo caso sono detti *iper-quadranti di sud-ovest*, già introdotti nella definizione 5.9.

**Teorema 9.6: misurabilità delle restrizioni** (*J-P Th 10.2*)

Sia  $X : (\Omega_1 \times \Omega_2, \mathcal{A}_1 \otimes \mathcal{A}_2) \mapsto (\mathbb{R}, \mathcal{B})$  misurabile. Allora:

- $X(\cdot, \omega_2) : (\Omega_1, \mathcal{A}_1) \rightarrow (\mathbb{R}, \mathcal{B})$  è misurabile  $\forall \omega_2 \in \Omega_2$  fissato;
- $X(\omega_1, \cdot) : (\Omega_2, \mathcal{A}_2) \rightarrow (\mathbb{R}, \mathcal{B})$  è misurabile  $\forall \omega_1 \in \Omega_1$  fissato.

In generale, il viceversa è falso. Le restrizioni sono misurabili rispetto alle loro  $\sigma$ -algebre.

**NB:** Se  $X = \mathbb{1}_A$  e  $A \in \mathcal{A}_1 \otimes \mathcal{A}_2$ , allora:

- $A_{\omega_2} = \{\omega_1 \in \Omega_1 : (\omega_1, \omega_2) \in A\} \in \mathcal{A}_1$ ;
- $A_{\omega_1} = \{\omega_2 \in \Omega_2 : (\omega_1, \omega_2) \in A\} \in \mathcal{A}_2$ .

**Dimostrazione**

a) Dapprima sarà dimostrata la tesi per  $X = \mathbb{1}_A$ , con  $A \in \mathcal{A}_1 \otimes \mathcal{A}_2$ .

Sia  $\mathcal{C} := \{A \in \mathcal{A}_1 \otimes \mathcal{A}_2 : \mathbb{1}_A(\omega_1, \cdot) \text{ è } \mathcal{A}_2\text{-misurabile } \forall \omega_1\}$  e si noti subito che

$\mathcal{C} \subseteq \mathcal{A}_1 \otimes \mathcal{A}_2$ .

Allora  $\mathcal{A}_1 \times \mathcal{A}_2 \subseteq \mathcal{C}$  perché  $\mathbb{1}_{\mathcal{A}_1 \times \mathcal{A}_2}(\omega_1, \cdot) = \mathbb{1}_{\mathcal{A}_1}(\omega_1)\mathbb{1}_{\mathcal{A}_2}(\cdot)$ , la quale ha  $A_2 \in \mathcal{A}_2$  ed è pertanto misurabile. Esaminando ora le proprietà di  $\mathcal{C}$ , si scopre che  $\Omega_1 \times \Omega_2 \in \mathcal{C}$  e che è chiusa per unione numerabile: infatti, presa una successione  $A_k \in \mathcal{C}$  con  $k \in \mathcal{N}$ , sappiamo che  $\bigcup_k A_k \in \mathcal{A}_1 \otimes \mathcal{A}_2$  se e solo se  $\mathbb{1}_{\bigcup_k A_k}(\omega_1, \cdot)$  è  $\mathcal{A}_2$ -misurabile; ma  $\mathbb{1}_{\bigcup_k A_k}(\omega_1, \cdot) = \mathbb{1}_{\bigcup_k A_k, \omega_1}(\cdot)$  (dove  $\omega_1$  è un parametro fissato) è effettivamente  $\mathcal{A}_2$ -misurabile perché  $A_{k, \omega_1} \in \mathcal{A}_2 \forall k \forall \omega_1$ . Similmente si dimostra la chiusura per intersezione numerabile.  $\mathcal{C}$  è dunque una  $\sigma$ -algebra; in particolare, vista la definizione di  $\mathcal{C}$ , si deduce che  $\mathcal{C} \equiv \mathcal{A}_1 \otimes \mathcal{A}_2$ .

- b) La tesi vale per  $X$  semplice per linearità: le VA di questo tipo sono infatti combinazioni lineari di indicatrici.
- c) La tesi vale per  $X$  misurabile e positiva poiché si può scrivere  $X = \lim_n X_n$  con  $X_n$  VA semplici e positive.
- d) La tesi vale per  $X$  generica perché  $X = X_+ - X_-$ , che sono entrambi addendi positivi. □

### 9.4 Il teorema di Fubini-Tonelli

Consideriamo gli spazi di probabilità  $(\Omega_1, \mathcal{A}_1, \mathbb{P}_1)$  e  $(\Omega_2, \mathcal{A}_2, \mathbb{P}_2)$ . Siano  $\Omega = \Omega_1 \times \Omega_2$  e  $\mathcal{A} = \mathcal{A}_1 \otimes \mathcal{A}_2$ .

Non esiste una  $\mathbb{P}$  “naturale” su  $\mathcal{A}_1 \otimes \mathcal{A}_2$ : le probabilità dei fattori non determinano un’unica probabilità.

Possiamo però chiederci se esista una  $\mathbb{P}$  su  $\mathcal{A}_1 \otimes \mathcal{A}_2$  che renda gli eventi in  $\mathcal{A}_1$  indipendenti dagli eventi in  $\mathcal{A}_2$  e assegni loro le probabilità  $\mathbb{P}_1$  e  $\mathbb{P}_2$ .

$A_1 \in \mathcal{A}_1$  corrisponde a  $A_1 \times \Omega_2 \in \mathcal{A} = \mathcal{A}_1 \otimes \mathcal{A}_2$ . Analogamente,  $A_2 \in \mathcal{A}_2$  corrisponde a  $\Omega_1 \times A_2 \in \mathcal{A} = \mathcal{A}_1 \otimes \mathcal{A}_2$ . Cerchiamo dunque  $\mathbb{P}$  tale che:

$$\underbrace{\mathbb{P}(A_1 \times \Omega_2, \Omega_1 \times A_2)}_{\mathbb{P}(A_1 \times A_2)} = \underbrace{\mathbb{P}(A_1 \times \Omega_2) \mathbb{P}(\Omega_1 \times A_2)}_{\mathbb{P}_1(A_1) \mathbb{P}_2(A_2)}$$

#### **Teorema 9.7: Fubini-Tonelli**<sup>27</sup> (*J-P Th 10.3*)

Siano  $(\Omega_1, \mathcal{A}_1, \mathbb{P}_1)$  e  $(\Omega_2, \mathcal{A}_2, \mathbb{P}_2)$  spazi di probabilità.

1.  $\exists!$  la probabilità  $\mathbb{P} : \mathcal{A}_1 \otimes \mathcal{A}_2 \rightarrow [0, 1]$  tale che:

$$\mathbb{P}(A_1 \times A_2) = \mathbb{P}_1(A_1) \mathbb{P}_2(A_2) \quad \forall A_1 \in \mathcal{A}_1 \quad \forall A_2 \in \mathcal{A}_2$$

---

<sup>27</sup>Il teorema è in realtà l’unione di due teoremi distinti formulati dai matematici italiani Guido Fubini e Leonida Tonelli, vissuti a cavallo tra il XIX e il XX secolo. In particolare, il “teorema di Fubini” è la parte relativa alle funzioni  $L^1$  e il “teorema di Tonelli” la parte relativa alle funzioni positive. Lo J-P lo chiama Tonelli-Fubini, ma non credete a queste fake news.



Scriveremo  $\mathbb{P} = \mathbb{P}_1 \otimes \mathbb{P}_2$ .

2. Sia  $X : \Omega_1 \times \Omega_2 \rightarrow [0, +\infty]$  (positiva) e  $\mathcal{A}_1 \otimes \mathcal{A}_2$ -misurabile. Allora<sup>28</sup>:

- (b)  $\varphi(\omega_2) := \int_{\Omega_1} X(\omega_1, \omega_2) d\mathbb{P}_1$  è  $\mathcal{A}_2$ -misurabile e  
 $\psi(\omega_1) := \int_{\Omega_2} X(\omega_1, \omega_2) d\mathbb{P}_2$  è  $\mathcal{A}_1$ -misurabile;

(c) si ha il seguente *scambio di integrali*:

$$\begin{aligned} \int_{\Omega_2} \left( \int_{\Omega_1} X(\omega_1, \omega_2) d\mathbb{P}_1 \right) d\mathbb{P}_2 &= \int_{\Omega_1} \left( \int_{\Omega_2} X(\omega_1, \omega_2) d\mathbb{P}_2 \right) d\mathbb{P}_1 \\ &= \int_{\Omega_1 \times \Omega_2} X d(\mathbb{P}_1 \otimes \mathbb{P}_2) \end{aligned}$$

3. Sia  $X \in L^1(\Omega_1 \times \Omega_2, \mathcal{A}_1 \otimes \mathcal{A}_2, \mathbb{P}_1 \otimes \mathbb{P}_2)$ . Allora:

- (a)  $X_{\omega_2} = X(\cdot, \omega_2) : \Omega_1 \rightarrow \mathbb{R}$  è  $L^1(\mathbb{P}_1)$  qc rispetto a  $\omega_2$  e  
 $X_{\omega_1} = X(\omega_1, \cdot) : \Omega_2 \rightarrow \mathbb{R}$  è  $L^1(\mathbb{P}_2)$  qc rispetto a  $\omega_1$ ;

- (b)  $\varphi(\omega_2) := \int_{\Omega_1} X(\omega_1, \omega_2) d\mathbb{P}_1$  definita qc è  $L^1(\mathbb{P}_2)$  e  
 $\psi(\omega_1) := \int_{\Omega_2} X(\omega_1, \omega_2) d\mathbb{P}_2$  definita qc è  $L^1(\mathbb{P}_1)$ ;

(c) si ha il seguente *scambio di integrali*:

$$\begin{aligned} \int_{\Omega_2} \left( \int_{\Omega_1} X(\omega_1, \omega_2) d\mathbb{P}_1 \right) d\mathbb{P}_2 &= \int_{\Omega_1} \left( \int_{\Omega_2} X(\omega_1, \omega_2) d\mathbb{P}_2 \right) d\mathbb{P}_1 \\ &= \int_{\Omega_1 \times \Omega_2} X d(\mathbb{P}_1 \otimes \mathbb{P}_2) \end{aligned}$$

### Dimostrazione

1. Sia  $A \in \mathcal{A}_1 \otimes \mathcal{A}_2$ .

Sia definita la *proiezione*  $A_{\omega_2} := \{\omega_1 \in \Omega_1 : (\omega_1, \omega_2) \in A\}$ , che appartiene alla sua  $\sigma$ -algebra:  $A_{\omega_2} \in \mathcal{A}_1 \quad \forall \omega_2 \in \Omega_2$ . Dunque esiste  $\mathbb{P}_1(A_{\omega_2}) \quad \forall \omega_2$ , anche se non sappiamo ancora se sia misurabile rispetto a  $\omega_2$ .

- Sia  $A$  un *rettangolo*, ovvero tale per cui esistono  $A_1 \in \mathcal{A}_1, A_2 \in \mathcal{A}_2$  tali che  $A = A_1 \times A_2$ . Allora:

$$A_{\omega_2} = \begin{cases} \emptyset & \text{se } \omega_2 \notin A_2 \\ A_1 & \text{se } \omega_2 \in A_2 \end{cases}$$

Passando alle probabilità:

$$\mathbb{P}_1(A_{\omega_2}) = \begin{cases} 0 & \text{se } \omega_2 \notin A_2 \\ \mathbb{P}_1(A_1) & \text{se } \omega_2 \in A_2 \end{cases} = \mathbb{P}_1(A_1) \cdot \mathbf{1}_{A_2}(\omega_2)$$

<sup>28</sup>Il punto (a) è omesso: "Non ve lo dico" -Gregoratti.

$\mathbb{P}_1(A_1)$  è costante, e dunque  $\mathbb{P}_1(A_{\omega_2})$  è misurabile rispetto a  $\omega_2$ .

- Sia  $\mathcal{D} := \{A = \mathcal{A}_1 \otimes \mathcal{A}_2 : \mathbb{P}_1(A_{\omega_2}) \text{ misurabile rispetto a } \omega_2\}$ . Notiamo che  $\mathcal{A}_1 \times \mathcal{A}_2 \subseteq \mathcal{D} \subseteq \mathcal{A}_1 \otimes \mathcal{A}_2$ , la quale è chiusa per intersezioni finite (l'intersezione di rettangoli è ovviamente un rettangolo). Inoltre si può verificare facilmente anche che  $\mathcal{D}$  è chiusa per limiti crescenti e differenze. Per le classi monotone (teorema 3.4) dunque  $\mathcal{D} = \mathcal{A}_1 \otimes \mathcal{A}_2$ .
- $\mathbb{P}_1(A_{\omega_2})$  è dunque misurabile in  $\omega_2 \forall A \in \mathcal{A}_1 \otimes \mathcal{A}_2$ .  
Quindi è possibile definire la seguente funzione:

$$\mathbb{P} : \mathcal{A}_1 \otimes \mathcal{A}_2 \rightarrow [0, 1], \quad \mathbb{P}(A) := \int_{\Omega_2} \mathbb{P}_1(A_{\omega_2}) d\mathbb{P}_2$$

- Per dimostrare sia una probabilità bisogna dimostrare le due proprietà. La probabilità totale unitaria è immediata:

$$\mathbb{P}(\Omega_1, \Omega_2) = \int_{\Omega_2} \mathbb{P}_1(\Omega_1) d\mathbb{P}_2 = \int_{\Omega_2} 1 d\mathbb{P}_2 = 1$$

Dimostriamo ora la  $\sigma$ -additività di  $\mathbb{P}$ . Siano  $A_n \in \mathcal{A}_1 \otimes \mathcal{A}_2$  tali che  $A_k \cap A_l = \emptyset \forall k \neq l$ . Abbiamo che  $A_{n,\omega_2} \in \mathcal{A}_1$  e che  $A_{k,\omega_2} \cap A_{l,\omega_2} = \emptyset \forall k \neq l$  (ovvero gli  $A_{k,\omega_2} \cap A_{l,\omega_2}$  sono disgiunti), quindi:

$$\begin{aligned} \mathbb{P} \left( \bigcup_{n=1}^{+\infty} A_n \right) &= \int_{\Omega_2} \mathbb{P}_1 \left( \left( \bigcup_{n=1}^{+\infty} A_n \right)_{\omega_2} \right) d\mathbb{P}_2 \\ &= \int_{\Omega_2} \mathbb{P}_1 \left( \left( \bigcup_{n=1}^{+\infty} A_{n,\omega_2} \right) \right) d\mathbb{P}_2 \quad (\text{per } \sigma\text{-additività di } \mathbb{P}_1) \\ &= \int_{\Omega_2} \sum_{n=1}^{+\infty} \mathbb{P}_1(A_{n,\omega_2}) d\mathbb{P}_2 \\ &\quad (\text{poiché la funzione e la serie sono positive}) \\ &= \sum_{n=1}^{+\infty} \int_{\Omega_2} \mathbb{P}_1(A_{n,\omega_2}) d\mathbb{P}_2 = \sum_{n=1}^{+\infty} \mathbb{P}(A_n) \end{aligned}$$

L'esistenza di  $\mathbb{P}$  è così dimostrata.

- Serve ora ottenere l'unicità.

$$\begin{aligned}
 \mathbb{P}(A_1 \times A_2) &= \int_{\Omega_2} \mathbb{P}_1((A_1 \times A_2)_{\omega_2}) \, d\mathbb{P}_2 \\
 &= \int_{\Omega_2} \mathbb{P}_1(A_1) \mathbb{1}_{A_2}(\omega_2) \, d\mathbb{P}_2 \quad (\text{per definizione di } \mathbb{P}_1) \\
 &= \mathbb{P}_1(A_1) \int_{\Omega_2} \mathbb{1}_{A_2}(\omega_2) \, d\mathbb{P}_2 \quad (\mathbb{P}_1(A_1) \text{ è una costante}) \\
 &= \mathbb{P}_1(A_1) \mathbb{P}_2(A_2)
 \end{aligned}$$

$\mathbb{P}$  è caratterizzata dai suoi valori su  $\mathcal{A}_1 \times \mathcal{A}_2$  che è chiusa per intersezioni finite, quindi per il corollario 3.5 alle classi monotone,  $\mathbb{P}$  è unica:

$$\mathbb{P}_1 \otimes \mathbb{P}_2(A) = \int_{\Omega_2} \mathbb{P}_1(A_{\omega_2}) \, d\mathbb{P}_2 = \int_{\Omega_1} \mathbb{P}_2(A_{\omega_1}) \, d\mathbb{P}_1$$

2. • Se  $X = \mathbb{1}_A$ , con  $A \in \mathcal{A}_1 \otimes \mathcal{A}_2$ , allora per il punto precedente valgono i punti (b) e (c) della tesi, con  $\mathbb{P} = \mathbb{E}[\mathbb{1}_A]$ .
  - Se  $X$  è semplice, (b) e (c) valgono per linearità del valore atteso a partire dal caso precedente.
  - Se  $X$  è generica ma positiva, ovvero  $X : \Omega_1 \times \Omega_2 \rightarrow [0, +\infty]$  misurabile rispetto a  $\mathcal{A}_1 \otimes \mathcal{A}_2$ , si considerino le variabili semplici  $X_n$  tali che  $X_n(\omega) \uparrow X(\omega) \forall \omega$ . Grazie alla convergenza monotona si può avere uno scambio di limiti:

$$\begin{aligned}
 \varphi(\omega_2) &:= \int_{\Omega_1} X(\cdot, \omega_2) \, d\mathbb{P}_1 = \int_{\Omega_1} \left( \lim_n X_n(\cdot, \omega_2) \right) \, d\mathbb{P}_1 = \\
 &= \lim_n \int_{\Omega_1} X_n(\cdot, \omega_2) \, d\mathbb{P}_1 = \lim_n \varphi_n(\omega_2)
 \end{aligned}$$

Essendo  $\varphi_n(\omega_2)$  misurabile su  $\mathcal{A}_2$  per il punto precedente, anche il limite è misurabile su  $\mathcal{A}_2$  e dunque vale (b). Analogamente vale anche (c).

3. (a) Sia  $X \in L^1(\mathbb{P}_1 \otimes \mathbb{P}_2)$ . Allora  $|X|$ ,  $X_+$ ,  $X_-$  sono tutte VA in  $L^1(\mathbb{P}_1 \otimes \mathbb{P}_2)$  e per loro vale (2) in quanto VA positive. Inoltre, cosa non sorprendente, si può dimostrare che  $|X|_{\omega_2} = |X_{\omega_2}|$ ,  $(X_+)_{\omega_2} = (X_{\omega_2})_+$  e  $(X_-)_{\omega_2} = (X_{\omega_2})_-$ .

Per il punto (2) vale che:

$$\begin{aligned}
 \int_{\Omega_2} \left( \int_{\Omega_1} |X_{\omega_2}| \, d\mathbb{P}_1 \right) \, d\mathbb{P}_2 &= \int_{\Omega_2} \varphi_{|X|}(\omega_2) \, d\mathbb{P}_2 \\
 &= \int_{\Omega_1 \times \Omega_2} |X| \, d\mathbb{P}_1 \otimes d\mathbb{P}_2 < +\infty
 \end{aligned}$$

Dunque  $\varphi_{|X|} < +\infty$   $\mathbb{P}_2$ -qc, vale a dire  $X_{\omega_2} \in L^1(\mathbb{P}_1)$  qc rispetto a  $\omega_2$ . Il punto (a) è dimostrato.

(b)  $\int_{\Omega_1} |X_{\omega_2}| d\mathbb{P}_1 = \varphi_{|X|}(\omega_2) \in L^1(\mathbb{P}_2)$ .

(c) Separando  $X$  come  $X = X_+ - X_-$ , per le quali vale il punto (2), la tesi è vera per linearità.  $\square$

Presentiamo ora un nuovo concetto di indipendenza, che dipende solo dalla legge delle variabili.

**Corollario 9.8**

Siano  $(\Omega, \mathcal{A}, \mathbb{P})$  spazio di probabilità,  $(E, \mathcal{E}), (F, \mathcal{F})$  spazi misurabili,  $X : \Omega \rightarrow E$  e  $Y : \Omega \rightarrow F$  variabili aleatorie, ovvero  $(X, Y) : \Omega^2 \rightarrow E \times F$  è una VA. Allora:

$$X \perp\!\!\!\perp Y \iff P^{(X,Y)} = P^X \otimes P^Y$$

$P^{(X,Y)}$  è detta *legge congiunta*, mentre  $P^X$  e  $P^Y$  sono dette *leggi marginali*. Si noti che mediante questo corollario l'indipendenza è stata "riscritta" nel linguaggio delle leggi ed è stata trasferita dal dominio al codominio.

**Dimostrazione**

Si ha la seguente catena di equivalenze:

$$\begin{aligned} X \perp\!\!\!\perp Y &\iff \mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A) \cdot \mathbb{P}(Y \in B) \quad \forall A \in \mathcal{E}, \forall B \in \mathcal{F} \\ &\iff P^{(X,Y)} = P^X(A) \cdot P^Y(B) \quad \forall A \in \mathcal{E}, \forall B \in \mathcal{F} \\ &\iff P^{(X,Y)} = P^X \otimes P^Y \end{aligned} \quad \square$$

**Corollario 9.9**

Sia  $(\Omega, \mathcal{A}, \mathbb{P})$  uno spazio di probabilità, e siano  $X : \Omega \rightarrow \mathbb{R}$  e  $Y : \Omega \rightarrow \mathbb{R}$  VAR appartenenti a  $L^1(\mathbb{P})$ . Se  $X \perp\!\!\!\perp Y$ , allora anche  $XY \in L^1(\mathbb{P})$ .

**Dimostrazione**

$$\begin{aligned} \mathbb{E}[|XY|] &= \int_{\mathbb{R}^2} |st| P^{(X,Y)} ds dt && \text{(per la regola del valore atteso)} \\ &= \int_{\mathbb{R}^2} |st| P^X \otimes P^Y ds dt && \text{(per il corollario 9.8)} \\ &= \int_{\mathbb{R}} \left( \int_{\mathbb{R}} |st| P^X ds \right) P^Y dt && \text{(per Fubini-Tonelli)} \\ &= \int_{\mathbb{R}} |s| P^X ds \cdot \int_{\mathbb{R}} |t| P^Y dt && \text{(portando fuori da integrale)} \\ &= \mathbb{E}[|X|] \mathbb{E}[|Y|] < +\infty \end{aligned}$$

Poiché  $XY \in L^1 \iff |XY| \in L^1 \iff \mathbb{E}[|XY|] < +\infty$ , la dimostrazione è conclusa.  $\square$

### 9.4.1 Prodotti di probabilità

Siano due VA:  $X_1 : (\Omega_1, \mathcal{A}_1, \mathbb{P}_1) \rightarrow (E, \mathcal{E})$  di legge  $P^{X_1}$ ,  $Y_2 : (\Omega_2, \mathcal{A}_2, \mathbb{P}_2) \rightarrow (F, \mathcal{F})$  di legge  $P^{Y_2}$ . Si costruisce quindi un nuovo spazio di probabilità:

$$\Omega = \Omega_1 \times \Omega_2, \quad \mathcal{A} = \mathcal{A}_1 \otimes \mathcal{A}_2, \quad \mathbb{P} = \mathbb{P}_1 \otimes \mathbb{P}_2$$

Su di esso definiamo le seguenti VA:

$$X : \Omega \rightarrow E, \quad X(\omega_1, \omega_2) = X_1(\omega_1) \quad \text{e} \quad Y : \Omega \rightarrow F, \quad Y(\omega_1, \omega_2) = Y_2(\omega_2)$$

Allora:

- $X$  è variabile aleatoria.  
Infatti, preso  $A \in \mathcal{E}$ :

$$(X \in A) = X^{-1}(A) = (X_1 \in A) \times \Omega_2 = X_1^{-1}(A) \times \Omega_2 = \mathcal{A}_1 \times \mathcal{A}_2 \subseteq \mathcal{A}_1 \otimes \mathcal{A}_2$$

- $Y$  è una VA (la dimostrazione è analoga);
- $P^X = P^{X_1}$  e  $P^Y = P^{Y_2}$ .

Infatti:

$$P^X(A) = \mathbb{P}(X \in A) = \mathbb{P}((X_1 \in A) \times \Omega_2) = \mathbb{P}_1(X \in A)\mathbb{P}_2(\Omega_2) = P^{X_1}(A) \quad \forall A$$

Si procede analogamente per  $P^Y$ ; dunque le leggi marginali corrispondono alle leggi di partenza.

- $P^{(X,Y)} = P^X \otimes P^Y$  e di conseguenza  $X \perp\!\!\!\perp Y$ .  
Infatti  $\forall A \in \mathcal{A}_1, \forall B \in \mathcal{A}_2$ :

$$\begin{aligned} P^{(X,Y)}(A \times B) &= \mathbb{P}(X \in A, Y \in B) = \mathbb{P}((X_1 \in A) \times (Y_2 \in B)) \\ &= \mathbb{P}_1(X_1 \in A)\mathbb{P}_2(Y_2 \in B) = P^{X_1}(A)P^{Y_2}(B) \end{aligned}$$

*Esempio.* Consideriamo due lanci di dadi separati.

Definiamo lo spazio di probabilità relativo al primo lancio:  $\Omega_1 = \{1, \dots, 6\}$ ,  $\mathcal{A}_1 = 2^{\Omega_1}$ ,  $\mathbb{P}_1$  probabilità uniforme:

$$p_k = \frac{1}{6} \quad \forall k = 1, \dots, 6$$

Analogamente definiamo  $\Omega_2, \mathcal{A}_2$  e  $\mathbb{P}_2$  per il secondo lancio. Siano poi  $X_1 : \Omega_1 \rightarrow \mathbb{R}$ ,  $X_1(\omega_1) = \omega_1$  e  $Y_2 : \Omega_2 \rightarrow \mathbb{R}$ ,  $Y_2(\omega_2) = \omega_2$  i risultati rispettivamente del primo e del secondo lancio.

Consideriamo ora i due lanci come un unico esperimento.

Possiamo allora definire lo spazio di probabilità  $\Omega = \Omega_1^2 = \{1, \dots, 6\}^2$ ,  $\mathcal{A} = \mathcal{A}_1 \otimes \mathcal{A}_2$ ,  $\mathbb{P} = \mathbb{P}_1 \otimes \mathbb{P}_2$  uniforme:

$$\mathbb{P}(\{\omega\}) = \mathbb{P}(\{\omega_1, \omega_2\}) = \mathbb{P}(\{\omega_1\} \times \{\omega_2\}) = \mathbb{P}_1(\{\omega_1\}) \cdot \mathbb{P}_2(\{\omega_2\}) = \frac{1}{6} \cdot \frac{1}{6} = \frac{1}{36} \quad \forall \omega$$

Di conseguenza abbiamo: 
$$\begin{cases} X(\omega_1, \omega_2) = X_1(\omega_1) = \omega_1 \\ Y(\omega_1, \omega_2) = Y_2(\omega_2) = \omega_2 \end{cases} \implies X \perp\!\!\!\perp Y$$

**Definizione 9.10**

Siano  $(\Omega, \mathcal{A}, \mathbb{P})$  spazio di probabilità,  $(F_i, \mathcal{F}_i)$  spazi misurabili con  $i \in I$  insieme (anche numerabile) di indici, e  $X_i : \Omega \rightarrow F_i \ \forall i \in I$ .  $\{X_i\}_{i \in I}$  è detta **famiglia di VA indipendenti**<sup>29</sup> se per ogni sottoinsieme finito  $J$  di  $I$ :

$$\mathbb{P} \left( \bigcap_{j \in J} (X_j \in B_j) \right) = \prod_{j \in J} \mathbb{P}(X_j \in B_j) \quad \forall B_j \in \mathcal{F}_j$$

**Osservazione 9.11**

$X_i \perp\!\!\!\perp X_j \ \forall i \neq j$ , ma anche fissando una qualsiasi stringa a valori in  $\mathbb{R}$ :

$$(X_{i_1}, \dots, X_{i_n}) \perp\!\!\!\perp (X_{j_1}, \dots, X_{j_n}) \quad \forall \{i_1, \dots, i_n\} \cup \{j_1, \dots, j_n\} = \emptyset$$

*Controesempio.* Date  $X$  e  $Y$  VA,  $X \sim Y \sim U(\{-1, 1\})$ ,  $X \perp\!\!\!\perp Y$  definiamo una terza VA  $Z : \Omega \rightarrow \{-1, 1\}$   $Z = XY$ .

Si può mostrare che  $X$  è indipendente da  $Z$  calcolando la probabilità dell'intersezione:

$$\mathbb{P}(X = 1, Z = 1) = \mathbb{P}(X = 1, Y = 1) = \mathbb{P}(X = 1)\mathbb{P}(Y = 1) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$$

e confrontandola con il prodotto delle probabilità:

$$\begin{aligned} \mathbb{P}(Z = 1) &= \mathbb{P}(XY = 1) = \mathbb{P}((X = 1, Y = 1) \cup (X = -1, Y = -1)) \\ &= \mathbb{P}(X = 1, Y = 1) + \mathbb{P}(X = -1, Y = -1) \\ &= \mathbb{P}(X = 1)\mathbb{P}(Y = 1) + \mathbb{P}(X = -1)\mathbb{P}(Y = -1) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2} \end{aligned}$$

---

<sup>29</sup>Ricorrendo ad un piccolo abuso di linguaggio si può anche dire, per semplicità di scrittura, che le suddette VA sono indipendenti, omettendo la precisazione “famiglia di”. Ovviamente si fa comunque riferimento a questa definizione: quando si parla di  $n$  VA indipendenti, non si chiede solo che siano indipendenti a coppie, ma che tutti i possibili sottoinsiemi siano indipendenti, come richiesto, appunto, dalla definizione di cui sopra.

Moltiplicando le probabilità si ottiene infatti lo stesso risultato:

$$\mathbb{P}(Z = 1)\mathbb{P}(X = 1) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$$

I procedimenti negli altri casi possibili sono analoghi.

Per simmetria anche  $Y \perp\!\!\!\perp Z$ , quindi le coppie sono indipendenti, ma  $(X, Y, Z)$  non può essere una famiglia di eventi indipendenti, infatti  $Z = h(X, Y)$ .

### 9.4.2 Prodotti numerabili

Dati i modelli  $X_k : (\Omega_k, \mathcal{A}_k, \mathbb{P}_k) \rightarrow (F_k, \mathcal{F}_k)$  di legge  $P^{X_k}$  con  $k \in \mathbb{N}$  si può costruire  $\tilde{X}_k \sim P^{X_k}$  con  $\{\tilde{X}_k\}_{k \in \mathbb{N}}$  famiglia di variabili aleatorie indipendenti. Siano:

$$\Omega = \times_{k=1}^{+\infty} \Omega_k, \quad \omega = (\omega_k)_1^{+\infty}$$

Allora si può definire la  $\sigma$ -algebra prodotto come:

$$\begin{aligned} \mathcal{A} &= \bigotimes_{k=1}^{+\infty} \mathcal{A}_k = \sigma \left\{ \underbrace{A_1 \times \cdots \times A_n \times \Omega_{n+1} \times \Omega_{n+2} \times \cdots}_{\text{cilindri misurabili}} \mid n \in \mathbb{N} \text{ e } A_k \in \mathcal{A}_k \right\} \\ &= \sigma \{ \Omega_1 \times \cdots \times \Omega_{n-1} \times A_n \times \Omega_{n+1} \times \cdots \mid n \in \mathbb{N} \text{ e } A_n \in \mathcal{A}_k \} \end{aligned}$$

**NB:** Questa definizione di  $\mathcal{A}$  non è chiusa per intersezioni.

#### Teorema 9.12: Super Fubini (*J-P Th 10.4*)

Siano  $(\Omega_k, \mathcal{A}_k, \mathbb{P}_k)$  spazi di probabilità,  $k \in \mathbb{N}$ ,  $\Omega = \times_{k=1}^{+\infty} \Omega_k$ ,  $\mathcal{A} = \bigotimes_{k=1}^{+\infty} \mathcal{A}_k$ .

Allora  $\exists! \mathbb{P} : \mathcal{A} \rightarrow [0, 1]$  tale che:

$$\mathbb{P}(A_1 \times \cdots \times A_n \times \Omega_{n+1} \times \cdots) = \prod_{k=1}^n \mathbb{P}_k(A_k) \quad \forall n \in \mathbb{N} \quad \forall A_k \in \mathcal{A}_k$$

Indichiamo questa probabilità come:

$$\mathbb{P} = \bigotimes_{n=1}^{+\infty} \mathbb{P}_n$$

*Esempio.* Verifichiamo che  $\exists! \mathbb{P}$  su  $\Omega = \{0, 1\}^{\mathbb{N}}$ ,  $\mathcal{A} = \sigma(E_k | k \in \mathbb{N})$ ,  $E_k = \{\omega | \omega_k = 1\}$  tale che:

$$\begin{cases} \mathbb{P}(E_k) = p & \forall k \in \mathbb{N} \\ \{E_k\}_k & \text{indipendenti} \end{cases}$$

Allora dato un  $\tilde{X}_k : \Omega \rightarrow F_k$   $\tilde{X}_k(\omega) = X_k(\omega_k)$  estensione naturale di  $X_k$  implica che  $\tilde{X}_k$  è misurabile e  $\tilde{X}_k \sim P^{X_k}$ .

Come visto nell'esempio si può mostrare formalmente che questa probabilità è proprio  $P^{X_k}$ .

**Corollario 9.13**

Dati  $X_k : (\Omega_k, \mathcal{A}_k, \mathbb{P}_k) \rightarrow (F_k, \mathcal{F}_k)$  di legge  $P^{X_k}$  con  $k \in \mathbb{N}$ , siano:

$$\Omega = \prod_{k=1}^{+\infty} \Omega_k, \quad \mathcal{A} = \bigotimes_{k=1}^{+\infty} \mathcal{A}_k \quad \text{e} \quad \mathbb{P} = \bigotimes_{n=1}^{+\infty} \mathbb{P}_n$$

Siano inoltre  $\tilde{X}_k$  le estensioni naturali delle  $X_k$ . Allora:

$$\tilde{X}_k \sim P^{X_k} \quad \text{con} \quad \left( \tilde{X}_k \right)_{k \in \mathbb{N}} \quad \text{indipendenti}$$

**Dimostrazione**

La controimmagine del  $k$ -esimo elemento è:

$$(\tilde{X}_k \in B) = \Omega_1 \times \dots \times \Omega_{k-1} \times (X_k \in B) \times \Omega_{k+1} \times \dots$$

Pertanto intersecando le prime  $n$  si creano dei *cilindri misurabili*:

$$\begin{aligned} \mathbb{P} \left( \bigcap_{k=1}^n (\tilde{X}_k \in B_k) \right) &= \mathbb{P}((X_1 \in B_1) \times \dots \times (X_n \in B_n) \times \Omega_{n+1} \times \dots) \\ &= \mathbb{P}(X_1 \in B_1) \dots \mathbb{P}(X_n \in B_n) \\ &= \mathbb{P}(\tilde{X}_1 \in B_1) \dots \mathbb{P}(\tilde{X}_n \in B_n) \quad \forall n, \quad \forall B_k \in \mathcal{F}_k \end{aligned}$$

L'ultimo passaggio fa uso della scomposizione di  $\mathbb{P}$  come da teorema 9.12. □

*Esempio.* Assegnata una legge  $P^X$  su  $(F, \mathcal{F})$  sappiamo costruire:

$$X : (\Omega, \mathcal{A}, \mathbb{P}) \rightarrow (F, \mathcal{F}) \quad \text{tale che} \quad X \sim P^X$$

Da qui possiamo calcolare la sua estensione naturale  $\tilde{X}$ :

$$\tilde{X}_k : (\tilde{\Omega}, \tilde{\mathcal{A}}, \tilde{\mathbb{P}}) \rightarrow (F, \mathcal{F}) \quad \text{tali che} \quad \tilde{X}_n \sim P^X \quad \text{con} \quad \tilde{X}_n \quad \text{indipendenti} \quad (\tilde{X}_n \text{ iid } P^X)$$



## 10 Vettori aleatori

La trattazione dei *vettori aleatori* è una parte fondamentale dello studio della probabilità: essi rappresentano infatti il caso più generale e utilizzato delle variabili aleatorie. Parte del capitolo servirà per reintrodurre la teoria precedentemente sviluppata sulle variabili monodimensionali, che si manterrà sostanzialmente tale e quale nel passaggio da  $\mathbb{R}$  a  $\mathbb{R}^n$ . Sarà inoltre introdotta la *covarianza*, che indica il legame probabilistico tra le componenti del vettore; da essa discenderà il *coefficiente di correlazione lineare*, molto utile anche in ambito statistico.

I **vettori aleatori** sono variabili aleatorie a valori in  $(\mathbb{R}^n, \mathcal{B}^n)$ , ovvero:

$$X = (x_1 \dots x_n) = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = [x_1 \dots x_n] = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} : (\Omega, \mathcal{A}, \mathbb{P}) \rightarrow (\mathbb{R}^n, \mathcal{B}^n)$$

Nel testo saranno usate interscambiabilmente la notazione colonna e quella riga e le parentesi tonde e quadre, a seconda della convenienza estetica del momento.<sup>30</sup>

Sappiamo già che:

- $\mathcal{B}^n = \sigma(\tau^n) = \mathcal{B}^{\otimes n} = \sigma\left(\bigtimes_{k=1}^n (-\infty, t_k] : t_k \in \mathbb{Q}\right)$

In quest'ultima scrittura gli insiemi generanti sono detti *iper-quadranti* di sud-ovest, già visti nell'esempio a pagina 110.

- $X$  misurabile  $\iff X_k$  misurabili  $\forall k$ ; ovvero, non ci sono problemi o perdite di misurabilità passando dal vettore alle componenti e viceversa.
- Se  $X : \Omega \rightarrow \mathbb{R}^n$  è misurabile e  $h : \mathbb{R}^n \rightarrow \mathbb{R}^k$  è boreliana, allora  $Y = h(X) : \Omega \rightarrow \mathbb{R}^k$  è misurabile.
- $X \stackrel{\text{qc}}{=} Y \iff X_k \stackrel{\text{qc}}{=} Y_k \forall k$ .
- $P^X$  dipende solo da  $[X]$ , la *classe di equivalenza* della variabile  $X$ .
- Se  $X_1, X_2 \in L^2$ , allora  $X_1 \cdot X_2 \in L^1$  e vale la *disuguaglianza di Cauchy-Schwarz*:

$$|\mathbb{E}[X_1 X_2]| \leq \sqrt{\mathbb{E}[X_1^2] \mathbb{E}[X_2^2]}$$

- Se  $X_1, X_2$  sono in  $L^1$  e  $X_1 \perp\!\!\!\perp X_2$ , allora  $X_1 X_2 \in L^1$ .
- Si ha la seguente serie di equivalenze:  $X_1 \perp\!\!\!\perp X_2$

---

<sup>30</sup>Ciò non toglie che se usi le tonde sei una brutta persona

$\iff \mathbb{P}(X_1 \in A, X_2 \in B) = \mathbb{P}_1(X_1 \in A)\mathbb{P}_2(X_2 \in B) \quad \forall A \in \mathcal{C}, \forall B \in \mathcal{D}$   
 con  $\sigma(\mathcal{C}) = \sigma(\mathcal{D}) = \mathcal{B}$  e  $\mathcal{C}, \mathcal{D}$  chiusi per intersezioni finite

$\iff h(X_1) \perp\!\!\!\perp g(X_2) \quad \forall h, \forall g$  misurabili

$\iff \mathbb{E}[h(X_1)g(X_2)] = \mathbb{E}[h(X_1)]\mathbb{E}[g(X_2)]$   
 $\forall h, \forall g$  misurabili e positive, misurabili e limitate, oppure continue e limitate

$\iff P^{(X_1, X_2)} = P^{X_1} \otimes P^{X_2}$ .

**NB:** L'ultima condizione è la più facile da verificare e la più utilizzata nelle applicazioni pratiche.

### Definizione 10.1

$F_X : \mathbb{R}^n \rightarrow [0, 1]$  è detta **funzione di ripartizione** in  $\mathbb{R}^n$  se:

$$F_X(t) = F_X(t_1, \dots, t_n) := \mathbb{P}(X_1 \leq t_1, \dots, X_n \leq t_n) = \mathbb{P}\left(X \in \bigtimes_{k=1}^n (-\infty, t_k]\right)$$

Equivalentemente,  $F_X(t_1, \dots, t_n)$  è la probabilità che  $X$  sia in un iper-quadrante di sud-ovest delimitato dai valori  $t_k$  assegnati.

Nel caso generale di  $n \geq 2$  la definizione risulta spesso complicata, quindi si tenterà di farne a meno nei successivi sviluppi della teoria.

## 10.1 Vettori aleatori discreti<sup>31</sup>

### Definizione 10.2

La funzione positiva  $p : S \rightarrow [0, 1]$ , con  $S$  al più numerabile, è detta **densità discreta di probabilità** se:

$$\sum_{x \in S} p(x) = 1 \quad \text{e} \quad P^X(B) = \sum_{x \in S \cap B} p(x)$$

### Definizione 10.3

Sia  $(\Omega, \mathcal{A}, \mathbb{P})$  uno spazio di probabilità e  $X : \Omega \rightarrow \mathbb{R}^n$  misurabile.

$X$  è detto **vettore aleatorio discreto** se la sua legge  $P^X$  è discreta su  $\mathcal{B}^n$ , ovvero esiste un insieme discreto  $S$ , al più numerabile, su cui è definita una densità discreta di probabilità  $p : S \rightarrow [0, 1]$ .

Si noti che la definizione non riguarda il dominio di  $X$  ma solamente il codominio, ovvero la legge  $P^X$ . Si noti inoltre che la definizione non esclude la possibilità che  $S$  contenga punti a probabilità nulla, i quali sono “inutili” nella costruzione della

<sup>31</sup>La trattazione dei vettori aleatori discreti non è presente sullo J-P; evidentemente pensavano (sbagliando) che fosse possibile arrivarci da soli a questo punto del corso.

legge. Per questo motivo  $S$  non è univocamente determinato, anche se, quando c'è possibilità di scegliere, molto spesso si prende  $S$  contenente solo punti con probabilità positiva, per evitare ridondanze e garantire comunque l'unicità di  $S$ .

### 10.1.1 Condizioni equivalenti

Le seguenti affermazioni sono equivalenti:

- $X$  è vettore aleatorio discreto;
- $\exists S \subset \mathbb{R}^n$ , al più numerabile, tale che  $\mathbb{P}(X \in S) = 1$ ;
- $\text{im}(X) = S$  al più numerabile a meno di elementi trascurabili, ovvero  $X \stackrel{\text{qc}}{=} \tilde{X}$  con  $\text{im}(\tilde{X}) = S$  al più numerabile; (per esempio,  $\tilde{X} = X\mathbb{1}_S(X) + x_0\mathbb{1}_{S^c}(X)$  con  $x_0$  arbitrario, e comunque irrilevante in quanto è raggiunto da  $X$  con probabilità nulla)
- $X_1, \dots, X_n$  VA *discrete*.

Dimostriamo ora la coimplicazione “ $X$  vettore aleatorio discreto  $\iff X_1, \dots, X_n$  VA discrete”, che è di gran lunga la più importante: è facile da verificare ed è peculiare dei vettori discreti (si vedrà più avanti che questa equivalenza non è valida nel caso continuo).

#### Dimostrazione

- ( $\implies$ ):

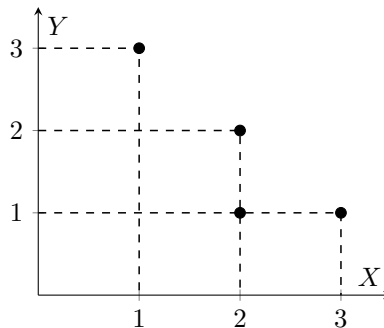


Figura 10.1: proiezione di un vettore aleatorio di dimensione 2

Se  $X$  è discreto,  $\exists S$  discreto tale che  $\mathbb{P}(X \in S) = 1$ . Sia ora  $S_k$  la *proiezione* di  $S$  sull'asse  $X_k$  (per definirla in astratto è necessario un prodotto scalare, ma una definizione geometrica è sufficiente per l'intuizione).

Poiché  $X \in S \implies X_k \in S_k$ , si ha la relazione tra eventi  $(X_k \in S_k) \supseteq (X \in S)$ : dunque  $\mathbb{P}(X_k \in S_k) \geq \mathbb{P}(X \in S) \stackrel{\text{hp}}{=} 1$ . Ma allora  $\mathbb{P}(X_k \in S_k) = 1$ , che è

la tesi.

- ( $\Leftarrow$ ): per ipotesi esistono  $S_1, \dots, S_n$ , supporti al più numerabili di  $X_1, \dots, X_n$ .

Si definisca la seguente *griglia*:

$$S = \bigtimes_{k=1}^n S_k$$

Essa è un insieme  $n$ -rettangolare nel quale sono sicuramente racchiusi tutti i punti di  $X$  su cui è concentrata la probabilità, più un numero indefinito di punti a probabilità nulla (che comunque non ostruiscono la dimostrazione). Dunque:

$$\Omega \supseteq (X \in S) = \bigcap_{k=1}^n (X_k \in S_k) \implies \mathbb{P}(X \in S) = \mathbb{P}\left(\bigcap_{k=1}^n (X_k \in S_k)\right)$$

Tutti gli eventi dell'intersezione hanno probabilità 1, pertanto anche la loro intersezione  $S$  ha probabilità 1.  $\square$

### 10.1.2 Legge congiunta e legge marginale

Cerchiamo ora un legame tra la legge  $P^X$  di  $X$ , detta *legge congiunta*, e le leggi  $P^{X_k}$  delle  $X_k$ , dette *leggi marginali*.

#### Proposizione 10.4

Sia  $X$  vettore aleatorio discreto con densità  $p$  sul supporto  $S = \bigtimes_{k=1}^n S_k$ .

Allora la VA  $X_k$  ha densità discreta:

$$p_k(x_k) = \sum_{\substack{x_1 \in S_1 \\ \vdots \\ x_{k-1} \in S_{k-1} \\ x_{k+1} \in S_{k+1} \\ \vdots \\ x_n \in S_n}} p(x_1, \dots, x_n)$$

In altre parole, chiedendo la legge marginale di  $X_k$ , stiamo tenendo fisso il valore di  $X_k$  e stiamo facendo la somma dei termini della legge congiunta *al variare di tutte le coordinate che **non** sto sommando*, ovvero tutte tranne  $X_k$ .

#### Dimostrazione

Per semplicità di notazione sarà dimostrato solo il caso  $n = 2$ ; il caso generale è analogo.

Si definisca  $p_1$  nel seguente modo, che può essere riscritto grazie alla formula

delle probabilità totali:

$$p_1(x_1) := \mathbb{P}(X_1 = x_1) = \sum_{x_2 \in S_2} \mathbb{P}(X_1 = x_1, X_2 = x_2) = \sum_{x_2 \in S_2} p(x_1, x_2)$$

Ciò è possibile in quanto gli eventi della forma  $(x_k \in S_k)$  formano una partizione di  $S$  (ancora, coprire tutto  $\Omega$  non è necessario in quanto fuori da  $S$  esistono solo punti a probabilità nulla). Similmente si ricava anche  $p_2(x_2)$ .  $\square$

### 10.1.3 Valore atteso per vettori discreti

#### Proposizione 10.5: regola del valore atteso

Siano  $X : (\Omega, \mathcal{A}, \mathbb{P}) \rightarrow (\mathbb{R}^n, \mathcal{B}^n)$  un vettore aleatorio discreto con densità  $p$  e  $h : \mathbb{R}^n \rightarrow \mathbb{R}$  una funzione misurabile, con  $h$  positiva oppure in  $L^1(P^X)$ . Allora:

$$\mathbb{E}[h(X)] = \int_{\Omega} h(X) d\mathbb{P} = \int_{\mathbb{R}^n} h(x) dP^X \stackrel{\downarrow}{=} \sum_{x \in S} h(x) p(x)$$

La dimostrazione è simile a quella per le precedenti versioni della regola del valore atteso.<sup>32</sup>

### 10.1.4 Indipendenza

#### Proposizione 10.6

Sia  $X$  un vettore aleatorio discreto. Allora  $\{X_k\}_{k=1, \dots, n}$  è una famiglia di VA *indipendenti* se e solo se:

$$p(x_1, \dots, x_n) = p_1(x_1) \cdots p_n(x_n) \quad \forall x_k \in S_k, \text{ con } k = 1, \dots, n$$

In altre parole, una condizione sia necessaria che sufficiente per l'indipendenza di  $X_1, \dots, X_n$  è che la legge congiunta si fattorizzi nelle leggi marginali. Questa proprietà è molto utile negli esercizi, dove si può riconoscere immediatamente se  $p(x_1, \dots, x_n)$  è il prodotto di  $n$  funzioni ciascuna in una variabile diversa.

#### Dimostrazione

- ( $\implies$ ):

$$\begin{aligned} p(x_1, \dots, x_n) &= \mathbb{P}(X_1 = x_1, \dots, X_n = x_n) \\ &= \mathbb{P}(X_1 = x_1) \cdots \mathbb{P}(X_n = x_n) \quad (\text{per l'indipendenza}) \\ &= p_1(x_1) \cdots p_n(x_n) \end{aligned}$$

---

<sup>32</sup>Pertanto un simpatico esercizio da amanuense

- ( $\Leftarrow$ ): Sia  $B = \prod_{k=1}^n B_k$  (con  $B_k \in \mathcal{B}$ ) un generico rettangolo e  $S = \prod_{k=1}^n S_k$  il supporto della legge  $p$  di  $X$ . Allora:

$$\mathbb{P}((X_1 \in B_1), \dots, (X_n \in B_n)) = \mathbb{P}(X \in B) = P^X(B) = \sum_{x \in S \cap B} p(x)$$

Scomponendo le sommatorie in tutti i  $B_1, \dots, B_n$ :

$$\begin{aligned} \sum_{x \in S \cap B} p(x) &= \sum_{x_1 \in S_1 \cap B_1} \dots \sum_{x_n \in S_n \cap B_n} p(x_1, \dots, x_n) \\ &= \sum_{x_1 \in S_1 \cap B_1} \dots \sum_{x_n \in S_n \cap B_n} p_1(x_1) \dots p_n(x_n) \quad (\text{per ipotesi}) \end{aligned}$$

Portando fuori dalle sommatorie interne i termini costanti uno alla volta, dopo numerosi passaggi otteniamo:

$$\left( \sum_{x_1 \in S_1} p_1(x_1) \right) \dots \left( \sum_{x_n \in S_n} p_n(x_n) \right) = \mathbb{P}(X_1 \in B_1) \dots \mathbb{P}(X_n \in B_n) \quad \square$$

*Esempio:* lancio di 2 dadi non truccati e indipendenti.

Dati  $X_1, X_2 \sim U(\{1, \dots, 6\})$ , con  $X_1 \perp\!\!\!\perp X_2$ , sono i risultati del lancio del primo e secondo dado, rispettivamente. Ricordiamo che non serve precisare  $(\Omega, \mathcal{A}, \mathbb{P})$  perché all'occorrenza lo sappiamo costruire, come visto nelle pagine precedenti.

L'indipendenza implica che:

$$p(x_1, x_2) = p_1(x_1) p_2(x_2) = \frac{1}{6} \cdot \frac{1}{6} = \frac{1}{36} \quad \forall x_1, x_2$$

Ora, dunque, sappiamo che l'uniformità di  $p(x_1, x_2)$  è non solo condizione necessaria, ma anche *sufficiente* per l'indipendenza di  $X_1$  e  $X_2$ .

Sia inoltre  $Y = X_1 + X_2$ . Allora  $Y \not\perp\!\!\!\perp X_1$ ; è possibile verificarlo notando che, per esempio:

$$\mathbb{P}(Y = 2, X_1 = 6) = 0$$

$$\mathbb{P}(Y = 2) \mathbb{P}(X_1 = 6) = [\mathbb{P}(X_1 = 1) \mathbb{P}(X_2 = 1)] \cdot \mathbb{P}(X_1 = 6) = \frac{1}{36} \cdot \frac{1}{6} = \frac{1}{216}$$

$$\implies \mathbb{P}(Y = 2, X_1 = 6) \neq \mathbb{P}(Y = 2) \mathbb{P}(X_1 = 6)$$

Oppure, intuitivamente, si osserva che conoscere certi valori di una delle due VA porta a un cambiamento del grado di fiducia nell'altra (come nel caso precedente), negando la possibilità d'indipendenza tra le due.

Cerchiamo ora la legge del vettore  $(Y, X_1)$ , che sappiamo essere discreto perché le sue componenti sono discrete. Trovare la legge di un vettore discreto equivale a riempire una tabella con tutte le possibili combinazioni delle componenti, ciascuna con la sua probabilità *congiunta* data dalla formula generale  $p(x, y) = \mathbb{P}(X_1 = x, Y = y)$ , che, ricordiamo, non può essere fattorizzata ulteriormente per la mancanza di indipendenza tra  $X_1$  e  $Y$ . In questo caso particolare le probabilità congiunte sono tutte o  $\frac{1}{36}$  (in quanto prodotto di due termini delle uniformi) o nulle (per gli eventi incompatibili come quello sopra riportato). Sommando per righe o per colonne i valori nella tabella si ottengono le probabilità *marginali* delle due variabili  $X_1$  e  $Y$ . Per esempio:

$$p_1(1) = \sum_{y=2}^{12} p(1, y)$$

Introduciamo infine un terzo dado, indipendenti dai primi due, i cui risultati sono rappresentati da  $X_3 \sim U(\{1, \dots, 6\})$ . È vero che  $Y \perp\!\!\!\perp X_3$ ? Sappiamo che  $\{X_1, X_2, X_3\}$  è una famiglia indipendente per costruzione, ovvero:

$$(X_1, X_2) \perp\!\!\!\perp X_3 \implies Y = h(X_1, X_2) \perp\!\!\!\perp X_3$$

La legge di  $(Y, X_3)$  è dunque data da:

$$q(y, z) = \mathbb{P}(Y = y, X_3 = z) \stackrel{\perp\!\!\!\perp}{=} p_2(y) p_1(z)$$

Calcolando le leggi marginali di  $(Y, X_1)$  si nota che sono le stesse di  $(Y, X_3)$ , ma che la congiunta è diversa! Questa è un'osservazione importante: conoscere le leggi marginali **non** significa conoscere la legge congiunta, perché su quest'ultima c'è ambiguità e si possono costruire potenzialmente infinite congiunte ugualmente valide.

## 10.2 Vettori aleatori continui

Riprendiamo la definizione di misura di Lebesgue adattandola al caso multidimensionale.

### 10.2.1 Misura di Lebesgue su $\mathbb{R}^n$

#### Definizione 10.7

La **misura di Lebesgue** su  $(\mathbb{R}^n, \mathcal{B}^n)$  è la funzione  $m_n : \mathcal{B}^n \rightarrow [0, +\infty]$  tale che:

1.  $m_n$  è  $\sigma$ -additiva
2.  $m_n(A_1 \times \dots \times A_n) = m(A_1) \cdots m(A_n) \quad \forall A_1, \dots, A_n \in \mathcal{B}$

Per  $A \in \mathcal{B}^n$ ,  $m_n(A)$  è detto *volume* di  $A$ .

**Proposizione 10.8**

La misura di Lebesgue multidimensionale esiste ed è unica ( $\exists! m_n$ ).

Può essere data una definizione equivalente di  $m_n$  basata sui rettangoli: la *misura di Lebesgue* su  $(\mathbb{R}^n, \mathcal{B}^n)$  è la funzione  $m_n : \mathcal{B}^n \rightarrow [0, +\infty]$  tale che:

1.  $m_n$  è  $\sigma$ -additiva
2.  $m_n \left( \prod_{k=1}^n (a_k, b_k] \right) = \prod_{k=1}^n (b_k - a_k) \quad \forall a_k, b_k : -\infty < a_k < b_k < +\infty$

Questa produttoria va immaginata come la costruzione di rettangoli, parallelepipedi o iperparallelepipedi, in cui iteriamo le dimensioni con l'indice  $k$ . Nel caso  $n = 2$  questa misura un'area, nel caso  $n = 3$  un volume.

Studiamo ora la misura di Lebesgue di alcuni insiemi notevoli:

- $\mathbb{R}^n$ :  $m_n(\mathbb{R}^n) = +\infty$  Per calcolare la dimensione di  $\mathbb{R}^n$  moltiplichiamo  $n$  volte la dimensione di  $\mathbb{R}$ , che è  $+\infty$ . Visto che intendiamo misurare la dimensione  $m_n$ , non c'è nessun fattore nullo, quindi non c'è forma di indecisione.
- Iper-quadranti di sud-ovest:  $m_n \left( \prod_{k=1}^n (-\infty, t_k] \right) = \prod_{k=1}^n m((-\infty, t_k]) = +\infty$   
Come nel caso precedente anche a degli iper-quadranti si applica lo stesso ragionamento, a patto che tutti i  $t_k > -\infty$ .
- Punto:  $m_n(\{x\}) = 0$ .
- Retta:  $m_n(\mathbb{R} \times \{x_2\}) = 0$ .

**Osservazione 10.9**

In generale, vale la seguente proposizione:  $m_n(A) = 0$  se  $\dim A < n$ . Si può infatti dimostrare, per limiti crescenti, che nell'apparente forma di indecisione  $[+\infty] \cdot [0]$  prevale lo  $[0]$ .

La teoria dell'integrazione per  $X : (\Omega, \mathcal{A}, \mathbb{P}) \rightarrow (\mathbb{R}, \mathcal{B})$  misurabile si ripete tale e quale per  $h : (\mathbb{R}^n, \mathcal{B}^n, m_n) \rightarrow (\mathbb{R}, \mathcal{B})$  boreliana, ad eccezione di due proprietà. Nell'ambito della misura di Lebesgue:

- $X$  limitata  $\not\Rightarrow X \in L^1(\mathbb{P})$ ;
- $L^2(\mathbb{P}) \not\subseteq L^1(\mathbb{P})$ .

Notiamo in particolare che vale Fubini-Tonelli<sup>33</sup> con  $m_2 = m \otimes m$  e, nel caso multidimensionale,  $m_n = m \otimes \cdots \otimes m$ .

---

<sup>33</sup>♡



Sia  $h$  boreliana,  $h \geq 0$  oppure  $h \in L^1(\mathcal{m}_n)$ . Allora:

$$\int_{\mathbb{R}^n} h \, d\mathcal{m}_n = \int_{\mathbb{R}^n} h(x_1, \dots, x_n) \, dx_1 \cdots dx_n = \int_{\mathbb{R}^n} h \, dx$$

Dunque anche nel caso multidimensionale l'integrale di Lebesgue è un'estensione dell'integrale di Riemann.

Notiamo inoltre che  $\int_{\mathbb{R}^n} h \, d\mathcal{m}_n$  dipende solo da  $[h]$ , la classe di equivalenza quasi certa di  $h$ , a cui appartengono le funzioni boreliane  $\tilde{h}$  tali che  $h = \tilde{h}$  qo.

Infine, sia  $A \in \mathcal{B}^n$  tale che  $\mathcal{m}_n(A) = 0$ . Allora  $\int_A h(x) \, dx = 0$ .

### 10.2.2 Densità continua di probabilità

#### Definizione 10.10

Una probabilità  $\mathbb{P}$  su  $(\mathbb{R}^n, \mathcal{B}^n)$  **ammette densità (continua)** rispetto a  $\mathcal{m}_n$  se  $\exists f : \mathbb{R}^n \rightarrow [0, +\infty)$ , boreliana, tale che:

$$\mathbb{P}(A) = \int_A f(x) \, dx \quad \forall A \in \mathcal{B}^n$$

Se  $\mathbb{P} = P^X$ , allora  $X$  è un **vettore aleatorio continuo** con densità  $f$ , e le sue componenti si dicono **congiuntamente continue**.

**NB:**  $f$  non è necessariamente una funzione continua; l'aggettivo si riferisce al vettore aleatorio associato a essa.

#### Teorema 10.11 (J-P Th 12.1)

- $f : \mathbb{R}^n \rightarrow \mathbb{R}$  è densità della probabilità  $\mathbb{P}$  su  $(\mathbb{R}^n, \mathcal{B}^n)$  se e solo se:

$$f \text{ boreliana, } f \geq 0, \quad \int_{\mathbb{R}^n} f(x) \, dx = 1$$

- In tal caso,  $[f]$  caratterizza  $\mathbb{P}$ , ovvero:  
 $f$  densità per  $\mathbb{P}$ ,  $\tilde{f}$  boreliana,  $\tilde{f} \geq 0$ ,  $f = \tilde{f}$  qo  $\implies \tilde{f}$  densità per  $\mathbb{P}$
- Una  $\mathbb{P}$  su  $(\mathbb{R}^n, \mathcal{B}^n)$ , con densità  $f$ , caratterizza  $[f]$ , ovvero: se  $f$  e  $\tilde{f}$  sono densità per  $\mathbb{P}$ , allora  $f \stackrel{\text{qc}}{=} \tilde{f}$ .

*Esempio:* caso  $(\mathbb{R}^2, \mathcal{B}^2)$ . Siano  $S$  e  $S'$  due cerchi di raggio 1, con e senza frontiera:

$$S = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq 1\}, \quad S' = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 < 1\}$$

Allora,  $\overline{S'} = S$ .

Sia inoltre  $f(x) = \frac{1}{\pi} \cdot \mathbf{1}_S(x, y)$ .  $f$  è boreliana,  $f \geq 0$ , e inoltre:

$$\int_{\mathbb{R}} f(x, y) dx dy = \frac{1}{\pi} \int_{\mathbb{R}^2} \mathbf{1}_S(x, y) dx dy = \frac{1}{\pi} m_2(S) = \frac{\pi}{\pi} = 1$$

Dunque  $f$  è una densità continua di probabilità.

Analogamente si verifica che  $\tilde{f}(x) = \frac{1}{\pi} \cdot \mathbf{1}_{S'}(x, y)$  è una densità continua di probabilità:

$$\mathbb{P}(A) = \int_A \frac{1}{\pi} \mathbf{1}_S dx dy = \frac{1}{\pi} m_2(S \cap A)$$

Da ciò si deduce che  $\mathbb{P}(A)$  è una probabilità uniforme:

$$\tilde{\mathbb{P}}(A) = \int_A \frac{1}{\pi} \mathbf{1}_{S'} dx dy = \mathbb{P}(A)$$

Dunque la probabilità non cambia in presenza o meno della frontiera.

Sia  $B$  una parabola. Se  $X \sim U(S)$ , allora:

$$\mathbb{P}(X \in B) = P^X(B) = \frac{m_2(B \cap S)}{\pi} = 0$$

poiché  $B$  ha dimensione 1, mentre in questo esempio la probabilità è calcolata in  $\mathbb{R}^2$ .

*Controesempio.* Siano  $X \sim \mathcal{N}(0, 1)$ ,  $Y = X^2 \sim \chi^2(1)$  e consideriamo il vettore aleatorio:

$$W = (X, Y) = (X, X^2) : (\Omega, \mathcal{A}, \mathbb{P}) \rightarrow (\mathbb{R}^2, \mathcal{B}^2)$$

Calcoliamo la probabilità che il vettore appartenga alla parabola  $B$ :

$$\mathbb{P}((X, X^2) \in B) = 1 \neq \int_B f(x, y) dx dy$$

Dunque  $W$  ha componenti continue *ma non è continuo!*

### 10.2.3 Valore atteso per vettori continui

#### Teorema 10.12

Sia  $X = (X_1, \dots, X_n)$  un vettore aleatorio continuo su  $(\Omega, \mathcal{A}, \mathbb{P})$  con legge  $P^X$  e densità  $f$ , e sia  $h : \mathbb{R}^n \rightarrow \mathbb{R}$  una funzione boreliana. Allora:

$$h \in L^1(P^X) \iff hf \in L^1(m_n)$$

Inoltre, nel caso in cui  $h$  sia positiva o  $L^1(P^X)$ :

$$\mathbb{E}[h(X)] = \int_{\Omega} h(X_1, \dots, X_n) d\mathbb{P} = \int_{\mathbb{R}^n} h dP^X = \int_{\mathbb{R}^n} h(x)f(x) dx$$

Notiamo che è la stessa regola del valore atteso del caso monodimensionale, con un integrale multiplo su  $\mathbb{R}^n$  al posto di quello in  $\mathbb{R}$ . Anche la dimostrazione è simile e verrà pertanto omessa.

### 10.2.4 Continuità delle componenti

#### Teorema 10.13 (J-P Th 12.2)

Sia  $X = (X_1, \dots, X_n) : (\Omega, \mathcal{A}, \mathbb{P}) \rightarrow (\mathbb{R}^n, \mathcal{B}^n)$  un vettore aleatorio continuo di densità  $f$ . Allora:

1.  $\forall k, X_k : (\Omega, \mathcal{A}, \mathbb{P}) \rightarrow (\mathbb{R}, \mathcal{B})$  è una VA *continua* con densità:

$$f_k(x_k) = \int_{\mathbb{R}^{n-1}} f(x_1, \dots, x_n) dx_1 \cdots dx_{k-1} dx_{k+1} \cdots dx_n$$

Abbiamo ottenuto un integrale su tutte le componenti che non ci interessano, ovvero tutte tranne  $x_k$ .

2. **Criterio sull'indipendenza:**

$\{X_k\}_{k=1, \dots, n}$  è una famiglia di VA indipendenti se e solo se:

$$f(x_1, \dots, x_n) = f_1(x_1) \cdots f_n(x_n)$$

Definendo un'opportuna famiglia di funzioni  $h_k : \mathbb{R} \rightarrow \mathbb{R} \forall k$ , le  $X_k$  sono indipendenti se e solo se:

$$f(x_1, \dots, x_n) = h_1(x_1) \cdots h_n(x_n)$$

In entrambe i casi la relazione dev'essere verificata rispetto alla misura di Lebesgue  $m_n$ , ovvero quasi ovunque.

Questo significa che la legge di una famiglia di VA indipendenti è sempre fattorizzabile nelle leggi delle componenti.

#### Dimostrazione

1. Si sta cercando la  $f$  tale che  $\mathbb{P}(X_k \in B) = \int_B f_k(X_k) dx_k \quad \forall B \in \mathcal{B} :$

$$\begin{aligned} \mathbb{P}(X_k \in B) &= \mathbb{P}\left(\underbrace{X \in \mathbb{R} \times \cdots \times \mathbb{R} \times \overbrace{B}^{k\text{-esima pos.}} \times \mathbb{R} \times \cdots \times \mathbb{R}}_{=C \in \mathcal{B}^n}\right) \\ &= \int_C f(x_1, \dots, x_n) dx_1 \cdots dx_n \end{aligned}$$

Per Fubini-Tonelli quest'ultimo integrale diventa:

$$\int_B \left( \int_{\mathbb{R}^{n-1}} f(x_1, \dots, x_n) dx_1 \cdots dx_{k-1} dx_{k+1} \cdots dx_n \right) dx_k,$$

dove la funzione dentro la parentesi è la  $f_k(x_k)$  cercata. Si può infatti verificare facilmente che rispetta tutte le proprietà richieste a una densità (misurabilità, positività, e integrale globale unitario).

2. ( $\Leftarrow$ ): Per ipotesi, la  $f$  è fattorizzabile nelle  $f_k$ . Allora:

$$\begin{aligned} \mathbb{P}(X_1 \in B_1, \dots, X_n \in B_n) &= \mathbb{P}\left(X \in \times_{k=1}^n B_k\right) \\ &\stackrel{\text{hp}}{=} \int_D f_1(x_1) \cdots f_n(x_n) dx_1 \cdots dx_n \quad \left(\text{Per } D = \times_{k=1}^n B_k\right) \\ &= \int_{\mathbb{R}^n} \mathbb{1}_D(x_1, \dots, x_n) f_1(x_1) \cdots f_n(x_n) dx \\ &\stackrel{\text{F-T}}{=} \left(\int_{\mathbb{R}} \mathbb{1}_{B_1}(x_1) f_1(x_1) dx_1\right) \cdots \left(\int_{\mathbb{R}} \mathbb{1}_{B_n}(x_n) f_n(x_n) dx_n\right) \\ &= \mathbb{P}(x_1 \in B_1) \cdots \mathbb{P}(x_n \in B_n) \end{aligned}$$

La tesi è dimostrata. Si noti che questo procedimento di *fattorizzazione*, sia del dominio (nei  $B_k$ ) che del codominio (nelle indicatrici), sarà utilizzato spesso nelle prossime dimostrazioni e pertanto è bene imparare a padroneggiarlo. Nelle future applicazioni i passaggi centrali di questo procedimento saranno omessi.

( $\Rightarrow$ ): Sia nuovamente  $D = \times_{k=1}^n B_k$ . Allora vale:

$$\begin{aligned} X_k \text{ indipendenti} &\iff P^X = P^{X_1} \otimes \dots \otimes P^{X_n} \\ &\iff P^X(D) = P^X(B_1 \times \dots \times B_n) \\ &= P^{X_1}(B_1) \cdots P^{X_n}(B_n) \quad \forall B_1, \dots, B_n \in \mathcal{B} \end{aligned}$$

Per definizione, il membro sinistro e destro di questa ultima uguaglianza possono essere rispettivamente scritti come:

$$\int_D f(x_1, \dots, x_n) dx_1 \cdots dx_n = \left(\int_{B_1} f_1(x_1) dx_1\right) \cdots \left(\int_{B_n} f_n(x_n) dx_n\right)$$

Per Fubini-Tonelli applicato “al contrario”, il membro destro può essere riscritto:

$$\int_D f(x_1, \dots, x_n) dx_1 \cdots dx_n = \int_D \underbrace{f_1(x_1) \cdots f_n(x_n)}_{=: g(x_1, \dots, x_n)} dx_1 \cdots dx_n$$

$g$  è dunque una densità su  $(\mathbb{R}^n, \mathcal{B}^n)$ . Sia  $\mathbb{Q} : \mathcal{B}^n \rightarrow [0, 1]$  la probabilità associata a tale densità (è infatti noto che per ogni densità esiste una corrispondente probabilità).

Per quanto detto prima si sa che:

$$P^X(B_1 \times \cdots \times B_n) = \mathbb{Q}(B_1 \times \cdots \times B_n) \quad \forall B_1, \dots, B_n \in \mathcal{B}$$

Questo vale per ogni rettangolo di  $\mathcal{B}^n$ . Poiché la *classe dei rettangoli*  $\mathcal{C} = \{B_1 \times \cdots \times B_n : B_k \in \mathcal{B} \forall k\}$  è chiusa per intersezioni finite e genera la sua  $\sigma$ -algebra (i.e.  $\sigma(\mathcal{C}) = \mathcal{B}^n$ ), è possibile estendere l'uguaglianza all'intera  $\sigma$ -algebra grazie al corollario 3.5 delle classi monotone:  $P^X(A) = \mathbb{Q}(A) \forall A \in \mathcal{B}^n$ . Abbiamo dunque:

$$f(x_1, \dots, x_n) = f_1(x_1) \cdots f_n(x_n)$$

Questo vale qo, perché  $P^X$  caratterizza la classe di  $f$ . L'uguaglianza ottenuta è la tesi.

La dimostrazione della formula delle  $h_k$  è invece lasciata al lettore come esercizio.  $\square$

### 10.3 Considerazioni pratiche

Siano  $(\Omega, \mathcal{A}, \mathbb{P})$  spazio di probabilità e  $(X, Y) : \Omega \rightarrow \mathbb{R}^2$  vettore aleatorio.  $(X, Y)$  è continuo se e solo se:

- $P^{(X,Y)}$  ammette densità rispetto alla misura di Lebesgue  $m_2$  e:

$$P^{(X,Y)}(B) = \mathbb{P}((X, Y) \in B) = \int_B f_{(X,Y)}(x, y) dx dy \quad \forall B \in \mathcal{B}^2$$

Inoltre si può affermare che:

$$\dim(B) < 2 \implies \mathbb{P}((X, Y) \in B) = 0$$

- $X$  e  $Y$  sono continue:

$$f_X(x) = \int_{-\infty}^{+\infty} f_{(X,Y)}(x, y) dy \quad \text{e} \quad f_Y(y) = \int_{-\infty}^{+\infty} f_{(X,Y)}(x, y) dx$$

**Densità**  $f_{(X,Y)} : \mathbb{R}^2 \rightarrow \mathbb{R}$  è una densità se:

1.  $f_{(X,Y)}$  è boreliana
2.  $f_{(X,Y)} \geq 0$
3.  $\int_{\mathbb{R}^2} f_{(X,Y)}(x, y) dx dy = 1$

**Valore atteso** Data  $h : \mathbb{R}^2 \rightarrow [0, +\infty)$  boreliana oppure  $h : \mathbb{R}^2 \rightarrow \mathbb{R}$  con  $h \in L^1(\mathbb{R}^2, \mathcal{B}^2, P^{(X,Y)})$ , allora:

$$\mathbb{E}[h(X, Y)] = \int_{\mathbb{R}^2} h(x, y) f_{(X,Y)}(x, y) dx dy$$

**Indipendenza**  $X \perp\!\!\!\perp Y$ :

$$\iff P^{(X,Y)} = P^X \otimes P^Y$$

$$\iff f_{(X,Y)}(x, y) \stackrel{\text{qc}}{=} f_X(x)f_Y(y)$$

$$\iff f_{(X,Y)}(x, y) \stackrel{\text{qc}}{=} h_1(x)h_2(y)$$

$$\implies S_{(X,Y)} = S_X \times S_Y$$

**Proposizione 10.14** (*J-P Th 12.6, 12, 7*)

Sia  $(X, Y) : \Omega \rightarrow \mathbb{R}^2$  vettore aleatorio continuo con densità  $f_{(X,Y)}$ , il supporto di  $P^{(X,Y)}$  è  $S \subseteq \mathbb{R}^2$ . Sia data  $h : \mathbb{R}^2 \rightarrow \mathbb{R}^2$  definita come:

$$(U, V) = h((X, Y)) = (h_1(X, Y), h_2(X, Y))$$

Siano inoltre valide le seguenti proprietà:

1.  $h \in C^1(S)$
2.  $\det(J_h)(x, y) \neq 0$
3.  $\exists g = h^{-1} : g : h(S) \rightarrow S \quad g(u, v) = (g_1(u, v), g_2(u, v))$

Allora:

1.  $(U, V)$  è un vettore aleatorio continuo
2.  $f_{(U,V)}(u, v) = f_{(X,Y)}(g_1(u, v), g_2(u, v)) |\det(J_g)(u, v)|$

Si procede ora con un cenno di dimostrazione<sup>34</sup>.

**Dimostrazione**

Si può mostrare che  $\mathbb{P}((U, V) \in B) \stackrel{?}{=} \int_B f_{(U,V)}(u, v) du dv$ :

$$\begin{aligned} \mathbb{P}((U, V) \in B) &= \mathbb{P}(h((X, Y)) \in B) \\ &= \mathbb{P}((X, Y) \in g(B)) \\ &= \int_{g(B)} f_{(X,Y)}(x, y) dx dy \quad (\text{cambiando la variabile}) \\ &= \int_B f_{(U,V)}(u, v) du dv \quad \square \end{aligned}$$

**Osservazione 10.15**

$f_{(X,Y)}$  contiene  $\mathbf{1}_S(x, y) \iff f_{(U,V)}$  contiene  $\mathbf{1}_S(g_1(x, y), g_2(x, y)) = \mathbf{1}_{h(S)}(u, v)$

<sup>34</sup>Lasciata come esercizio allo studente più motivato di noi.

**Osservazione 10.16**

È possibile trovare l'inversa di  $h$ ,  $g = h^{-1}$ :

$$(u, v) \in h(S) \implies \begin{cases} u = h_1(x, y) \\ v = h_2(x, y) \end{cases} \implies \begin{cases} x = g_1(u, v) \\ y = g_2(u, v) \end{cases} \implies \exists! (x, y) \in S$$

Si richiede l'unicità dell'inversa.

**10.4 Covarianza**

Data la VA  $X : (\Omega, \mathcal{A}, \mathbb{P}) \rightarrow (\mathbb{R}, \mathcal{B})$  con legge  $P^X$ , è possibile sintetizzare le informazioni date dalla legge con valore medio  $\mathbb{E}[X]$  e varianza  $Var(X)$ . Per ottenere lo stesso risultato con un vettore aleatorio a due componenti  $(X, Y) : (\Omega, \mathcal{A}, \mathbb{P}) \rightarrow (\mathbb{R}^n, \mathcal{B}^n)$  con legge  $P^{(X, Y)}$  è necessario aggiungere un valore che rappresenti il legame tra le due VA, ovvero di quanto le due variabili siano l'una dall'altra dipendenti.

**Definizione 10.17**

Siano  $X$  e  $Y$  VAR su  $(\Omega, \mathcal{A}, \mathbb{P})$  e in  $L^2(\mathbb{P})$ . Si definisce **covarianza** il seguente valore:

$$Cov(X, Y) := \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]$$

La covarianza è ben definita perché la funzione dentro il valore atteso è prodotto di due VA che sono  $L^2$  per ipotesi; tale prodotto è dunque  $L^1$  per il teorema di Cauchy-Schwarz.

Applicando opportunamente la regola del valore atteso si ottengono, per ogni  $X$  e  $Y$ , i seguenti risultati:

$$Cov(X, X) = Var(X) \quad \text{e} \quad Cov(X, Y) = Cov(Y, X) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y].$$

Una covarianza positiva indica che a valori di  $X$  sopra la media corrispondono *prevalentemente* (questo indica il valore atteso) valori di  $Y$  sopra la media, e viceversa. I valori del vettore  $(X, Y)$ , se visualizzati su un grafico bidimensionale, formano dunque una “macchia” allungata parallelamente alla bisettrice I-III quadrante. Se la covarianza è negativa, la distribuzione degli  $(X, Y)$  è invece allungata nella direzione della bisettrice II-IV quadrante (figura 10.2).

**10.4.1 Indipendenza e scorrelazione****Teorema 10.18** (*J-P Th 12.3*)

Siano  $X, Y \in L^2(\mathbb{P})$  tali che  $X \perp\!\!\!\perp Y$ . Allora  $Cov(X, Y) = 0$ .

Due variabili che hanno covarianza nulla si dicono *scorrelate*.

**Dimostrazione**

Per  $X, Y$  limitate (che infatti sono in  $L^2$ ):

$$Cov(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] \stackrel{X \perp\!\!\!\perp Y}{=} \mathbb{E}[X]\mathbb{E}[Y] - \mathbb{E}[X]\mathbb{E}[Y] = 0.$$

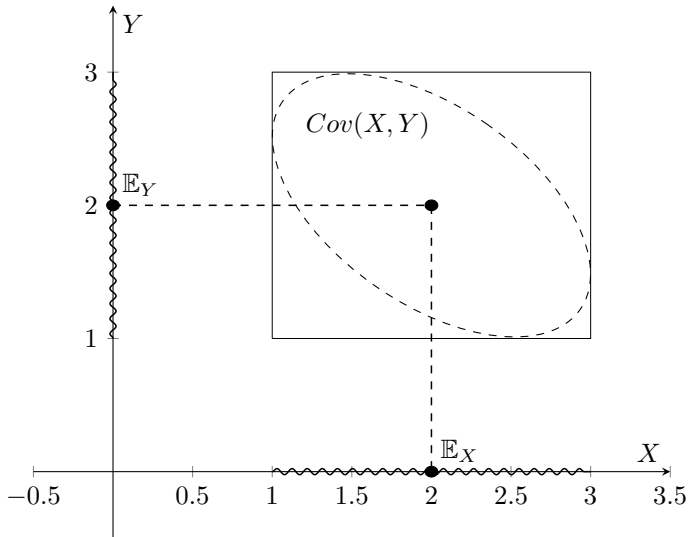


Figura 10.2: covarianza (negativa) per due VA

Per  $X, Y$  illimitate l'uguaglianza è comunque verificata costruendo due successioni crescenti  $X_n, Y_n$  che tendono, rispettivamente, a  $X$  e a  $Y$ ; la convergenza dominata garantisce la possibilità di scambiare limite e valore atteso.  $\square$

**NB:** In generale, il viceversa del teorema *non è vero!*

$$\text{Cov}(X, Y) = 0 \not\Rightarrow X \perp\!\!\!\perp Y$$

L'indipendenza è una condizione più forte della correlazione; ovvero, esistono coppie di variabili scorrelate ma non indipendenti. Più avanti entreremo più in dettaglio sul significato della (s)correlazione.

*Esempio.* Consideriamo una probabilità concentrata su 4 punti, ovvero:

$$(X, Y) \sim U(\{(\pm 1, 0), (0, \pm 1)\})$$

La figura 10.3 è simmetrica rispetto all'origine, nel senso che non c'è allungamento in una delle direzioni diagonali, quindi  $\text{Cov}(X, Y) = 0$  (risultato che, peraltro, si può anche verificare con rapidi conti). Tuttavia  $X \not\perp\!\!\!\perp Y$  perché la figura non forma un rettangolo. Infatti, prendendo una sezione qualsiasi dell'"area" in cui è condensata la probabilità, si osserva immediatamente che conoscere un'informazione su una delle due variabili (per esempio  $X = 1$ ) può influenzare il grado di fiducia sull'altra (per esempio è vero che  $\mathbb{P}(Y = 0 | X = 1) = 1$ ).



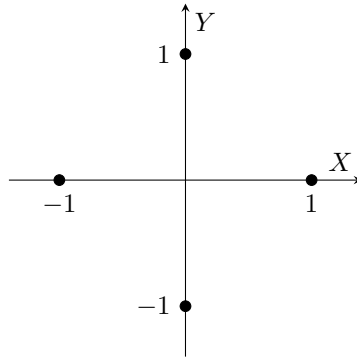


Figura 10.3: VA  $(X, Y)$  concentrata su 4 punti

Due VA sono indipendenti se e solo se il grafico del supporto (ovvero l'insieme dei punti in cui è condensata la probabilità) forma un rettangolo, sia esso discreto (come in questo caso) o continuo. Infatti prendendo una sezione del rettangolo, ovvero fissando un valore di una delle due VA, il grado di fiducia dell'altra VA non cambia perché il segmento-sezione ha ugual lunghezza lungo tutto il supporto.

*Esempio.* Consideriamo il cerchio  $S = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq 1\}$  e il vettore aleatorio  $(X, Y) \sim U(S)$ .

Notiamo immediatamente che  $(X, Y)$  ha probabilità concentrata su  $S$ .

Anche qui  $Cov(X, Y) = 0$  per simmetria centrale, ma  $X \not\perp Y$  perché il supporto non è un rettangolo. Verifichiamo questa conclusione mostrando un punto in cui le probabilità non si fattorizzano, per esempio:

$$\mathbb{P}\left(Y \geq \frac{1}{2} \mid X = 1\right) = 0 \neq \mathbb{P}\left(Y \geq \frac{1}{2}\right)$$

### Proposizione 10.19

$$|Cov(X, Y)| \leq \sqrt{Var(X) Var(Y)}$$

La proposizione è un'immediata conseguenza della disuguaglianza di Cauchy-Schwarz.

#### 10.4.2 Coefficiente di correlazione lineare

##### Definizione 10.20

Siano  $X$  e  $Y$  VAR su  $(\Omega, \mathcal{A}, \mathbb{P}) \rightarrow (\mathbb{R}, \mathcal{B})$  in  $L^2(\mathbb{P})$  e tali che  $Var(X) > 0$  e

$Var(Y) > 0$ . Si definisce **coefficiente di correlazione lineare** il valore:

$$\rho := \frac{Cov(X, Y)}{\sqrt{Var(X) Var(Y)}}$$

Sappiamo che  $|\rho| \leq 1$  per la proposizione precedente e che  $X \perp Y \implies \rho = 0$  per il teorema 10.18.

Notiamo inoltre che  $\rho$  è adimensionale in quanto indica un rapporto tra grandezze omogenee, al contrario di media, varianza e altri oggetti probabilistici.

### Proposizione 10.21

Siano  $X, Y \in L^2(\mathbb{P})$  tali che  $Var(X) > 0$  e  $Var(Y) > 0$ .

Si ha  $|\rho| = 1$  se e solo se esistono  $a, b \in \mathbb{R}, a \neq 0$ , tali che  $Y = aX + b$ . In tal caso, vale:

$$Y = \rho \sqrt{\frac{Var(Y)}{Var(X)}} X + \mathbb{E}[Y] - \rho \sqrt{\frac{Var(X)}{Var(Y)}} \mathbb{E}[X]$$

Il risultato (1) è coerente con quanto precedentemente detto sulla covarianza. Una covarianza positiva (o negativa) indica infatti una distribuzione di punti stretta e allungata nella direzione I-III (o II-IV) quadrante, e il caso  $|\rho| = 1$  indica il caso estremo di questo stringimento e allungamento: tutti i punti si allineano su una retta, che, per inciso, permette di calcolare deterministicamente il valore di  $Y = aX + b = h(X)$ ; questo infatti è il caso opposto rispetto all'indipendenza tra due VA (dove infatti  $Cov(X, Y) = 0$ ).

### Dimostrazione

Sarà dimostrato solo il punto (2). Sapendo che da  $Y = aX + b$  discendono  $Var(Y) = a^2 Var(X)$  e  $\mathbb{E}[Y] = a\mathbb{E}[X] + b$ , è sufficiente sostituire nella formula queste due espressioni per verificare facilmente che danno un'identità.  $\square$

**NB:** I casi in cui  $Y$  è espresso come una retta orizzontale o verticale (rispetto a un sistema di riferimento  $OXY$ ) andrebbero studiati a parte, ma sono comunque di facile trattazione e non costituiscono dunque un ostacolo alla generalità della proposizione.

#### 10.4.3 Proprietà

### Proposizione 10.22

La covarianza è *bilineare*, ovvero lineare su entrambi gli argomenti:

$$\begin{aligned} Cov(aX + bY, cZ + dW) &= \\ &= ac \cdot Cov(X, Z) + ad \cdot Cov(X, W) + bc \cdot Cov(Y, Z) + bd \cdot Cov(Y, W) \end{aligned}$$

La dimostrazione è banale ma non immediata: si effettua con noiosi conti su  $Cov(aX + bY, Z)$ . Per simmetria la proposizione sarà dunque vera anche per l'altro argomento, senza necessità di complicare ulteriormente i calcoli.

**Corollario 10.23**

$$Var(X + Y) = Var(X) + Var(Y) + 2Cov(X, Y)$$

La dimostrazione è banale e immediata calcolando  $Var(X + Y)$  come  $Cov(X + Y, X + Y)$ .

**10.5 Vettore media e matrice varianza**

**Definizione 10.24**

Sia  $X = (X_1, \dots, X_n) : (\Omega, \mathcal{A}, \mathbb{P}) \rightarrow (\mathbb{R}^n, \mathcal{B}^n)$  un vettore aleatorio.

Il **vettore media** di  $X$ , se le componenti sono  $L^1$ , è il vettore delle medie delle componenti:

$$\mathbb{E}[X] := \begin{bmatrix} \mathbb{E}[X_1] \\ \vdots \\ \mathbb{E}[X_n] \end{bmatrix}$$

e, se le componenti sono  $L^2$ , la **matrice varianza**, o matrice delle covarianze, di  $X$  è:

$$C := \begin{bmatrix} Var(X_1) & \dots & Cov(X_1, X_n) \\ \vdots & \ddots & \vdots \\ Cov(X_n, X_1) & \dots & Var(X_n) \end{bmatrix} \quad \text{ovvero } C_{ij} = Cov(X_i, X_j)$$

**Teorema 10.25 (J-P Th 12.4)**

Sia  $X$  vettore aleatorio con componenti  $X_1, \dots, X_n \in L^2$ .

Allora  $C$  è *semi-definita positiva*<sup>35</sup> (indicato con  $C \geq 0$ ), ovvero:

$$C = C^T \quad \text{e} \quad a^T C a \geq 0 \quad \forall a = (a_1, \dots, a_n) \in \mathbb{R}^n$$

**Dimostrazione**

La simmetria è ovvia perché  $Cov(X, Y) = Cov(Y, X)$ .

---

<sup>35</sup>Si noti che in questo corso si include la simmetria della matrice nella definizione di semi-definita positività, sebbene in altri testi e corsi sia trattata come una proprietà a parte.

Grazie alla bilinearità della covarianza si ha:

$$\begin{aligned}
 a^T C a &= \sum_{i,j} a_i C_{ij} a_j && \text{(scomponendo i vettori)} \\
 &= \sum_{i,j} a_i \text{Cov}(X_i, X_j) a_j && \text{(passando alle covarianze)} \\
 &= \sum_i \sum_j \text{Cov}(a_i X_i, a_j X_j) && \text{(per bilinearità)} \\
 &= \text{Cov} \left( \sum_i a_i X_i, \sum_j a_j X_j \right) && \text{(per bilinearità)} \\
 &= \text{Var} \left( \sum_i a_i X_i \right) \geq 0 \quad \forall a \in \mathbb{R}^n && \square
 \end{aligned}$$

**Teorema 10.26**

Sia  $X$  vettore aleatorio con componenti  $X_1, \dots, X_n \in L^2$ . Allora:

$$X \in \text{range}(C) + \mathbb{E}[X] \quad \text{qc}$$

Qui  $\text{range}(C) = \text{col}(C)$  è lo spazio colonna di  $C$ , ovvero lo spazio vettoriale avente come base le colonne di  $C$ .

Lo spazio indicato dal teorema è uno *spazio affine* di  $\mathbb{R}^n$ , ovvero ottenuto mediante una trasformazione lineare (la matrice  $C$ ) e una traslazione (l'aggiunta di  $\mathbb{E}[X]$ ). Si noti inoltre che è possibile che  $X$  non possa assumere tutti i valori di  $\mathbb{R}^n$ ; ciò succede nel caso in cui due o più colonne di  $C$  sono linearmente dipendenti, ovvero se la matrice non ha rango massimo.

Questo teorema ha un'importante conseguenza.

**Corollario 10.27**

Se  $X$  è un vettore aleatorio continuo, allora la sua matrice varianza  $C$  è invertibile; equivalentemente, se  $C$  non è invertibile allora  $X$  non è continuo.

Si noti che questa condizione è solo necessaria e non sufficiente. Per esempio, si può verificare che il controesempio mostrato a pagina 130, cioè  $W = (X, X^2)$  con  $X \sim \mathcal{N}(0, 1)$ , ha matrice varianza  $C$  invertibile pur non essendo continuo.

**Teorema 10.28** (*J-P Th 12.5*)

Siano  $X : (\Omega, \mathcal{A}, \mathbb{P}) \rightarrow (\mathbb{R}^n, \mathcal{B}^n)$  un vettore aleatorio con varianza  $C_X$ , e  $Y = AX + b$  un vettore aleatorio, con  $A \in \mathbb{R}^{m \times n}$  e  $b \in \mathbb{R}^m$ . Allora:

$$\mathbb{E}[Y] = A \mathbb{E}[X] + b \quad \text{e} \quad C_Y = A C_X A^T$$

Quest'ultima formula è molto utile in contesti pratici per ridurre le dimensioni dei vettori, al fine di facilitare i conti.

### Dimostrazione

1. Calcoliamo la  $i$ -esima componente del vettore media sfruttando la linearità del valore atteso:

$$\mathbb{E}[Y_i] = \mathbb{E} \left[ \sum_j a_{ij} X_j + b_i \right] = \sum_j a_{ij} \mathbb{E}[X_j] + b_i = (A \mathbb{E}[X] + b)_i$$

2. Per quanto riguarda la matrice varianza:

$$\begin{aligned} (C_Y)_{ij} &= \text{Cov}(Y_i, Y_j) \\ &= \text{Cov} \left( \sum_k a_{ik} X_k + b_i, \sum_m a_{jm} X_m + b_j \right) \\ &= \sum_{k,m} a_{ik} a_{jm} (C_X)_{ij} && \text{(bilinearità)} \\ &= (A C_X A^T)_{ij} && \square \end{aligned}$$



---

## 11 Funzione caratteristica

Dalla teoria sviluppata fin qui, sappiamo che una probabilità su  $(\mathbb{R}, \mathcal{B})$  è caratterizzata dalla funzione di ripartizione e dalla densità (sia essa discreta o continua), e che la densità caratterizza anche una probabilità nel caso multidimensionale  $(\mathbb{R}^n, \mathcal{B}^n)$ . In questo capitolo introdurremo un nuovo modo di caratterizzare la probabilità, la *funzione caratteristica*. Vedremo che questo strumento presenta una serie di notevoli vantaggi: permette di trattare indistintamente i casi continui e discreti, semplifica il calcolo dei valori attesi e facilita lo studio di alcune aree d'interesse come l'indipendenza, la continuità e la combinazione lineare di variabili aleatorie.

### 11.1 Definizioni

Riprendiamo la definizione di prodotto scalare e introduciamo alcune notazioni alternative:

$$\langle X|Y \rangle = \langle X, Y \rangle = X \cdot Y = (X, Y) := \sum_{k=1}^n X_k Y_k$$

Tra le varie notazioni, in questo testo si è scelto di utilizzare la prima<sup>36</sup>:  $\langle X|Y \rangle$ .

#### Definizione 11.1

Sia  $\mathbb{P}$  una probabilità su  $(\mathbb{R}, \mathcal{B})$ , la **funzione caratteristica** (o **trasformata di Fourier**) di  $\mathbb{P}$  è la funzione:

$$\hat{\mathbb{P}} = \varphi : \mathbb{R}^n \rightarrow \mathbb{C} \quad \varphi(u) := \int_{\mathbb{R}^n} e^{i\langle u|x \rangle} \mathbb{P}(dx)$$

Dall'analisi sappiamo che  $z = e^{i\vartheta} = \cos(\vartheta) + i \sin(\vartheta)$ , che è forma esponenziale del numero complesso, con  $\vartheta \in \mathbb{R}$   $|e^{i\vartheta}| = 1 \quad \forall \vartheta \in \mathbb{R}$

Ne discende che:

$$e^{i\langle u|x \rangle} = \cos(\langle u|x \rangle) + i \sin(\langle u|x \rangle), \quad |e^{i\langle u|x \rangle}| = 1, \quad \forall u \quad \forall x$$

Si consulti l'appendice A a pagina 257 per eventuali delucidazioni circa l'integrazione di variabili complesse.

Si ricorda che vale la *disuguaglianza triangolare* estesa agli integrali:

$$\left| \int h d\mathbb{P} \right| \leq \int |h| d\mathbb{P}$$

Si fa inoltre presente che  $\exists \varphi(u) \quad \forall u$  perché  $e^{i\langle u|x \rangle} \in \mathcal{L}^1(\mathbb{P}) \quad \forall u$  in quanto  $|e^{i\langle u|x \rangle}| = 1 \in \mathcal{L}^1(\mathbb{P}) \quad \forall u$ .

---

<sup>36</sup>Contro il volere del popolo, in quanto è quella dei fisici

**Definizione 11.2**

Sia  $X : (\Omega, \mathcal{A}, \mathbb{P}) \rightarrow (\mathbb{R}^n, \mathcal{B}^n)$  vettore aleatorio di legge  $P^X$ . La funzione caratteristica di  $X$  è  $\widehat{P}^X = \varphi_X : \mathbb{R}^n \rightarrow \mathbb{C}$  definita come:

$$\varphi_X(u) = \int_{\mathbb{R}^n} e^{i\langle u|x \rangle} P^X(dx) = \mathbb{E} \left[ e^{i\langle u|x \rangle} \right] = \int_{\Omega} e^{i\langle u|x \rangle} d\mathbb{P}$$

*Esempio.*  $X \sim Po(\lambda)$ , con  $\lambda > 0$  (VA con distribuzione Poisson). Ricaviamo la sua funzione caratteristica:

$$\begin{aligned} \varphi_X(u) &= \mathbb{E}[e^{iu x}] = \sum_{k=0}^{+\infty} e^{iuk} e^{-\lambda} \frac{\lambda^k}{k!} && \text{(applicando la definizione)} \\ &= e^{-\lambda} \sum_{k=0}^{+\infty} \frac{(\lambda e^{iu})^k}{k!} && \text{(per def. di esponenziale complesso)} \\ &= e^{-\lambda} e^{\lambda e^{iu}} && \text{(raccogliendo)} \\ &= \exp\{\lambda(e^{iu} - 1)\} \end{aligned}$$

*Esempio.*  $X \sim \mathcal{E}(\lambda)$ , con  $\lambda > 0$  (VA aleatoria con distribuzione esponenziale). Ricaviamo la sua funzione caratteristica:

$$\begin{aligned} \varphi_X(u) &= \mathbb{E}[e^{iu x}] = \int_{-\infty}^{+\infty} e^{iut} f_X(t) dt && \text{(applicando la definizione)} \\ &= \int_0^{+\infty} e^{iut} \lambda e^{-\lambda t} dt && \text{(per linearità)} \\ &= \lambda \int_0^{+\infty} e^{t(iu-\lambda)} dt \\ &= \lambda \left. \frac{e^{t(iu-\lambda)}}{iu-\lambda} \right|_0^{+\infty} && (e^{iut} \text{ limitata}) \\ &= \frac{\lambda}{\lambda - iu} \end{aligned}$$

Nell'ultima uguaglianza si è sfruttato il fatto che in un prodotto tra un esponenziale con argomento una quantità negativa (infinitesimo) e un esponenziale con argomento complesso (limitato) vince il primo.

## 11.2 Proprietà

**Teorema 11.3** (*J-P Th 14.1*)

La funzione caratteristica  $\widehat{\mathbb{P}}$  di una probabilità  $\mathbb{P}$  su  $(\mathbb{R}^n, \mathcal{B}^n)$  caratterizza  $\mathbb{P}$ , ovvero:

$$\widehat{\mathbb{P}}(u) = \widehat{\mathbb{Q}}(u) \quad \forall u \in \mathbb{R}^n \iff \mathbb{P}(B) = \mathbb{Q}(B) \quad \forall B \in \mathcal{B}^n$$



In particolare, dato un vettore aleatorio  $X$ , la funzione caratteristica  $\varphi_X$  caratterizza la legge  $P^X$ .

**Teorema 11.4: proprietà analitiche di  $\varphi$**  (*J-P Th 13.1*)

Data  $\varphi$  funzione caratteristica di una qualsiasi probabilità  $\mathbb{P}$  su  $(\mathbb{R}^n, \mathcal{B}^n)$ , allora è vero che:

1.  $\varphi$  continua;
2.  $|\varphi| \leq 1$ ;
3.  $\varphi(0) = 1$ .

**Dimostrazione**

$$3. \varphi(0) = \int_{\mathbb{R}^n} e^{i\langle 0|x \rangle} \mathbb{P}(dx) = \mathbb{P}(\mathbb{R}^n) = 1.$$

$$2. |\varphi(u)| = \left| \int_{\mathbb{R}^n} e^{i\langle u|x \rangle} \mathbb{P}(dx) \right| \leq \int_{\mathbb{R}^n} |e^{i\langle u|x \rangle}| \mathbb{P}(dx) = 1 \quad \forall u.$$

$$1. l \rightarrow +\infty \implies u_l \rightarrow u \implies \varphi(u_l) \rightarrow \varphi(u) \text{ dove } \varphi(u_l) = \int_{\mathbb{R}^n} e^{i\langle u_l|x \rangle} \mathbb{P}(dx).$$

Per quest'ultimo punto occorre tener presente che:

- $\langle u_l|x \rangle \rightarrow \langle u|x \rangle$  per continuità del prodotto scalare  $\forall x$ ;
- $e^{i\langle u_l|x \rangle} \rightarrow e^{i\langle u|x \rangle}$  per continuità dell'esponenziale  $\forall x$ ;
- $l \rightarrow +\infty \implies \int_{\mathbb{R}^n} e^{i\langle u_l|x \rangle} \mathbb{P}(dx) = \varphi(u)$  per convergenza dominata.  $\square$

**11.2.1 Calcolo dei momenti**

**Definizione 11.5**

Si dice **momento  $n$ -esimo** di una variabile aleatoria  $X$  il valore atteso della sua  $n$ -esima potenza  $\mathbb{E}[X^n]$ .

**Teorema 11.6** (*J-P Th 13.2*)

Sia  $X = (X_1, \dots, X_n) : (\Omega, \mathcal{A}, \mathbb{P}) \rightarrow (\mathbb{R}^n, \mathcal{B}^n)$  vettore aleatorio.

Se  $\exists n \in \mathbb{N}$  tale per cui  $x_k \in L^m(\mathbb{P}) \quad \forall k = 1, \dots, n$ , allora  $\varphi_X \in \mathcal{C}^m(\mathbb{R}^n, \mathbb{C})$  e vale:

$$\frac{\partial^m}{\partial u_{k_1} \dots \partial u_{k_m}} \varphi(u) = i^m \mathbb{E}[X_{k_1} \dots X_{k_m} e^{i\langle u|x \rangle}] \quad \forall u \in \mathbb{R}^n$$

Si tratta dunque di uno *scambio di limiti* fra la derivata e il valore atteso, ovvero, euristicamente:

$$\frac{\partial^n}{\partial u_{k_1} \dots \partial u_{k_n}} \mathbb{E}[\cdot] = \mathbb{E} \left[ \frac{\partial^n}{\partial u_{k_1} \dots \partial u_{k_n}} \cdot \right]$$

Inoltre è utile notare che con questo teorema è possibile calcolare il momento ricorrendo solo alle derivate senza bisogno dell'integrale.

**Osservazione 11.7**

In particolare, il valore  $\varphi(0)$  può essere molto utile per ricavare direttamente i momenti:

$$\frac{\partial^m}{\partial u_{k_1} \cdots \partial u_{k_m}} \varphi(0) = i^m \mathbb{E}[X_{k_1} \cdots X_{k_m}]$$

**Osservazione 11.8**

Sia  $X_k \in L^2 \subset L^1 \forall k$ . Allora:

- $\mathbb{E}[X_k] = \frac{1}{i} \frac{\partial}{\partial u_k} \varphi(0)$ ;
- $\mathbb{E}[X_k^2] = -\frac{\partial^2}{\partial u_k^2} \varphi(0)$ , da cui si può agevolmente calcolare  $Var(X_k)$ ;
- $\mathbb{E}[X_k X_j] = -\frac{\partial^2}{\partial u_k \partial u_j} \varphi(0)$ , da cui si può agevolmente calcolare  $Cov(X_k, X_j)$ .

*Esempio.* Sia  $X \sim \mathcal{E}(\lambda) \lambda > 0$  (VA con distribuzione esponenziale). Da un esempio precedente, sappiamo che:  $X \in L^m \forall m > 1$  e  $\varphi_X(u) = \frac{\lambda}{\lambda - iu}$ . Allora:

$$\begin{aligned} \mathbb{E}[X] &= \frac{1}{i} \varphi'(0) = \frac{1}{i} \frac{d}{du} \left( \frac{\lambda}{\lambda - iu} \right) \Big|_{u=0} = \frac{1}{i} \lambda (-1) \frac{-i}{(\lambda - iu)^2} \Big|_{u=0} \\ &= \frac{\lambda}{(\lambda - iu)^2} \Big|_{u=0} = \frac{1}{\lambda} \end{aligned}$$

$$\begin{aligned} \mathbb{E}[X^2] &= -\varphi''(0) = -\frac{d}{du} \left[ \frac{\lambda i}{(\lambda - iu)^2} \right] \Big|_{u=0} = -\lambda i (-2) \frac{1}{(\lambda - iu)^3} (-i) \Big|_{u=0} \\ &= \frac{2\lambda}{(\lambda - iu)^3} \Big|_{u=0} = \frac{2}{\lambda^2} \end{aligned}$$

$$Var(X) = \mathbb{E}[X^2] - \mathbb{E}^2[X] = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}$$

*Esempio.* Calcoliamo la funzione caratteristica della distribuzione normale standard tramite il teorema dei momenti.

Sia  $X \sim \mathcal{N}(0, 1)$  allora vale che  $X \in L^p \forall p \geq 1$ . Sostituendo nella definizione di funzione caratteristica la densità della normale:

$$\varphi(u) = \mathbb{E}[e^{iux}] = \int_{\mathbb{R}} e^{iut} \frac{e^{-\frac{t^2}{2}}}{\sqrt{2\pi}} dt = \int_{\mathbb{R}} \cos(ut) \frac{e^{-\frac{t^2}{2}}}{\sqrt{2\pi}} dt + i \int_{\mathbb{R}} \sin(ut) \frac{e^{-\frac{t^2}{2}}}{\sqrt{2\pi}} dt$$

Si osserva che il secondo integrale è nullo in quanto l'integranda è una funzione dispari: si perde perciò la parte immaginaria.

Per risolvere quest'integrale si usa un trucco: applicando il teorema dei momenti, si ottiene la derivata  $\varphi'(u)$  in funzione di  $\varphi(u)$ :

$$\begin{aligned}\varphi'(u) &= i \mathbb{E} [x e^{iux}] = i \int_{\mathbb{R}} t e^{iut} \frac{e^{-\frac{t^2}{2}}}{\sqrt{2\pi}} dt \\ &= i \left( \underbrace{\int_{\mathbb{R}} t \cos(ut) \frac{e^{-\frac{t^2}{2}}}{\sqrt{2\pi}} dt}_{\text{funzione dispari}} + i \int_{\mathbb{R}} t \sin(ut) \frac{e^{-\frac{t^2}{2}}}{\sqrt{2\pi}} dt \right) \\ &= - \int_{\mathbb{R}} t \sin(ut) \frac{e^{-\frac{t^2}{2}}}{\sqrt{2\pi}} dt \\ &= - \frac{1}{\sqrt{2\pi}} \left\{ -e^{-\frac{t^2}{2}} \sin(ut) \Big|_{-\infty}^{+\infty} + \int_{-\infty}^{+\infty} e^{-\frac{t^2}{2}} \cos(ut) u dt \right\} = -u\varphi(u)\end{aligned}$$

L'integrale si è così trasformato in un problema di Cauchy ben posto, in cui la condizione iniziale è dettata dal teorema 11.4:

$$\begin{cases} \varphi'(u) = -u\varphi(u) \\ \varphi(0) = 1 \end{cases} \implies \varphi(u) = e^{-\frac{u^2}{2}}$$

È immediato verificare che vale l'osservazione 11.7, infatti il valore atteso è:

$$\mathbb{E}[X] = \frac{1}{i} \varphi'(0) = 0 \cdot \varphi(0) = 0$$

Il calcolo della varianza è altrettanto semplice:

$$\text{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}^2[X] = -\varphi''(0) = \varphi(0) - 0^2 \cdot \varphi(0) = 1$$

Visti i risultati negli esempi precedenti dove la parte immaginaria della funzione caratteristica è andata annullandosi, si potrebbe pensare di definire una funzione caratteristica senza i numeri complessi.

### Definizione 11.9

La funzione  $M_X : \mathbb{R}^n \rightarrow \mathbb{R}$  tale che:

$$M_X(u) = \mathbb{E} [e^{\langle u, X \rangle}] = \int_{\mathbb{R}^n} e^{\langle u, X \rangle} P^X(dx)$$

è detta **funzione generatrice dei momenti**.

Dalla definizione segue immediatamente che:

$$\frac{\partial^m}{\partial u_{k_1} \cdots \partial u_{k_m}} M_X(0) = \mathbb{E}[X_{k_1} \cdots X_{k_m}]$$

Non considerando i numeri complessi, in tutti i casi in cui questi non si annullano l'integrale non esiste oppure la funzione non può essere definita su tutto  $\mathbb{R}^n$ .

### 11.2.2 Trasformazioni affini

**Teorema 11.10** (*J-P Th 13.3*)

Sia il vettore aleatorio  $X : (\Omega, \mathcal{A}, \mathbb{P}) \rightarrow (\mathbb{R}^n, \mathcal{B}^n)$ .

Siano inoltre  $A \in \mathbb{R}^{m \times n}$ ,  $b \in \mathbb{R}^m$ , e il vettore aleatorio trasformato  $Y = AX + b : (\Omega, \mathcal{A}, \mathbb{P}) \rightarrow (\mathbb{R}^m, \mathcal{B}^m)$ , allora:

$$\varphi_Y(u) = e^{i\langle u|b \rangle} \varphi_X(A^T u) \quad \forall u \in \mathbb{R}^m$$

#### Dimostrazione

$$\begin{aligned} \varphi_Y(u) &= \mathbb{E} \left[ e^{i\langle u|Y \rangle} \right] = \mathbb{E} \left[ e^{i\langle u|AX+b \rangle} \right] && \text{(per definizione)} \\ &= e^{i\langle u|b \rangle} \mathbb{E} \left[ e^{i\langle u|AX \rangle} \right] && \text{(per linearità di } \mathbb{E} \text{)} \\ &= e^{i\langle u|b \rangle} \mathbb{E} \left[ e^{i\langle A^T u|X \rangle} \right] && \text{(prop. del prodotto scalare)} \\ &= e^{i\langle u|b \rangle} \varphi_X(A^T u) \end{aligned} \quad \square$$

*Esempio:* funzione caratteristica di una variabile aleatoria normale non standard. Sia  $X \sim N(\mu, \sigma^2)$ . Allora  $X = \mu + \sigma Z$ , con  $Z = \frac{X-\mu}{\sigma} \sim N(0, 1)$ . Dunque:

$$\varphi_X(u) = e^{iu\mu} \varphi_Z(\sigma u) = e^{iu\mu} e^{-\frac{\sigma^2}{2} u^2} = e^{iu\mu - \frac{\sigma^2}{2} u^2}$$

*Esempio:* funzione caratteristica della somma di variabili aleatorie.

Siano  $X = (X_1, \dots, X_n) \sim \varphi_X$  e  $Y = \sum_{k=1}^n X_k$ . Calcoliamo le funzioni caratteristiche di  $Y$  e di  $X_k$ :

$$Y = \begin{bmatrix} 1, & \dots, & 1 \end{bmatrix} \begin{bmatrix} X_1 \\ \vdots \\ X_n \end{bmatrix} \quad \varphi_Y(u) = \varphi_X \left( \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} u \right) = \varphi_X(u, \dots, u)$$

Per estrarre una sola componente poniamo  $v = [0, \dots, 0, u, 0, \dots, 0]$ , dove la  $u$  occupa la  $k$ -esima posizione.

Allora  $uX_k = \langle v|X \rangle$  e dunque:

$$\varphi_{X_k}(u) = \mathbb{E}[e^{iuX_k}] = \mathbb{E}[e^{i\langle v|X \rangle}] = \varphi_X(v) = \varphi_X(0, \dots, 0, u, 0, \dots, 0)$$

## 11.3 Risultati notevoli

Riportiamo i risultati ottenuti dagli esercizi omettendo i passaggi.

### 11.3.1 Alcune funzioni caratteristiche notevoli

- **Bernoulli:**

$$X \sim Be(p) \leftrightarrow \varphi_X(u) = pe^{iu} + 1 - p, \text{ con } p \in [0, 1]$$

- **Binomiale:**

$$X \sim Bi(n, p) \leftrightarrow \varphi_X(u) = (pe^{iu} + 1 - p)^n, \text{ con } p \in [0, 1] \text{ e } n \in \mathbb{N}$$

(Nota: in quanto somma di  $n$  bernoulliane iid, la binomiale ha la funzione caratteristica uguale al prodotto delle loro funzioni caratteristiche, come previsto dalla teoria.)

- **Uniforme simmetrica:**

$$X \sim U(-a, a) \leftrightarrow \varphi_X(u) = \frac{\sin(au)}{au} =: \text{sinc}(au), \text{ con } \varphi_X(0) = 1, a \in [0, +\infty)$$

Tale funzione è detta *seno cardinale*.

- **Uniforme generica:**

$$X \sim U(a, b) \leftrightarrow \varphi_X(u) = e^{i\frac{a+b}{2}u} \text{sinc}\left(\frac{b-a}{2}u\right), \text{ con } -\infty < a < b < +\infty$$

(Nota: è stata ottenuta come trasformazione affine del caso precedente.)

Si rimanda alle appendici B e C di pagina 260 e 268 per una lista più completa delle funzioni caratteristiche.

### 11.3.2 Somma di VA indipendenti

- **Bernoulliane.** Date  $X_1, \dots, X_n \sim Be(p)$  indipendenti:

$$X_1 + \dots + X_n \sim Bi(n, p)$$

- **Gaussiane.** Date  $X_1, \dots, X_n \sim \mathcal{N}(0, 1)$  indipendenti:

$$X_1 + \dots + X_n \sim \mathcal{N}(0, n) \quad \text{e} \quad X_1^2 + \dots + X_n^2 \sim \chi^2(n)$$

- **Esponenziali.** Date  $X_1, \dots, X_n \sim \mathcal{E}(\lambda)$  indipendenti:

$$X_1 + \dots + X_n \sim \Gamma(n, \lambda)$$

- **Gamma.** Date  $X \sim \Gamma(\alpha, \lambda), Y \sim \Gamma(\beta, \lambda)$  indipendenti:

$$X + Y \sim \Gamma(\alpha + \beta, \lambda)$$

- **Poisson.** Date  $X \sim Po(\lambda_X), Y \sim Po(\lambda_Y)$  indipendenti:

$$X + Y \sim Po(\lambda_X + \lambda_Y)$$

## 11.4 Continuità e indipendenza di vettori aleatori

In questo capitolo mostreremo che, se  $X$  è un vettore aleatorio *continuo*,  $X_k$  continuo  $\forall k$ , mentre in generale  $X_k$  continuo  $\forall k \not\Rightarrow X$  vettore aleatorio *continuo*, tranne in alcuni casi specifici.

### Teorema 11.11

Sia  $X = (X_1, \dots, X_n) : (\Omega, \mathcal{A}, \mathbb{P}) \rightarrow (\mathbb{R}^n, \mathcal{B}^n)$  vettore aleatorio tale che le componenti  $X_k$  sono indipendenti. Allora:

$$X \text{ continuo} \iff X_k \text{ continuo } \forall k$$

### Dimostrazione

( $\implies$ ): ovvio.

( $\impliedby$ ): dato un qualsiasi evento  $B_1 \times \dots \times B_n \in \mathcal{B}^n$ , possiamo fattorizzare la legge congiunta grazie all'indipendenza tra le VA:

$$\begin{aligned} P^X(B_1 \times \dots \times B_n) &= P^{X_1}(B_1) \dots P^{X_n}(B_n) \\ &= \int_{B_1} f_1(x_1) dx_1 \dots \int_{B_n} f_n(x_n) dx_n \\ &\stackrel{\text{F-T}}{=} \int_D f_1(x_1) \dots f_n(x_n) dx_1 dx_n \quad \square \end{aligned}$$

**NB:** La semplice continuità non è sufficiente ad affermare che le componenti continue implicano un vettore continuo; introducendo l'indipendenza rendiamo l'affermazione necessaria e sufficiente.

### Teorema 11.12

Siano  $X_1 : (\Omega, \mathcal{A}, \mathbb{P}) \rightarrow (\mathbb{R}^n, \mathcal{B}^n), X_2 : (\Omega, \mathcal{A}, \mathbb{P}) \rightarrow (\mathbb{R}^m, \mathcal{B}^m)$  vettori aleatori **indipendenti**. Allora:

$$X = (X_1, X_2) : (\Omega, \mathcal{A}, \mathbb{P}) \rightarrow (\mathbb{R}^{n+m}, \mathcal{B}^{n+m}) \text{ continuo} \iff X_1 \text{ e } X_2 \text{ continui}$$

### Teorema 11.13

Sia  $X = (X_1, \dots, X_n) : (\Omega, \mathcal{A}, \mathbb{P}) \rightarrow (\mathbb{R}^n, \mathcal{B}^n)$  vettore aleatorio. Allora:

$$\{X_k\}_{k=1}^n \text{ famiglia di VA } \perp \iff \varphi_X(u_1, \dots, u_n) = \prod_{k=1}^n \varphi_{X_k}(u_k) \quad \forall u \in \mathbb{R}^n$$

**Dimostrazione**

- ( $\implies$ ): per ipotesi,  $X_k$  sono indipendenti. Allora:

$$\begin{aligned}
 \varphi_X(u_1, \dots, u_n) &= \mathbb{E}[\exp\{i\langle u | X \rangle\}] \\
 &= \mathbb{E}\left[\exp\left\{i \sum_{k=1}^n u_k X_k\right\}\right] && \text{(esplicitando)} \\
 &= \mathbb{E}\left[\prod_{k=1}^n \exp\{i u_k X_k\}\right] && \text{(proprietà degli exp)} \\
 &= \prod_{k=1}^n \mathbb{E}[\exp\{i u_k X_k\}] && \text{(per indipendenza)} \\
 &= \prod_{k=1}^n \varphi_{X_k}(u_k)
 \end{aligned}$$

- ( $\impliedby$ ): sia  $Q = P^{X_1} \otimes \dots \otimes P^{X_n}$  probabilità su  $(\mathbb{R}^n, \mathcal{B}^n)$ . Allora:

$$\hat{Q}(u) = \prod_{k=1}^n \hat{P}^{X_k}(u_k) = \prod_{k=1}^n \varphi_{X_k}(u_k)$$

Notiamo che anche la legge congiunta si fattorizza allo stesso modo:

$$\hat{P}^X(u) = \varphi(u) = \prod_{k=1}^n \varphi_{X_k}(u_k) \implies \hat{Q}(u) = \hat{P}^X(u) \quad \forall u \in \mathbb{R}^n$$

Dunque abbiamo  $Q = P^X$  e  $P^X = P^{X_1} \otimes \dots \otimes P^{X_n}$ , da cui segue che le  $X_k$  sono indipendenti.  $\square$

**Corollario 11.14** (*J-P Th 15.2*)

Siano  $X_1, \dots, X_n$  VA indipendenti. Allora:

$$Y = \sum_{k=1}^n X_k \implies \varphi_Y(u) = \varphi_{\sum_k X_k}(u) = \prod_{k=1}^n \varphi_{X_k}(u) \quad \forall u \in \mathbb{R}$$

**Definizione 11.15**

Siano  $X_1, X_2$  variabili aleatorie indipendenti e  $Y = X_1 + X_2$ .

$P^Y$  è detto anche **prodotto di convoluzione** di  $P^{X_1}$  e  $P^{X_2}$ :

$$P^Y := P^{X_1} * P^{X_2}$$

È utile notare che il prodotto di convoluzione è commutativo, infatti come  $Y = X_1 + X_2 = X_2 + X_1$  la probabilità  $P^Y = P^{X_1} * P^{X_2} = P^{X_2} * P^{X_1}$ .

**Teorema 11.16** (*J-P Th 15.1, 15.3*)

Siano  $X_1, X_2$  VAR indipendenti,  $Y = X_1 + X_2$ . Allora:

1.  $P^Y(B) = (P^{X_1} * P^{X_2})(B) = \int_{\mathbb{R} \times \mathbb{R}} \mathbb{1}_B(s+t) P^{X_1}(ds) P^{X_2}(dt)$ ;
2.  $X_1$  e  $X_2$  sono continue  $\implies Y$  continua e vale il seguente risultato<sup>37</sup>:

$$f_Y(y) = \int_{\mathbb{R}} f_{X_1}(y-t) f_{X_2}(t) dt = (f_1 * f_2)(y)$$

Intuitivamente:  $\mathbb{P}(y < Y < y + dy) \simeq f_Y(y) dy$ . Ovvero, la funzione densità  $f_Y(y)$  rappresenta la probabilità che il valore della VA  $Y$  cada in un intorno infinitesimo del valore  $y$ . Questo getta luce sul significato della formula fornita dal teorema: ciascuna delle due densità indica la probabilità che la corrispondente VA assuma circa il valore  $(y - t)$  e  $t$ , rispettivamente; quindi, il loro prodotto rappresenta la probabilità che la loro somma sia  $y$ . L'integrale serve per "raccolgere" tutti i diversi valori possibili che può assumere  $y$ .

---

<sup>37</sup>"Non che questo integrale sia facile, di nuovo evviva la funzione caratteristica"



---

## 12 Variabili e vettori aleatori gaussiani

Le *variabili aleatorie gaussiane* e, più in generale, i *vettori aleatori gaussiani* sono una parte fondamentale della teoria della probabilità, sia per la loro grande versatilità modellistica (possono essere utilizzate in innumerevoli ambiti), che per le particolari proprietà di cui essi godono. In questo capitolo sarà introdotta una nuova e migliore definizione di *gaussianità*, basata sulla funzione caratteristica recentemente illustrata, e ne saranno analizzate le peculiarità, in particolare le forti relazioni biunivoche tra proprietà del vettore e proprietà delle componenti (generalmente non valide per vettori continui arbitrari). La trattazione delle variabili gaussiane permetterà inoltre di definire due utili distribuzioni ad esse correlate, la *t di Student* e la *chi-quadro*, e due nuovi indicatori statistici, la *media campionaria* e la *varianza campionaria*.

### 12.1 Definizioni

Sappiamo che una VA  $X$  è gaussiana, ovvero  $X \sim \mathcal{N}(\mu, \sigma^2)$ , con  $\mu \in \mathbb{R}, \sigma^2 > 0$  se e solo se:

1.  $X$  è una VAR continua con  $f_X(t) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(t-\mu)^2}{2\sigma^2}\right\}$ ,  $\mu \in \mathbb{R}, \sigma^2 > 0$  per definizione di gaussiana; oppure
2.  $X$  è una VAR continua con  $\varphi(u) = \exp\{iu\mu - \frac{1}{2}\sigma^2 u^2\}$ ,  $\mu \in \mathbb{R}, \sigma^2 > 0$ , come visto nel paragrafo 11.2.1.

Sappiamo anche, dal teorema 11.10, che le trasformazioni affini di VAR gaussiane rimangono gaussiane:

$$Y = aX + b \implies Y \sim \mathcal{N}(a\mu + b, a^2\sigma^2)$$

Tuttavia, nel caso  $a = 0$ , la coimplicazione (1) non funziona: infatti in questo caso  $\sigma_Y^2 = 0$  e, poiché questo valore compare al denominatore nella densità  $f_X$ , la distribuzione diventa priva di significato. La coimplicazione (2) non comporta questo problema: quando  $a = 0$  si ha semplicemente che:

$$\varphi_X(t) = \mathbb{E}[e^{iuX}] = e^{iu\mu}, \text{ ovvero } X \sim \delta_\mu \quad (X \stackrel{\text{qc}}{=} \mu)$$

In questo modo si è ottenuta una VA perfettamente sensata e ben definita.

D'ora in avanti utilizzeremo dunque questa definizione per una VAR normale:

$$\mu \in \mathbb{R}, \sigma^2 \geq 0, X \sim \mathcal{N}(\mu, \sigma^2) \iff \varphi(u) = \exp\left\{iu\mu - \frac{1}{2}\sigma^2 u^2\right\}.$$

Generalizziamo ora la definizione al caso multidimensionale.

**Definizione 12.1**

Sia  $X = (X_1, \dots, X_n) : (\Omega, \mathcal{A}, \mathbb{P}) \rightarrow (\mathbb{R}^n, \mathcal{B}^n)$  è un **vettore aleatorio gaussiano**<sup>38</sup> (o **normale**) di parametri  $\mu \in \mathbb{R}^n$ ,  $C \in \mathbb{R}^{n \times n}$  con  $C \geq 0$  se:

$$\varphi_X(u) = \exp \left\{ i \langle u | \mu \rangle - \frac{1}{2} \langle u | C u \rangle \right\} \quad \forall u \in \mathbb{R}^n$$

Scriveremo  $X \sim \mathcal{N}(\mu, C)$ .

Le componenti  $X_1, X_2, \dots, X_n$  di un vettore aleatorio gaussiano vengono dette **congiuntamente normali**.

**Osservazione 12.2**

Questa nuova definizione è valida anche per  $n \geq 2$  perché il prodotto di esponenziali è la somma degli esponenti:

$$e^{-\frac{\|u\|^2}{2}} = \varphi_{(Z_1, \dots, Z_n)}(u_1, \dots, u_n) \quad Z_k \sim \mathcal{N}(0, 1) \quad \forall k, Z_k \text{ indipendenti}$$

Inoltre, una distribuzione gaussiana  $n$ -dimensionale può essere normalizzata nel seguente modo:

$$\begin{aligned} \varphi_X(u) &= \exp \left\{ i \langle u | \mu \rangle - \frac{1}{2} \langle u | C u \rangle \right\} \\ &= \exp \left\{ i \langle u | \mu \rangle - \frac{1}{2} \sum_k (\sqrt{C} u)_k^2 \right\} \\ &= \exp \left\{ i \langle u | \mu \rangle - \frac{1}{2} |\sqrt{C} u|^2 \right\} \\ &= \exp \left\{ i \langle u | \mu \rangle - \frac{1}{2} \langle \sqrt{C} u | \sqrt{C} u \rangle \right\} \\ &= \exp \{ i \langle u | \mu \rangle \} \varphi_Z(\sqrt{C} u) \\ &= \varphi_{\sqrt{C} Z + \mu} \end{aligned}$$

Rimandiamo all'appendice A (pagina 258) per una breve trattazione sul significato di radice di matrice.

**NB:**  $X \sim \mathcal{N}(\mu, C)$  ha come supporto tutto il sottospazio  $\text{im}(C) + \mu = \text{range}(C) + \mu$ , con cui si intende lo spazio colonna di  $C$  traslato rispetto all'origine mediante il vettore  $\mu$ . Inoltre  $\text{im}(\sqrt{C}) + \mu = \text{im}(C) + \mu$ .

<sup>38</sup>Nello J-P questa definizione è trattata come un teorema (il 16.1).

## 12.2 Media e varianza

### Proposizione 12.3

Sia il vettore aleatorio  $X \sim \mathcal{N}(\mu, C)$ . Allora:

$$X_k \in L^p \quad \forall k = 1, \dots, n \quad \forall p \geq 1$$

Inoltre  $\mu = \mathbb{E}[X]$ , e  $C$  è la matrice varianza di  $X$ .

### Dimostrazione

Visto che  $(X_1, \dots, X_n) \sim \sqrt{C}(Z_1, \dots, Z_n) + \mu$  e  $Z_k \in L^p \forall k$ , essendo  $L^p$  uno spazio vettoriale, è possibile applicare la trasformazione  $\sum_j (\sqrt{C})_{kj} Z_j + \mu_k$  rimanendo all'interno di esso. Dunque  $X_k \in L^p \forall k$ .

Il valore atteso è quindi  $\mu$  perché:

$$\mathbb{E}[\sqrt{C}(Z_1, \dots, Z_n) + \mu] = \mathbb{E}[\mu] = \mu$$

Invece, la varianza è  $C$  perché:

$$\begin{aligned} C_{\sqrt{C}(Z_1, \dots, Z_n) + \mu} &= C_{\sqrt{C}(Z_1, \dots, Z_n)} = \sqrt{C} C_{(Z_1, \dots, Z_n)} \sqrt{C}^T \\ &= \sqrt{C} I \sqrt{C}^T = C \end{aligned}$$

Questo conclude la dimostrazione.  $\square$

*Esempio.* Riprendiamo l'esempio di pagina 146, applicandolo al caso vettoriale  $Z \sim \mathcal{N}(0, I)$ . Per la definizione 11.9 di funzione generatrice dei momenti abbiamo:

$$-i \frac{\partial}{\partial u_k} \varphi(0) = \mu_k, \quad -\frac{\partial^2}{\partial u_k \partial u_l} \varphi(0) = c_{kl} + \mu_k \mu_l$$

La funzione caratteristica della normale standard multidimensionale risulta essere:

$$\varphi(u) = \exp \left\{ -\frac{\langle u|u \rangle}{2} \right\} = \exp \left\{ -\frac{u_1^2 + \dots + u_n^2}{2} \right\}$$

Si può innanzitutto mostrare che rispecchia le proprietà della funzione caratteristica, ovvero continuità,  $|\varphi| \leq 1$  e  $\varphi(0) = 1$ , in quanto è un esponenziale.

Si verifica immediatamente che il valore atteso è il vettore nullo, derivando in 0 lungo qualsiasi  $u_k$  si ottiene 0:

$$-i \frac{\partial}{\partial u_k} \varphi(u) = u_k \exp \left\{ -\frac{u_1^2 + \dots + u_n^2}{2} \right\} \implies -i \frac{\partial}{\partial u_k} \varphi(0) = 0 \cdot \exp \left\{ -\frac{0}{2} \right\} = 0$$

Nel caso della varianza, derivo innanzitutto lungo  $u_k$  e  $u_l$  diversi:

$$-\frac{\partial^2}{\partial u_k \partial u_l} \varphi(u) = -u_k u_l \exp \left\{ -\frac{u_1^2 + \dots + u_n^2}{2} \right\} \implies -\frac{\partial^2}{\partial u_k \partial u_l} \varphi(0) = 0$$

Quando  $u_k = u_l$  si deve impiegare la regola del prodotto:

$$-\frac{\partial^2}{\partial u_k^2} \varphi(u) = -u_k^2 \exp \left\{ -\frac{u_1^2 + \dots + u_n^2}{2} \right\} + \exp \left\{ -\frac{u_1^2 + \dots + u_n^2}{2} \right\}$$

Sostituendo i valori numerici:

$$-\frac{\partial^2}{\partial u_k^2} \varphi(0) = -0^2 \cdot \exp \left\{ -\frac{0}{2} \right\} + \exp \left\{ -\frac{0}{2} \right\} = 0 + 1 = 1$$

ottenendo così la matrice identità.

### 12.3 Vettori gaussiani e componenti

#### Teorema 12.4

Sia  $X : (\Omega, \mathcal{A}, \mathbb{P}) \rightarrow (\mathbb{R}^n, \mathcal{B}^n)$  vettore aleatorio<sup>39</sup>. Allora:

$$X \sim \mathcal{N} \iff \langle a|X \rangle \sim \mathcal{N} \quad \forall a \in \mathbb{R}^n$$

#### Dimostrazione

- ( $\implies$ ):

Sostituendo  $\langle a|X \rangle$  nella funzione caratteristica di  $\mathcal{N}(\mu, C)$ :

$$\varphi_{\langle a|X \rangle}(u) = \mathbb{E}[e^{i\langle a|X \rangle u}] = \exp \left\{ i\langle a|\mu \rangle u - \frac{1}{2} \langle a|Ca \rangle u^2 \right\}$$

Si ottiene così una trasformazione affine:

$$\langle a|X \rangle \sim \mathcal{N}(\langle a|\mu \rangle, \langle a|Ca \rangle)$$

- ( $\impliedby$ ):

Per la proposizione 12.3:

$$\langle a|X \rangle \sim \mathcal{N} \quad \forall a \in \mathbb{R}^n \implies X_k \sim \mathcal{N} \quad \forall k, \quad X_k \in L^p \quad \forall k \quad \forall p \geq 1$$

Per ipotesi siano  $\mu = \mathbb{E}[X]$ ,  $C = C_X$ ,  $Y = \langle a|X \rangle$  e  $Y \sim \mathcal{N}$ . Ne consegue che:

$$\varphi_Y = \exp \left\{ i\mu_Y u - \frac{1}{2} \sigma_Y^2 u^2 \right\} = \exp \left\{ iu\langle a|\mu \rangle - \frac{1}{2} \langle a|Ca \rangle u^2 \right\}$$

$$\implies \varphi_X(a) = \varphi_Y(1) \implies X \sim \mathcal{N}(\mu, C) \quad \square$$

---

<sup>39</sup>Nello J-P questo teorema è trattato come la definizione di vettore gaussiano; poiché la definizione 12.1 utilizzata in questo testo coimplica il presente teorema, entrambi gli approcci sono coerenti.

**Teorema 12.5**

Se  $X = (X_1, \dots, X_n) \sim \mathcal{N}$ , allora  $X_k \sim \mathcal{N} \quad \forall k$ .

Questo risultato non è sorprendente: abbiamo infatti già visto che un vettore continuo ha componenti continue.

**NB:** Il contrario *non* è vero in generale:  $X_k \sim \mathcal{N} \quad \forall k \not\Rightarrow X \sim \mathcal{N}$ .

**Proposizione 12.6: trasformazioni affini**

Siano  $X \sim \mathcal{N}(\mu, C)$ ,  $\mu \in \mathbb{R}^n$ ,  $C = C^T \geq 0$ ,  $C \in \mathbb{R}^{n \times n}$ . Allora:

$$Y = AX + b, A \in \mathbb{R}^{m \times n}, b \in \mathbb{R}^m \implies Y \sim \mathcal{N}(A\mu + b, ACA^T)$$

ovvero trasformazioni affini dei vettori normali sono normali.

**Dimostrazione**

$$\begin{aligned} \varphi_Y(u) &= \mathbb{E} \left[ e^{i(AX+b)u} \right] = \mathbb{E} \left[ e^{iAXu} e^{ibu} \right] && \text{(per definizione)} \\ &= \mathbb{E} \left[ e^{iAXu} \right] e^{ibu} = \varphi_X(Au) e^{ibu} && \text{(linearità di } \mathbb{E} \text{)} \\ &= \exp \left\{ i \langle Au | \mu \rangle - \frac{1}{2} \langle Au | CAu \rangle + ibu \right\} && \text{(proprietà di exp)} \\ &= \exp \left\{ i(A\mu + b)u - \frac{1}{2} AC^T A^T u^2 \right\} && \text{(raccogliendo } u \text{)} \\ &= \exp \left\{ i(A\mu + b)u - \frac{1}{2} ACA^T u^2 \right\} && \text{(simmetria di } C \text{)} \end{aligned}$$

In questo modo ci si è ricondotti alla funzione caratteristica di una gaussiana:

$$Y \sim \mathcal{N}(A\mu + b, ACA^T) \quad \square$$

**Proposizione 12.7**

L'indipendenza delle componenti gaussiane è necessaria e sufficiente per la gaussianità del vettore:

$$\begin{cases} X_k \sim \mathcal{N} \quad \forall k = 1, \dots, n \\ X_k \text{ indipendenti} \end{cases} \iff \begin{cases} X = [X_1 \quad \dots \quad X_n] \sim \mathcal{N} \\ X_k \text{ indipendenti} \end{cases}$$

In tal caso, dati  $X_k \sim \mathcal{N}(\mu_k, \sigma_k^2)$ , allora:

$$X \sim \mathcal{N} \left( \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_n \end{bmatrix}, \begin{bmatrix} \sigma_1^2 & & \\ & \ddots & \\ & & \sigma_n^2 \end{bmatrix} \right)$$

**Dimostrazione**

Della dimostrazione ( $\Leftarrow$ ) siamo già a conoscenza.  
 Per ( $\Rightarrow$ ) sfruttiamo la funzione caratteristica della normale:

$$\begin{aligned} \varphi_{X_k}(u) &= \prod_{k=1}^n \varphi_{X_k}(u_k) && \text{(VA indipendenti)} \\ &= \prod_{k=1}^n \exp \left\{ i\mu_k u_k - \frac{1}{2}\sigma_k^2 u_k^2 \right\} && (X_k \sim \mathcal{N}) \\ &= \exp \left\{ i \langle \mu | u \rangle - \frac{1}{2} \left\langle u \left| \begin{bmatrix} \sigma_1^2 & & 0 \\ & \ddots & \\ 0 & & \sigma_n \end{bmatrix} u \right. \right\rangle \right\} && \square \end{aligned}$$

**NB:** Su alcuni testi si può trovare *univariato* al posto di “scalare” e *multivariato* al posto di “vettoriale”.

**Proposizione 12.8** (*J-P Th 16.3*)

Siano i vettori aleatori gaussiani  $X_1 : \Omega \rightarrow \mathbb{R}^n \sim \mathcal{N}(\mu_1, C_1)$  e  $X_2 : \Omega \rightarrow \mathbb{R}^m \sim \mathcal{N}(\mu_2, C_2)$ . Allora:

$$X_1 \perp\!\!\!\perp X_2 \iff X = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \sim \mathcal{N} \left( \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} C_1 & 0 \\ 0 & C_2 \end{bmatrix} \right)$$

con  $X : \Omega \rightarrow \mathbb{R}^{n+m}$ .

**Teorema 12.9** (*J-P Th 16.4*)

Sia il vettore aleatorio  $X = [X_1 \ \dots \ X_n] \sim \mathcal{N}$ .

$X_1, \dots, X_n$  sono indipendenti se e solo se sono scorrelate a coppie, ovvero la matrice varianza è diagonale:

$$C_X = \begin{bmatrix} \sigma_1^2 & & 0 \\ & \ddots & \\ 0 & & \sigma_n^2 \end{bmatrix}$$

Dunque, per le variabili congiuntamente normali, *scorrelazione* e *indipendenza* sono equivalenti.

**Dimostrazione**

La dimostrazione ( $\Rightarrow$ ) è ovvia perché vale a prescindere dalla natura del

vettore.

Nel caso ( $\Leftarrow$ ), data  $C_X$  diagonale:

$$\begin{aligned} \varphi_X(u) &= \exp \left\{ i \langle u | \mu \rangle - \frac{1}{2} \langle u | C_X u \rangle \right\} && \text{(per definizione)} \\ &= \exp \left\{ i \sum_{k=1}^n u_k \mu_k - \frac{1}{2} \sum_{k=1}^n \sigma_k^2 u_k^2 \right\} && \text{(def. di prodotto scalare)} \\ &= \prod_{k=1}^n \varphi_{x_k}(u_k) && \text{(proprietà di exp) } \square \end{aligned}$$

## 12.4 Normalizzazione

Nel caso unidimensionale una gaussiana è normalizzata tramite:

$$X \sim \mathcal{N}(\mu, \sigma^2), \sigma^2 > 0 \rightarrow Z = \frac{X - \mu}{\sigma} \sim \mathcal{N}(0, 1)$$

$X = \mu + \sigma Z$ , ovvero  $X$  è una trasformazione affine di  $Z$ , inoltre  $X$  e  $Z$  convivono in  $(\Omega, \mathcal{A}, \mathbb{P})$ .

Nel caso multidimensionale  $X \sim \mathcal{N}(\mu, C)$  si normalizza tramite  $X \sim \mu + \sqrt{C}Z$  con  $Z \sim \mathcal{N}(0, I_n)$ , ovvero  $X$  e  $Z$  rispettano la stessa legge. Purtroppo con quest'espressione non si ha l'uguaglianza tra i due membri, che può ora essere definita prestando attenzione ai casi degeneri.

### Teorema 12.10 (*J-P Th 16.2*)

Sia un vettore aleatorio gaussiano  $X : (\Omega, \mathcal{A}, \mathbb{P}) \rightarrow (\mathbb{R}^n, \mathbb{B}^n)$ ,  $X \sim \mathcal{N}(\mu, C)$ ,  $\mu \in \mathbb{R}^n$ ,  $C \in \mathbb{R}^{n \times n}$ ,  $C \geq 0$ . Allora:

$$\begin{cases} \exists Y : \Omega \rightarrow \mathbb{R}^n, Y \sim \mathcal{N}(0, \text{diag}(\lambda_1, \dots, \lambda_n)) & \lambda_k \geq 0 \quad \forall k \\ \exists U \in \mathbb{R}^{n \times n} \text{ ortogonale tale che } X = \mu + UY \end{cases}$$

Il teorema, enunciato in questa forma, vale a prescindere da  $Y$ .

### Dimostrazione

Per ipotesi,  $X \sim \mathcal{N}(\mu, C)$ ,  $C \geq 0$ . Poiché  $C$  è simmetrica e definita positiva, possiamo scomporla così:

$$C = U \Lambda U^T \text{ con } U^{-1} = U^T \text{ (ortogonale), } \Lambda = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}, \lambda_k \geq 0$$

Ponendo  $Y = U^T(X - \mu)$ , allora si ha  $Y \sim \mathcal{N}(0, \underbrace{U^T C U}_{\Lambda}) = \mathcal{N}(0, \Lambda)$  e dunque vale  $X = \mu + UY$ . □

**NB:**  $\mu + UY$  e  $\mu + \sqrt{C}Z$  sono normalizzazioni diverse: mentre la prima è esattamente identica, la seconda presenta solamente la stessa distribuzione di  $X$ , quindi ha più gradi di libertà.

Ripassiamo le condizioni equivalenti all'invertibilità per una generica matrice  $A$ :

$$A \text{ invertibile} \iff \ker(A) = \{0\} \iff \text{im}(A) = \mathbb{R}^n \iff \det(A) \neq 0$$

Considerando ora una matrice  $C \geq 0$ , oltre alle condizioni precedenti abbiamo:

$$\begin{aligned} C \text{ invertibile} &\iff \exists \Lambda^{-1} \iff \ker(\Lambda) = \{0\} \iff \text{im}(\Lambda) = \mathbb{R}^n \iff \\ &\iff \lambda_k > 0 \quad \forall k \iff \det(C) > 0 \iff C > 0 \end{aligned}$$

**Teorema 12.11**

Sia  $X \sim \mathcal{N}(\mu, C)$ ,  $\mu \in \mathbb{R}^n$ ,  $C \in \mathbb{R}^{n \times n}$ ,  $C \geq 0$ .

Allora  $X$  è continuo se e solo se  $C$  è invertibile. In tal caso si ha:

$$f_X(x) = \frac{1}{\sqrt{(2\pi)^n \det(C)}} \cdot \exp \left\{ -\frac{1}{2} \langle x - \mu \mid C^{-1}(x - \mu) \rangle \right\}$$

Nel caso particolare  $n = 1$  valgono i risultati già visti, in quanto  $C = \sigma^2 \neq 0$ . Per  $n \geq 2$ , quando  $C$  è invertibile il relativo vettore può muoversi in tutto lo spazio  $n$ -dimensionale e ha densità massima quando  $X = \mu$ .

*Esempio:* normalizzazione di una gaussiana in due dimensioni. Definiamo quattro diverse gaussiane:

1.  $Z = \mathcal{N}(0, I_2)$ ;
2.  $Y = \Lambda Z = \begin{bmatrix} \sqrt{\lambda_1} & 0 \\ 0 & \sqrt{\lambda_2} \end{bmatrix} Z = \mathcal{N} \left( 0, \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} \right)$ ;
3.  $UY$ , dove  $U \in \mathbb{R}^{n \times n}$  ortogonale;
4.  $X = UY + \mu$ .

Com'è possibile osservare dalla figura 12.1 mentre  $Y$  è una rototraslazione di  $X$ ,  $Z$  è una VA che ha la stessa distribuzione ma non è simile a  $X$ .

Nella figura sono stati disegnati dei quadrati per evidenziare le aree, ma operando una sezione della gaussiana bidimensionale, ovvero delle curve di livello, si otterranno delle ellissi, anche se le componenti sono tra di loro indipendenti.

Nel caso la matrice  $C$  non sia invertibile questo significa che nella normalizzazione, per ottenere  $Y$ , saranno eliminate le componenti che hanno  $\sigma^2 = 0$ ; si tornerà così in una situazione che ammette densità.



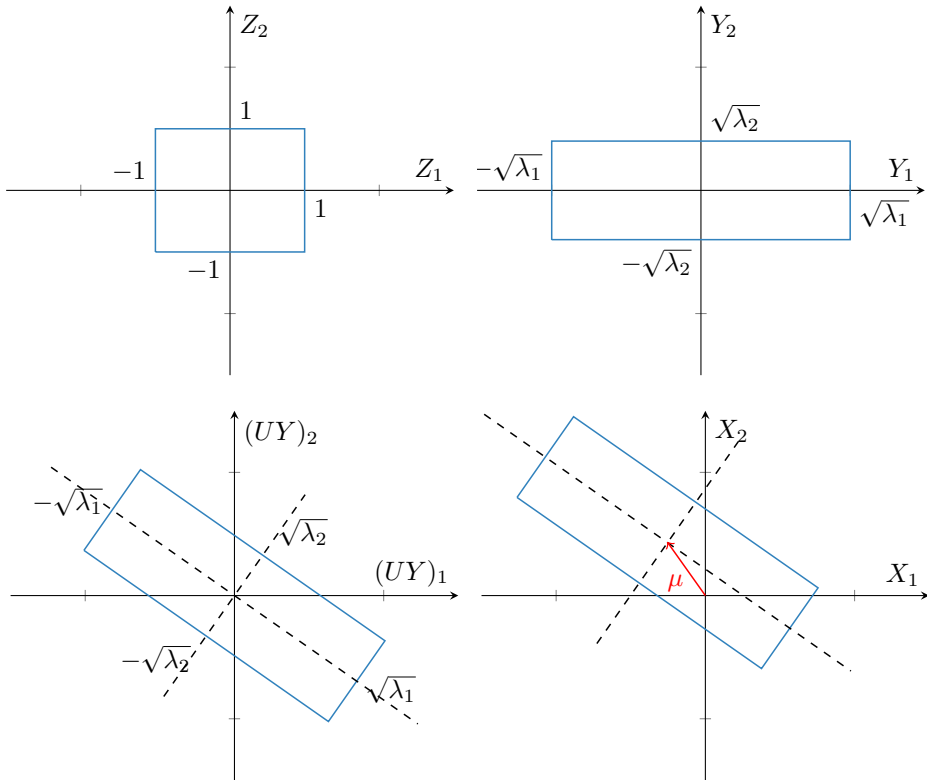


Figura 12.1: procedura per la normalizzazione nel caso bidimensionale

## 12.5 Leggi correlate alla gaussiana

### 12.5.1 La legge chi-quadro

coincide con la distribuzione della somma di variabili aleatorie indipendenti normalizzate al quadrato.

#### Definizione 12.12

La legge **chi-quadro**  $\chi^2(n)$ , con  $n \in \mathbb{N}$ , è la legge di  $Q = Z_1^2 + \dots + Z_n^2$  dove  $Z \sim \mathcal{N}(0, I_n)$ , ovvero della somma di  $n$  VA indipendenti normalizzate al quadrato.

Dalla definizione si verifica facilmente che:

$$\mathbb{E}[Q] = n, \quad \text{Var}(Q) = n\text{Var}(Z_1^2) = n(\mathbb{E}[Z_1^4] - \mathbb{E}[Z_1^2]^2) = 2n$$

**Proposizione 12.13**

$\chi^2(n) = \Gamma(\frac{n}{2}, \frac{1}{2})$ , ovvero  $\chi^2(n)$  è assolutamente continua rispetto alla misura di Lebesgue e ha densità:

$$f(x) = \frac{1}{\Gamma(\frac{n}{2}) 2^{n/2}} x^{n/2-1} e^{-x/2} \mathbf{1}_{(0,+\infty)}(x)$$

**12.5.2 La legge  $t$  di Student****Definizione 12.14**

La legge  $t$  di Student<sup>40</sup> a  $n$  gradi di libertà  $t(n)$ , con  $n \in \mathbb{N}$ , è la legge della variabile aleatoria così definita:

$$T = \frac{Z}{\sqrt{\frac{Q}{n}}} \quad \text{dove } Z \sim \mathcal{N}(0, 1), Q \sim \chi^2(n), Z \perp\!\!\!\perp Q$$

**Proposizione 12.15**

$t(n)$  è assolutamente continua rispetto alla misura di Lebesgue e ha densità:

$$f(x) = \frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2})\sqrt{\pi n}} \frac{1}{(1 + \frac{x^2}{n})^{\frac{n+1}{2}}}$$

**12.6 Media e varianza campionaria****Definizione 12.16**

Le VA  $X_1, \dots, X_n$  costituiscono un **campione casuale** se sono indipendenti e identicamente distribuite (iid).

**Definizione 12.17**

Per una successione di VA iid  $X_1, \dots, X_n$ , definiamo le seguenti due grandezze:

$$\text{media campionaria: } \bar{X} = \frac{1}{n} \sum_{k=1}^n X_k$$

$$\text{varianza campionaria: } S^2 = \frac{1}{n-1} \sum_{k=1}^n (X_k - \bar{X})^2$$

Queste saranno affrontate più nel dettaglio nel capitolo 17.5 con le loro applicazioni alla statistica.

<sup>40</sup>Dove Student è lo pseudonimo di un certo William Gosset. Questi lavorava alla fabbrica di birra "Guinness" e gli venne vietato di pubblicare articoli per evitare di profanare i segreti della fabbrica.

**Proposizione 12.18**

Sia  $X_1, \dots, X_n$  un campione casuale estratto da una popolazione  $\mathcal{N}(\mu, \sigma^2)$ .

Se  $X_1, \dots, X_n$  sono indipendenti allora:

$$X_k \sim \mathcal{N}(\mu, \sigma^2) \quad \forall k \iff \begin{bmatrix} X_1 \\ \vdots \\ X_n \end{bmatrix} \sim \mathcal{N} \left( \begin{bmatrix} \mu \\ \vdots \\ \mu \end{bmatrix}, \begin{bmatrix} \sigma^2 & & \\ & \ddots & \\ & & \sigma^2 \end{bmatrix} \right)$$

**Teorema 12.19**

Sia  $X_1, \dots, X_n$  un campione casuale estratto da una popolazione  $\mathcal{N}(\mu, \sigma^2)$ ,  $\sigma^2 > 0$ ,  $n \geq 2$ . Allora:

- $\bar{X} \sim \mathcal{N} \left( \mu, \frac{\sigma^2}{n} \right)$ .

(Grazie al teorema centrale del limite, che vedremo più avanti, questo risulterà approssimativamente vero anche se la VA non è normale.)

- $S^2 \sim \frac{\sigma^2}{n-1} \chi^2(n-1)$ , ovvero  $\frac{n-1}{\sigma^2} S^2 \sim \chi^2(n-1)$ .

Di conseguenza si può affermare che:

$$\mathbb{E}[S^2] = \sigma^2 \quad \text{e} \quad \text{Var}(S^2) = \frac{2\sigma^4}{n-1}$$

In particolare, la varianza della varianza campionaria tende a 0 al crescere di  $n$ .

- $\bar{X} \perp\!\!\!\perp S^2$ .

- $\frac{\bar{X} - \mu}{\sqrt{\frac{S^2}{n}}} \sim t(n-1)$ .

**Dimostrazione**

1. Dato  $X$  normale anche una sua trasformazione affine (lineare) è normale:

$$X = \begin{bmatrix} X_1 \\ \vdots \\ X_n \end{bmatrix} \sim \mathcal{N} \left( \begin{bmatrix} \mu \\ \vdots \\ \mu \end{bmatrix}, \begin{bmatrix} \sigma^2 & & \\ & \ddots & \\ & & \sigma^2 \end{bmatrix} \right) \implies \bar{X} = \begin{bmatrix} 1 & \dots & 1 \\ n & & n \end{bmatrix} \cdot X \sim \mathcal{N}$$

2. • Siano  $\mu = 0$ ,  $\sigma^2 = 1$ ,  $X \sim \mathcal{N}(0, I_n)$ ,  $U \in \mathbb{R}^{n \times n}$  ortogonale:

$$U = \begin{bmatrix} \frac{1}{\sqrt{n}} & \cdots & \frac{1}{\sqrt{n}} \\ & & A \end{bmatrix}$$

Solo la prima riga di  $U$  è rilevante;  $A$  denota un blocco ininfluenza.

Possiamo quindi normalizzare: definendo  $Z = UX$ , visto che  $U$  è ortogonale, si ottiene:

$$Z \sim \mathcal{N}(0, I_n) \quad \text{e} \quad Z_1 = \frac{1}{\sqrt{n}} \sum_{k=1}^n x_k = \sqrt{n}\bar{X}$$

Calcoliamo di conseguenza la somma dei quadrati:

$$\sum_{k=1}^n Z_k^2 = \|Z\|^2 - Z_1^2 \stackrel{U^T=U^{-1}}{=} \|X\|^2 - n\bar{X}^2 = (n-1)S^2$$

perché:

$$S^2 = \frac{1}{n-1} \sum_{k=1}^n (X_k - \bar{X})^2 = \frac{1}{n-1} \left( \sum_{k=1}^n X_k^2 - n\bar{X}^2 \right)$$

Concludendo, visto che  $\sigma^2 = 1$ :

$$(n-1)S^2 \sim \chi^2(n-1)$$

Inoltre vale la seguente affermazione:

$$\bar{X} = \frac{z_1}{\sqrt{n}} \perp\!\!\!\perp S^2 = h(z_2, \dots, z_n) \quad \text{con} \quad z_1 \perp\!\!\!\perp [z_2 \quad \cdots \quad z_n]$$

- Nel caso non standard, ci possiamo semplicemente ricondurre alla dimostrazione precedente tramite una normalizzazione. Siano  $\mu \in \mathbb{R}$  e  $\sigma^2 > 0$ , si considera:

$$Y_k = \frac{X_k - \mu}{\sigma} \implies \bar{Y} = \frac{\bar{X} - \mu}{\sigma}, \quad S_Y^2 = \frac{S^2}{\sigma^2}, \quad S^2 \sim \frac{\sigma^2}{n-1} \chi^2(n-1)$$

Quindi  $S^2$  risulta:

$$S^2 = \sigma^2 S_Y^2 \perp\!\!\!\perp \bar{X} = \mu + \sigma \bar{Y}$$

3. Banale corollario del punto (2).

4. Sfruttando i risultati del punto (1) e (2):

$$\frac{\bar{X} - \mu}{\sqrt{\frac{S^2}{n}}} = \frac{\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}}{\sqrt{\frac{S^2(n-1)}{\sigma^2} \cdot \frac{1}{n-1}}} \sim t(n-1)$$

Infatti il numeratore ha legge  $\mathcal{N}(0, 1)$  e  $\frac{S^2(n-1)}{\sigma^2}$  ha legge  $\chi^2(n-1)$ : per definizione di  $t$  di Student la frazione ha dunque legge  $t(n-1)$ .  $\square$

## 12.7 Considerazioni pratiche

Definiamo un vettore normale:

$$X = (X_1, \dots, X_n) \sim \mathcal{N}(\mu, C), \quad \mu \in \mathbb{R}^n, \quad C \in \mathbb{R}^{n \times n}, \quad C \geq 0$$

È noto, come visto a pagina 146, che:

$$\iff \varphi_{(X_1, \dots, X_n)}(u) = \exp \left\{ i \langle u | \mu \rangle - \frac{1}{2} \langle u | C u \rangle \right\}$$

Inoltre, secondo il teorema 12.4:

$$\iff \langle a | X \rangle \sim \mathcal{N} \quad \forall a \in \mathbb{R}^n$$

Le sue principali proprietà sono:

1.  $X_k \sim N(\mu_k, C_{kk})$
2.  $S_X = \text{im}(C) + \mu = \text{range}(C) + \mu = [\ker(C)]^\perp + \mu$
3.  $X$  vettore aleatorio continuo  $\iff \det(C) \neq 0 \iff C > 0$   
In tal caso la sua legge è:

$$f_X(x) = \frac{1}{\sqrt{(2\pi)^n \det(C)}} \exp \left\{ -\frac{1}{2} \langle x - \mu | C^{-1}(x - \mu) \rangle \right\}$$

4.  $A \in \mathbb{R}^{m \times n}, b \in \mathbb{R}^m \implies AX + b \sim N(A\mu + b, ACA^T)$  (proposizione 12.6)
5.  $X_i \perp\!\!\!\perp X_j \iff \text{Cov}(X_i, X_j) = 0$

**Caso bidimensionale** Sia  $(X, Y)$  un vettore gaussiano:

$$\begin{bmatrix} X \\ Y \end{bmatrix} \sim N \left( \begin{bmatrix} \mu_X \\ \mu_Y \end{bmatrix}, \begin{bmatrix} \sigma_X^2 & \text{Cov}(X, Y) \\ \text{Cov}(X, Y) & \sigma_Y^2 \end{bmatrix} \right)$$

La sua funzione caratteristica è:

$$\varphi_{(X,Y)}(u, v) = \exp \left\{ i(\mu_X u + \mu_Y v) - \frac{1}{2}(u^2 \sigma_X^2 + 2Cov(X, Y)uv + v^2 \sigma_Y^2) \right\}$$

In questo caso la proprietà (3) può anche essere espressa come:

$$\begin{bmatrix} X \\ Y \end{bmatrix} \iff \det(C) \neq 0 \iff \begin{cases} \sigma_X > 0 \\ \sigma_Y > 0 \\ |\rho_{X,Y}| < 1 \end{cases}$$

La legge in questo caso diventa<sup>41</sup>:

$$f_{(X,Y)}(x, y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho_{(X,Y)}^2}} \cdot \exp \left\{ -\frac{1}{2(1-\rho_{(X,Y)})} \left[ \frac{(x-\mu_X)^2}{\sigma_X^2} - 2\rho_{(X,Y)} \frac{(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y} + \frac{(y-\mu_Y)^2}{\sigma_Y^2} \right] \right\}$$

---

<sup>41</sup>un mostro

---

## 13 Leggi condizionate

In questo capitolo riprenderemo il concetto di *probabilità condizionata* giustificandone esistenza e unicità, che prima erano date per scontate. Saranno riformulate le definizioni già viste, passando dal linguaggio degli eventi a quello delle variabili aleatorie ed estendendole a una trattazione più astratta e generale. Analizzando alcuni casi particolari, ritroveremo quanto precedentemente osservato su densità e probabilità condizionate.

### Definizione 13.1

Siano  $(\Omega, \mathcal{A}, \mathbb{P})$  spazio di probabilità,  $(E, \mathcal{E}), (F, \mathcal{F})$  spazi misurabili,  $X : \Omega \rightarrow E, Y : \Omega \rightarrow F$  VA.

La funzione  $Q : E \times \mathcal{F} \rightarrow [0, 1]$  è una **legge condizionata** (o distribuzione condizionata) se:

1.  $\forall B \in \mathcal{F}$  fissato, la funzione  $Q(\cdot, B) : E \rightarrow [0, 1]$  è  $\mathcal{E}$ -misurabile;
2.  $\forall x \in E$  fissato, la funzione  $Q(x, \cdot) : \mathcal{F} \rightarrow [0, 1]$  è una probabilità;
3.  $\forall A \in \mathcal{E}, \forall B \in \mathcal{F}, \mathbb{P}(X \in A, Y \in B) = \int_E \mathbb{1}_A(x) Q(x, B) P^X(dx)$ .

Notiamo che il punto (3) è una generalizzazione al caso continuo della formula delle probabilità totali. Perché l'integrale abbia senso, è necessario che sia rispettata l'ipotesi di misurabilità di cui al punto (1).

Il concetto di *legge condizionata*, pur diverso nella definizione, è analogo a quello di probabilità condizionata che avevamo già visto a suo tempo.

Intuitivamente scriveremo:  $P^{Y|X}(B|x) = \mathbb{P}(Y \in B|X = x) =: Q(x|B)$

Vale inoltre la seguente proprietà, simile alla definizione di probabilità condizionata vista a pagina 23 del capitolo 2:

$$P^{Y|X}(y|x) = \mathbb{P}(Y = y|X = x) = \frac{\mathbb{P}(Y = y, X = x)}{\mathbb{P}(X = x)} = \frac{P^{(X,Y)}(x, y)}{P^X(x)}$$

Si può dimostrare che l'*esistenza* di  $Q$  è garantita nel caso particolare di  $(F, \mathcal{F}) = (\mathbb{R}^n, \mathcal{B}^n)$ . Non ci addentriamo nelle ipotesi aggiuntive richieste su uno spazio misurabile  $(F, \mathcal{F})$  generico. Vale inoltre l'*unicità* di  $Q$  a meno di insiemi a misura nulla; ovvero, se  $Q$  e  $\tilde{Q}$  sono due leggi condizionate di  $Y$  data  $X$ , allora è vero che,  $\forall B \in \mathcal{F}, Q(x, B) = \tilde{Q}(x, B)$  quasi certamente rispetto a  $P^X$  (scritto anche come  $P^X$ -qc).

### 13.1 Valori attesi rispetto a una legge condizionata

Sia  $(F, \mathcal{F}) = (\mathbb{R}^n, \mathcal{B}^n)$ . Introduciamo alcune grandezze delle leggi condizionate, le cui definizioni sono omologhe a quelle viste per le normali distribuzioni.

**Definizione 13.2**

Il **valore atteso condizionato** di  $Y_k$  data  $X$  è definito come:

$$\mathbb{E}[Y_k|X = x] = m(x) := \int_{\mathbb{R}^n} y Q(x, dy)$$

La **varianza condizionata** di  $Y_k$  data  $X$  è definita come:

$$\text{Var}(Y_k|X = x) = q^2(x) := \int_{\mathbb{R}^n} (y - \mathbb{E}[Y_k|X = x])^2 Q(x, dy)$$

La **covarianza condizionata** di  $Y_k$  e  $Y_l$  data  $X$  è definita come:

$$\text{Cov}(Y_k, Y_l|X = x) = \int_{\mathbb{R}^n} (y_k - \mathbb{E}[Y_k|X = x]) (y_l - \mathbb{E}[Y_l|X = x]) Q(x, dy)$$

Nel caso discreto le definizioni sono del tutto simili:

- $\mathbb{E}[Y|X = x] = \sum_{y \in S_Y} y p_{Y|X}(y|x)$
- $\text{Var}[Y|X = x] = \sum_{y \in S_Y} (y - \mathbb{E}[Y|X = x])^2 p_{Y|X}(y|x)$
- $\mathbb{E}[h(Y)|X = x] = \sum_{y \in S_Y} h(y) p_{Y|X}(y|x)$

**13.1.1 Indipendenza****Proposizione 13.3**

$X$  e  $Y$  sono indipendenti se e solo se  $P^{Y|X}(B|x) = P^Y(B)$  qc rispetto a  $P^X$ ,  $\forall B \in \mathcal{F}$ .

La condizione è simile, *mutatis mutandis*<sup>42</sup>, a una delle proprietà della probabilità condizionata: una VA è indipendente da un'altra se e solo se la sua legge rimane identica dopo il condizionamento. L'unica differenza è il linguaggio utilizzato, prima quello di indipendenza di eventi e di probabilità, ora quello di indipendenza di VA e di legge.

*Esempio.* Dato  $Y = h(X)$  con  $h : E \rightarrow F$  misurabile, allora  $P^{Y|X}(B|x) = \mathbb{1}_B(h(x)) = \delta_{h(x)}(B)$ .

*Esempio:* paradosso di Borel. Siano  $X, Y$  VA indipendenti di legge  $\mathcal{E}(\lambda)$ , con  $\lambda > 0$ . Sia inoltre  $Z = \frac{X-1}{Y}$ .

In tal caso,  $Y|X = 1 \not\sim Y|Z = 0$ , ovvero, in termini di leggi condizionate,  $P^{Y|X}(\cdot|1) \not\sim P^{Y|Z}(\cdot|0)$ .

<sup>42</sup>“Con le opportune modifiche”, dice Wikipedia. Verri non vi ha proprio insegnato niente?



Questo problema sorge dal fatto che l'evento  $A = (X = 1) = (Z = 0)$  abbia probabilità nulla, e che dunque le seguenti espressioni non sono ben definite:

$$P^{Y|X}(y|1) = \frac{P^{(X,Y)}(1, y)}{P^X(1)}, \quad P^{Y|Z}(y|0) = \frac{P^{(Z,Y)}(0, y)}{P^Z(0)}$$

Pertanto, quando d'ora in avanti ci si riferirà ad eventi della forma  $(Z = 0)$ , li si penserà sempre come ottenuti dal limite per  $\varepsilon \rightarrow 0$  degli eventi della forma  $(-\varepsilon < Z < \varepsilon)$ , che sono eventi non improbabili e quindi non generano ambiguità.

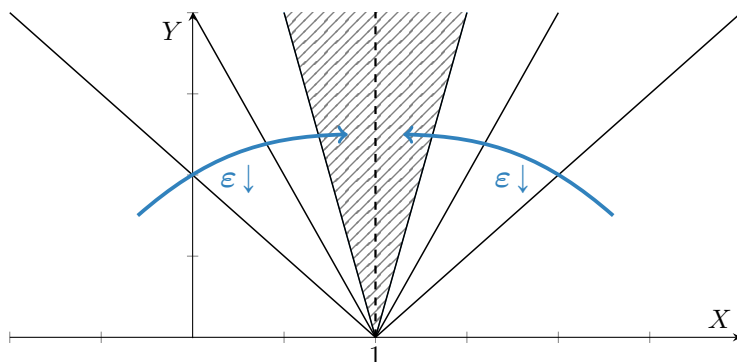


Figura 13.1: la retta  $(Z = 0)$  ottenuta come limite dei coni  $(-\varepsilon < Z < \varepsilon)$

**NB:** Come già notato, vale  $P^{(X,Y)}(x, y) = P^{Y|X}(y|x) \cdot P^X(X)$ .

Dunque conoscere  $P^{(X,Y)}$  (*probabilità congiunta*) equivale a conoscere  $P^X$  (*probabilità marginale*) e  $P^{Y|X}(\cdot|x)$  (*probabilità condizionata*).

Occupiamoci allora di legare la probabilità congiunta alla probabilità condizionata in alcuni casi di particolare rilevanza.

## 13.2 Casi discreti

### 13.2.1 Con due variabili aleatorie

*Esempio.* Si lanci un dado, seguito dal lancio di un numero di monete pari al risultato del dado. Qual è la legge congiunta del numero di teste e del risultato del dado?

Definiamo le VA discrete  $D$ , risultato del dado, e  $T$ , numero di teste. Ovviamente  $D \sim U(\{1, \dots, 6\})$ . Scriviamo  $T$  come legge condizionata: come già noto, si ha  $T|D = n \sim B\left(n, \frac{1}{2}\right)$ . I supporti delle due VA sono rispettivamente

$S_D = \{1, \dots, 6\}$  e  $S_T = \{0, \dots, 6\}$ . Il supporto del vettore aleatorio sarà dunque trapezoidale.

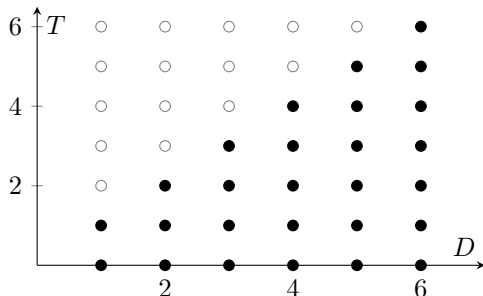


Figura 13.2: legge di  $T|D$ . I pallini pieni sono le combinazioni possibili.

Calcoliamo la probabilità richiesta mediante la definizione di probabilità condizionata (2.1):

$$P^{(D,T)}(n, k) = P^{T|D}(k|n) \cdot P^D(n) = \binom{n}{k} \frac{1}{2^n} \frac{1}{6}$$

Da questo possiamo ottenere che:

$$P^T(k) = \sum_{n=1}^6 \frac{1}{6} \binom{n}{k} \frac{1}{2^n} \mathbb{1}_{[0,n]}(k) = \frac{1}{6} \sum_{n=\max(1,k)}^6 \binom{n}{k} \frac{1}{2^n}$$

Tramite la legge condizionata, molto più semplice, siamo riusciti a risalire alla legge di  $T$ .

### 13.2.2 Con due vettori aleatori

Siano  $X$  vettore aleatorio discreto in  $\mathbb{R}^n$  e  $Y$  vettore aleatorio discreto in  $\mathbb{R}^m$ . Definiamo  $Q$  come:

$$Q(x, B) := \begin{cases} \mathbb{P}(Y \in B | X = x) & \text{per } x \in S_X \\ \tilde{Q}(B) & \text{per } x \notin S_X \end{cases}$$

$\tilde{Q}(B)$  è un'arbitraria probabilità su  $\mathcal{B}^m$ .

Verifichiamo che  $Q$  rispetti la definizione:

1.  $Q(\cdot, B)$  è misurabile in  $x$ : è costante su  $S_X^C$  e al più numerabile su  $S_X$ .
2.  $Q(x, \cdot)$  è una probabilità: per la teoria se  $x \in S_X$ , per costruzione se  $x \notin S_X$ .

3.

$$\begin{aligned}
\int_E \mathbb{1}_A(x) Q(x, B) P^X(dx) &= \sum_{x \in S_X} \mathbb{1}_A(x) Q(x, B) P^X(x) \\
&= \sum_{x \in S_X} \mathbb{1}_A(x) \mathbb{P}(Y \in B | X = x) P^X(x) \\
&= \mathbb{P}(X \in A, Y \in B)
\end{aligned}$$

Notiamo che  $\tilde{Q}$  non ha ruolo nella sommatoria, in quanto è una probabilità assunta solo su eventi fuori dal supporto, i quali per definizione hanno probabilità nulla.

### 13.2.3 Caso rettangolare

Sia  $C \in \mathcal{A} \times \mathcal{B}$ , con  $A, B \in \mathcal{B}$ . Allora:

$$\begin{aligned}
\mathbb{P}(X \in A, Y \in B) &= P^{(X,Y)}(A, B) \\
&= \sum_{x \in A \cap S_X} \sum_{y \in B \cap S_Y} P^{(X,Y)}(x, y) \\
&= \sum_{x \in A \cap S_X} \left[ \sum_{y \in B \cap S_Y} P^{(Y|X)}(y|x) \right] P^X(x) \\
&= \sum_{x \in S_X} \mathbb{1}_A(x) \mathbb{P}(Y \in B | X = x) \mathbb{P}(X = x) \\
&= \sum_{x \in S_X} \mathbb{1}_A(x) P^{Y|X}(B|x) P^X(x)
\end{aligned}$$

### 13.2.4 Caso generico

Sia  $C \in \mathcal{B}^2$  un generico boreliano. Allora:

$$\begin{aligned}
\mathbb{P}((X, Y) \in C) &= P^{(X,Y)}(C) \\
&= \sum_{(x,y) \in C} P^{(X,Y)}(x, y) && \text{(con } x \in S_X, y \in S_Y) \\
&= \sum_{x \in S_X} \left[ \sum_{y: (x,y) \in C} P^{Y|X}(y|x) \right] P^X(x) \\
&= \sum_{x \in S_X} \mathbb{P}((X, Y) \in C | X = x) \\
&= \sum_{x \in S_X} \mathbb{P}((x, Y) \in C) && \text{(dipende solo da } Y)
\end{aligned}$$

## 14 Caso congiuntamente continuo

### 14.1 Caso generico

Sia  $(X, Y)$  vettore aleatorio continuo in  $\mathbb{R}^n \times \mathbb{R}^m$ . Denotiamo con  $f_X$  e  $f_Y$  le due densità marginali e con  $f_{(X,Y)}$  la densità congiunta. Il supporto di  $X$  si può scrivere come  $S_X = \{s \in \mathbb{R}^n : f_X(s) > 0\}$ .

Definiamo  $f_{Y|X}$  come:

$$f_{Y|X}(t) := \begin{cases} \frac{f_{(X,Y)}(s, t)}{f_X(s)} & s \in S_X \\ \tilde{f}(t) & s \notin S_X \end{cases}$$

Come prima,  $\tilde{f}(t)$  è un'arbitraria probabilità su  $\mathbb{R}^m$ .

Si può dimostrare che  $f_{Y|X}(t|s)$  è una densità di probabilità su  $\mathbb{R}^m$  per ogni  $s$  fissato. La dimostrazione è lasciata come esercizio al lettore.<sup>43</sup>

#### Proposizione 14.1

$$P^{Y|X}(B|s) = \int_B f_{Y|X}(t|s) dt, \quad \text{ovvero } Y|X = s \sim f_{Y|X}(\cdot|s)$$

Verifichiamo che  $Q = P^{Y|X}$  rispetti la definizione. Per le proprietà (1) e (2) si rimanda al caso discreto, poiché la dimostrazione è molto simile (per il punto (1) si sfrutta Fubini-Tonelli). La dimostrazione della (3) è la seguente:

$$\begin{aligned} \int_{\mathbb{R}^n} \mathbb{1}_A(s) Q(s, B) P^X(ds) &= \int_{\mathbb{R}^n} \mathbb{1}_A(s) \left[ \int_B f_{Y|X}(t|s) dt \right] P^X(ds) \\ &= \int_A \int_B f_{(Y|X)}(t|s) f_X(s) ds dt \\ &= \int_{A \cap S_X} \int_B \frac{f_{(X,Y)}(s, t)}{f_X(s)} f_X(s) ds dt \\ &= \int_{A \cap S_X} \int_B f_{(X,Y)}(s, t) ds dt \\ &= \mathbb{P}(X \in A, Y \in B) \end{aligned}$$

#### Proposizione 14.2

Sia  $(X, Y)$  vettore continuo in  $\mathbb{R}^n \times \mathbb{R}^m$ . Allora  $Y|X = s$  è un vettore continuo con densità:

$$f_{Y|X}(t|s) = \frac{f_{(X,Y)}(s, t)}{f_X(s)} \quad \forall s \in S_X$$

<sup>43</sup>Ho sempre sognato di scriverlo, è un sogno che si avvera.

**Proposizione 14.3**

Sia  $(X, Y)$  vettore aleatorio in  $\mathbb{R}^n \times \mathbb{R}^m$ ; inoltre,  $X$  è continua con densità  $f_X$  e  $Y|X = s$  è pure continua con densità  $f_{Y|X}(\cdot|s) \forall s \in S_X$ . Allora  $(X, Y)$  è continuo con densità:

$$f_{(X,Y)}(s, t) = f_X(s)f_{(Y|X)}(t|s) \quad \forall s \in \mathbb{R}^n, \quad \forall t \in \mathbb{R}^m$$

**Dimostrazione**

Prendiamo due eventi  $A$  e  $B$  sulle rispettive  $\sigma$ -algebre e riscriviamone la probabilità<sup>44</sup>:

$$\begin{aligned} \mathbb{P}(X \in A, Y \in B) &= P^{(X,Y)}(A \times B) \\ &= \int_{\mathbb{R}^m} \mathbb{1}_A(X) Q_{(X,B)} P^X(dx) \\ &= \int_A \left( \int_B f_{Y|X}(t|s) dt \right) f_X(s) ds \\ &= \int_{A \times B} f_X(s) f_{Y|X}(t|s) ds dt \end{aligned} \quad \square$$

**14.1.1 Caso gaussiano per una coppia di variabili**

In due dimensioni:

$$\begin{bmatrix} X \\ Y \end{bmatrix} \sim \mathcal{N}(\mu, C) = \mathcal{N} \left( \begin{bmatrix} \mu_X \\ \mu_Y \end{bmatrix}, \begin{bmatrix} \sigma_X^2 & \rho\sigma_X\sigma_Y \\ \rho\sigma_X\sigma_Y & \sigma_Y^2 \end{bmatrix} \right)$$

Che legge ha  $Y|X = s$ ?

- Nel caso in cui  $\sigma_X = 0$ , si ha  $X \sim \mathcal{N}(\mu_X, 0) = \delta_{\mu_X} \implies \mathbb{P}(X = \mu_X) = 1 \implies Y|X = \mu_X \sim Y \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$
- Nel caso in cui  $\sigma_X > 0$ , si ha  $Y|X = s \sim \mathcal{N}(m(s), q^2)$ , dove  $m(s)$  indica la media (forma compatta per  $\mathbb{E}[Y|X = s]$ ). In particolare:

$$m(s) = \mu_Y + \frac{Cov(X, Y)}{Var(X)}(s - \mu_X) = \mu_Y + \rho \frac{\sigma_Y}{\sigma_X}(s - \mu_X)$$

$$q^2 = Var(Y|X = s) = Var(Y) - \frac{Cov(X, Y)^2}{Var(X)} = \sigma_Y^2(1 - \rho^2)$$

**Osservazione 14.4**

1.  $y = m(x)$  è la retta di regressione lineare (oggetto sviscerato in statistica).
2.  $q^2$  non dipende da  $s$ .

---

<sup>44</sup>Cosa ti ha fatto pensare che questa dimostrazione sarà diversa dalle 10 precedenti?

3. Casi notevoli:

(a)  $\rho = 0 \iff X \perp\!\!\!\perp Y \implies Y|X = s \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$ ;

(b)  $|\rho| = 1 \iff Y = AX + b \implies Y|X = s \sim \mathcal{N}(\mu_Y + \rho \frac{\sigma_Y}{\sigma_X}(s - \mu_X), 0)$ .

Questo è il caso di totale dipendenza, opposto al punto (1);

(c)  $\sigma_Y^2 = 0 \iff Y \sim \mathcal{N}(\mu_Y, 0) \implies Y|X = s \sim \mathcal{N}(\mu_Y, 0)$ ;

(d)  $q^2 \leq \sigma_Y^2$  e solo nel caso  $X \perp\!\!\!\perp Y$  vale l'uguaglianza:  $q^2 = \sigma_Y^2$ ;

(e)  $Y|X = s \sim \mathcal{N} \forall s$ .

### 14.1.2 Caso gaussiano per una coppia di vettori

#### Proposizione 14.5

Siano  $X \sim \mathcal{N}(\mu_X, C_X)$   $n$ -dimensionale e  $Y \sim \mathcal{N}(\mu_Y, C_Y)$   $m$ -dimensionale.

Allora:

$$\begin{bmatrix} X \\ Y \end{bmatrix} \sim \mathcal{N} \left( \begin{bmatrix} \mu_X \\ \mu_Y \end{bmatrix}, \begin{bmatrix} C_X & C_{XY} \\ C_{YX} & C_Y \end{bmatrix} \right)$$

Inoltre vale ovviamente  $C_{XY} = C_{YX}^T$ .

#### Proposizione 14.6

Se  $C_X$  è invertibile, ovvero  $\det(C_X) > 0$ , allora:

- $Y|X = s \sim \mathcal{N}(m(s), Q)$ ;
- $m(s) = \mathbb{E}[Y|X = s] = \mu_Y + C_{YX}C_X^{-1}(s - \mu_X)$ ;
- $Q = \text{Var}(Y|X = s) = C_Y - C_{YX}C_X^{-1}C_{XY}$ .

### 14.2 Caso misto

Sia  $(X, Y)$  vettore aleatorio in  $\mathbb{R}^{n+m}$ .

Siano inoltre  $X$  VA discreta di densità  $P_X$  e  $Y|X = s$  continua di densità

$$f_{Y|X}(\cdot|s) \forall s \in S_X.$$

Avendo una componente continua e una discreta, fissiamo una delle due e calco-

liamo l'altra:

$$\begin{aligned}
 \mathbb{P}(X \in A, Y \in B) &= \mathbb{P}^{(X,Y)}(A \times B) \\
 &= \int_{\mathbb{R}^n} \mathbf{1}_A(s) Q(s, B) P^X(ds) \\
 &= \sum_{s \in S_X} \mathbf{1}_A(s) \left( \int_B f_{Y|X}(t|s) dt P^X(s) \right) \\
 &= \sum_{s \in A \cap S_X} \left( \int_B f_{Y|X}(t|s) P^X(s) dt \right)
 \end{aligned}$$

Che legge ha  $Y$ ? Per scoprirlo si ponga  $A = \mathbb{R}^n$ , da cui discende che  $(X \in \mathbb{R}^n) = \Omega$ . Applichiamo ora la formula di sopra:

$$\begin{aligned}
 P(Y \in B) &= \sum_{s \in S_X} \left( \int_B f_{Y|X}(t|s) P^X(s) dt \right) \\
 &= \int_B \left( \sum_{s \in S_X} f_{Y|X}(t|s) P^X(s) \right) dt \quad (\text{per Fubini-Tonelli}) \\
 &= \int_B f_Y(t) dt
 \end{aligned}$$

Quindi  $Y$  è un vettore continuo con densità  $f_Y$ .





---

## 15 Attesa condizionata

Nel capitolo precedente è stata affrontata la trattazione del valore atteso condizionato dalla conoscenza di un determinato valore della variabile condizionante. Ora si vuole ampliare e generalizzare tale concetto, trasformando quel valore atteso in una nuova variabile aleatoria, detta *attesa condizionata*, dipendente appunto dal valore (questa volta *non noto a priori*) della variabile condizionante. Per sviluppare la teoria in questa direzione sarà necessario introdurre la  $\sigma$ -algebra generata da una variabile aleatoria, che non è altro che la rappresentazione formale dell'informazione fornita dalla variabile aleatoria stessa. L'attesa condizionata gode di una serie di utilissime proprietà e applicazioni in vari ambiti, dal calcolo di valori attesi e varianze complesse all'approssimazione di variabili aleatorie a partire da informazioni limitate.

### 15.1 Condizionamento data una variabile aleatoria

Ricapitoliamo quanto detto nelle pagine precedenti. Dato  $(X, Y)$  vettore aleatorio in  $E \times \mathbb{R}^n$ , con  $Y$  limitata:

- **a priori** è noto che  $Y$  può assumere valori “concentrati” su  $\mathbb{E}[Y]$  (in quanto valore atteso);
- **in medias res** se si scopre che  $X = s$  (ovvero, l'esperimento aleatorio sulla VA  $X$  produce il valore  $s$ ), è noto che  $Y$  può assumere valori “concentrati” su  $m(s) = \mathbb{E}[Y|X = s]$ .

Guardiamo ora gli eventi da una diversa prospettiva:

- **a priori** quando scoprirò  $X$ , saprò che  $Y$  può assumere valori “concentrati” su  $m(X) = \mathbb{E}[Y|X]$ , che è a sua volta una VA e sarà chiamata **attesa condizionata**.

Il fatto che le previsioni riassunte da  $m(X)$  siano effettuate *a priori* (prima della conoscenza dell'evento  $X = s$ ), e su tutti i possibili valori di  $X$ , è proprio quello che la rende una variabile aleatoria a tutti gli effetti, a differenza di  $\mathbb{E}[Y|X = s]$ , che invece è un normale valore atteso e dunque una quantità numerica ben definita.

**NB:**  $Q(x, B) \rightarrow m(x) \rightarrow m(X) = \mathbb{E}[Y|X]$ :

Dalla teoria precedente sappiamo già calcolare i valori attesi condizionati  $\mathbb{E}[Y|X = s] \forall s \in S_X$ , e di conseguenza *sappiamo già calcolare anche  $m(X)$* : basta sostituire ogni istanza di  $s$  con una  $X$ . L'obiettivo di questo capitolo è però capire il significato e le proprietà di questo oggetto e svincolarlo da  $Q$ ; infatti  $m(s)$  non è univocamente determinato a causa dell'arbitrarietà nella definizione di  $Q(s, B)$ , e ciò potrebbe causare problemi nello sviluppo della teoria.

*Esempio.* Riprendiamo l'esperimento del dado e della moneta di pagina 169, ma con le leggi condizionate.

Lanciato un dado ( $D \sim U(\{1, \dots, 6\})$ ), lanciamo una moneta  $n$  volte in base al numero sorteggiato:

$$T|D = n \sim Bi\left(n, \frac{1}{2}\right) \quad \forall n = 1, \dots, 6$$

$$\mathbb{E}[T|D = n] = \frac{n}{2} = m(n) \implies \mathbb{E}[T|D] = m(D) = \frac{D}{2}$$

Come è evidente in questo caso la trattazione mediante il valore atteso condizionato è molto più semplice.

**NB:** Se la VA condizionante è discreta, l'attesa condizionata ha una semplice e ovvia formulazione:

$$\mathbb{E}[Y|X] = \sum_{s \in S_X} \mathbb{E}[Y|X = s] \mathbf{1}_{(X=s)}$$

Ovvero,  $\mathbb{E}[Y|X]$  assume il valore  $\mathbb{E}[Y|X = s]$  solo nel caso in cui risulti che  $X = s$ , e questo per tutti gli  $s$  del supporto.

Scoprire che  $X = X(\omega) = x$  equivale a scoprire che  $\omega \in (X = x)$ , cioè a scoprire che si verificano tutti gli eventi nella controimmagine. Più in generale, ricordiamo che  $\omega \in (x \in B)$ , dove  $B$  è un evento contenuto nella  $\sigma$ -algebra  $\mathcal{E}$ , e  $X(\omega) = (x \in B)$  sono sinonimi per definizione di controimmagine.

### 15.1.1 $\sigma$ -algebra generata da una variabile

#### Definizione 15.1

Siano  $\Omega$  spazio campionario,  $(E, \mathcal{E})$  spazio misurabile e  $X : \Omega \rightarrow E$  VA.

Si dice  **$\sigma$ -algebra generata da  $X$**  la collezione così definita:

$$X^{-1}(\mathcal{E}) = \sigma(X) := \{A \subseteq \Omega : A = (X \in B) = X^{-1}(B) \text{ per qualche } B \in \mathcal{E}\}$$

$\sigma(X)$  è, insomma, la collezione degli eventi (del dominio) la cui immagine è un evento (del codominio).

Scoprire il valore di  $X$  equivale allo scoprire se ciascuno degli eventi in  $\sigma(X)$  si è verificato o meno: in altre parole,  $\sigma(X)$  rappresenta l'*informazione rivelata* da  $X$ .

#### Proposizione 15.2: proprietà di $\sigma(X)$

1.  $\sigma(X)$  è una  $\sigma$ -algebra;

2.  $\sigma(X)$  è la più piccola  $\sigma$ -algebra su cui  $X$  è misurabile;
3.  $X$  è  $\mathcal{A}$ -misurabile  $\iff \sigma(X) \subseteq \mathcal{A}$ .

*Esempio.* Siano  $\Omega = \{0, 1\}^{\mathbb{N}}$ ,  $X_k : \Omega \rightarrow \mathbb{R}$  successione di VA,  $X_k(\omega) = \omega_k \forall k \in \mathbb{N}$ . Consideriamo l'evento  $(X_3 = 1) = \{\omega \in \Omega : \omega_3 = 1\} =: E_3$ . Si ha che  $\sigma(X_3) = \{A \subseteq \Omega, A = (X \in B) \text{ tale che } B \in \mathcal{B}\} = \{E_3, E_3^C, \Omega, \emptyset\}$ .

Consideriamo poi le VA  $X_1, X_2$  e  $X_3$ : la  $\sigma$ -algebra generata dal vettore che le riunisce è  $\sigma(X_1, X_2, X_3) = \{A \subseteq \Omega, A = ((X_1, X_2, X_3) \in B) \text{ tale che } B \in \mathcal{B}^3\} = \sigma(E_1, E_2, E_3)$ . Chiaramente nella costruzione di quest'ultima trascuriamo le combinazioni di eventi impossibili, come  $E_1$  e  $E_1^C$  contemporaneamente.

**Lemma 15.3: misurabilità di Doob** (*J-P Th 23.2*)

Siano  $\Omega, (E, \mathcal{E})$  spazio misurabile, e le VA  $X : \Omega \rightarrow E, Y : \Omega \rightarrow \mathbb{R}$ . Allora:

$$Y \text{ è } \sigma(X)\text{-misurabile} \iff \exists h : E \rightarrow \mathbb{R} \text{ misurabile tale che } Y = h(X)$$

Abbozziamo la dimostrazione. Intuitivamente si nota che  $Y$   $\sigma(X)$ -misurabile  $\iff \sigma(Y) \subseteq \sigma(X)$ : acquisire informazioni su  $\sigma(X)$  significa inevitabilmente acquisire informazioni su  $\sigma(Y)$ . In particolare, scoprire il valore di  $X$  significa scoprire il valore di  $Y$ ; in altre parole quindi  $Y = h(X)$ .

**15.1.2 Calcolo di valori attesi**

**Proposizione 15.4**

Siano  $(\Omega, \mathcal{A}, \mathbb{P})$  spazio di probabilità,  $(E, \mathcal{E})$  spazio misurabile,  $X : \Omega \rightarrow E$  e  $Y : \Omega \rightarrow \mathbb{R}$  variabili aleatorie con  $Y$  limitata o positiva. Allora  $Z = \mathbb{E}[Y|X] = m(X)$  è l'unica VAR  $\sigma(X)$ -misurabile, a meno di eventi trascurabili<sup>45</sup>, tale che:

$$\int_A Y \, d\mathbb{P} = \int_A Z \, d\mathbb{P} \quad \forall A \in \sigma(X) \quad \text{ovvero} \quad \mathbb{E}[\mathbb{1}_A Y] = \mathbb{E}[\mathbb{1}_A Z] \quad \forall A \in \sigma(X)$$

Si noti che la richiesta di limitatezza (che implica l'appartenenza a  $L^1$ ) o di positività di  $Y$  è necessaria per garantire l'esistenza del valore atteso a prescindere dal valore  $s$  che sarà assunto da  $X$ . Inoltre, la  $\sigma(X)$ -misurabilità di  $Z = m(X)$  è ovvia in quanto essa è composizione di una VA  $\sigma(X)$ -misurabile per definizione.

Scegliendo  $A = \Omega \in \sigma(X)$ , si ottiene il seguente importantissimo caso particolare:

$$\mathbb{E}[Y] \stackrel{!}{=} \mathbb{E}[\mathbb{E}[Y|X]] = \begin{cases} \sum_{s \in S_X} \mathbb{E}[Y|X = s] p_X(s) & \text{se } X \text{ è discreto in } \mathbb{R}^n \\ \int_{\mathbb{R}^n} \mathbb{E}[Y|X = s] f_X(s) \, ds & \text{se } (X, Y) \text{ è continuo in } \mathbb{R}^n \end{cases}$$

<sup>45</sup>Questa precisazione è inevitabile in quanto gli eventi trascurabili non influiscono sul valore atteso, rendendo impossibile individuare la loro presenza o assenza nel caso generale.

La prima uguaglianza esprime in maniera formale un concetto intuitivo: il valore atteso, stabilito *a priori*, di  $Y$  dato un certo valore di  $X$ , *non ancora noto a priori* e che dunque varia su tutto il dominio di  $X$  senza fornire *nessuna informazione aggiuntiva*, non può che essere il valore atteso di  $Y$  stessa. La seconda parte dell'uguaglianza fornisce nuove e utilissime formule di calcolo per valori attesi particolarmente complessi, ora ridotti a un semplice integrale di leggi condizionate da un valore noto, le quali sono facilmente ricavabili attraverso i metodi precedentemente visti. (Un'altra applicazione di minore utilità pratica per questo corso è l'utilizzo di un generico  $A \neq \Omega$  per il calcolo di leggi condizionate.)

*Esempio.* Riprendiamo una terza volta l'esempio delle pagine 169 e 178. Lancio un dado che risulta in  $n$ , che è seguito dal lancio di  $n$  monete, e conteggio le teste ottenute.

È già noto che  $D \sim U(\{1, \dots, 6\})$ ,  $T|D = n \sim Bi(n, \frac{1}{2})$  e  $\mathbb{E}[T|D] = \frac{D}{2}$ , dunque:

$$\mathbb{E}[T] = \mathbb{E}[\mathbb{E}[T|D]] = \mathbb{E}\left[\frac{D}{2}\right] = \frac{1}{2} \mathbb{E}[D] = \frac{1}{2} \cdot \frac{1+6}{2} = \frac{7}{4}$$

*Esempio.* Sia il vettore aleatorio  $(X, Y) \sim \mathcal{N}$  con  $\sigma_X, \sigma_Y > 0$ . Allora:

$$\mathbb{E}[Y] = \mathbb{E}[\mathbb{E}[Y|X]] = \mathbb{E}\left[\mu_Y + \rho \frac{\sigma_X}{\sigma_Y}(X - \mu_X)\right] = \mu_Y$$

Si può infatti dimostrare che nel caso gaussiano la formula è valida nonostante le variabili gaussiane non siano limitate, a dimostrazione del fatto che positività e limitatezza siano solo condizioni sufficienti e non necessarie per la sua validità.

**NB:**  $\mathbb{E}[Y|X]$  dipende da  $Y$  e da  $\sigma(X)$ , ma non direttamente da  $X$ . Per esempio,  $\mathbb{E}[Y|X] = \mathbb{E}[Y|X^3] = \mathbb{E}[Y|h(X)]$  con  $h$  generica funzione invertibile: infatti, poiché  $\sigma(X)$  rappresenta l'informazione fornita conoscendo  $X$ , da  $h(X)$  invertibile si può risalire a tutti e soli valori di  $X$ , da cui discende l'uguaglianza delle  $\sigma$ -algebre. Inoltre, l'uguaglianza degli integrali di  $Y$  e  $Z$  non implica necessariamente l'uguaglianza delle VA stesse. Infatti in generale  $Y$  non è  $\sigma(X)$ -misurabile: questo succede solo nei (rari) casi in cui  $Y = h(X)$ .

### 15.1.3 Formule applicative

In generale, date due VA continue:

$$\begin{aligned} Y|X = x &\sim f_{Y|X}(\cdot|x) \\ f_{(X,Y)}(x, y) &= f_{Y|X}(y|x) \cdot f_X(x) \\ \mathbb{E}[Y|X = x] &= m(x) = \int_{S_Y} y f_{Y|X}(y|x) dy \\ \text{Var}(Y|X = x) &= q^2(x) = \int_{S_Y} (y - m(x))^2 f_{Y|X}(y|x) dy \end{aligned}$$

Nel caso di VA discrete è sufficiente sostituire la densità discreta  $p$  alla densità continua  $f$  e svolgere gli integrali come sommatorie.

## 15.2 Condizionamento data una $\sigma$ -algebra

**Teorema 15.5** (*J-P Th 23.4*)

Siano  $(\Omega, \mathcal{A}, \mathbb{P})$  spazio di probabilità,  $Y : \Omega \rightarrow \mathbb{R}$  VA tale che  $Y \in L^1(\mathcal{A})$ , e  $\mathcal{G}$  una *sotto- $\sigma$ -algebra* di  $\mathcal{A}$  (ovvero una  $\sigma$ -algebra tale che  $\mathcal{G} \subseteq \mathcal{A}$ ). Allora  $\exists Z : \Omega \rightarrow \mathbb{R}$  VAR tale che:

1.  $Z \in L^1$ ;
2.  $Z$  è  $\mathcal{G}$ -misurabile;
3.  $\int_G Z d\mathbb{P} = \int_G Y d\mathbb{P} \quad \forall G \in \mathcal{G}$ ;
- 3'. (condizione equivalente<sup>46</sup> alla (3))  $\mathbb{E}[WZ] = \mathbb{E}[WY]$ ,  $\forall W$   $\mathcal{G}$ -misurabile e limitata.

Inoltre, se  $\exists \tilde{Z} : \Omega \rightarrow \mathbb{R}$  che rispetta queste 3 proprietà,  $Z \stackrel{\text{qc}}{=} \tilde{Z}$ : in altre parole  $Z$  è *unica* a meno (come al solito) di eventi trascurabili.  $Z$  viene chiamata **attesa condizionata di  $Y$  data  $\mathcal{G}$**  e denotata con  $\mathbb{E}[Y|\mathcal{G}]$ .

Modellisticamente parlando,  $\mathbb{E}[Y|\mathcal{G}]$  rappresenta la modifica di  $\mathbb{E}[Y]$  in base alle nuove informazioni ottenute *in medias res*, che sono il verificarsi o meno degli eventi di  $\mathcal{G}$ , similmente a quanto detto precedentemente per  $\mathbb{E}[Y|X]$ , l'attesa condizionata da una VA.

È bene evidenziare ulteriormente che, per quanto detto sopra, in realtà  $\mathbb{E}[Y|\mathcal{G}]$  non è una VA in senso stretto, bensì una classe di equivalenza di tutte le VA che sono ad essa uguali quasi certamente. Ma questo non causa particolari problemi in quanto leggi, valori attesi, etc. sono invarianti di una classe di equivalenza di VA.

**NB:** Scegliendo  $\mathcal{G} = \sigma(X)$ , si ottiene  $\mathbb{E}[Y|\sigma(X)] = \mathbb{E}[Y|X]$ .

*Esempio.* Sia  $X \sim (0, \sigma^2)$ . Quanto vale  $\mathbb{E}[X|X^2]$ ?

Il vettore  $(X, X^2)$  non è congiuntamente continuo, quindi non può essere calcolato mediante formule che coinvolgono la legge congiunta. Ma se si intuisce la possibile soluzione, si possono verificare le 3 proprietà richieste dal teorema 15.5; la soluzione così ricavata, come sappiamo, è unica.

La risposta è  $\mathbb{E}[X|X^2] = 0$ : si immagina notando che, conoscendo il valore al quadrato della VA,  $X$  potrebbe essere una gaussiana centrata sulla radice positiva o sulla radice negativa del quadrato: il valore atteso diventa dunque il punto

<sup>46</sup>Scegliendo  $W = \mathbf{1}_A$  si vede immediatamente che (3') implica (3); l'opposto è vero per convergenza dominata.

medio tra queste due gaussiane equiprobabili, ovvero 0. La dimostrazione delle 3 proprietà è lasciata al lettore come esercizio.

**Teorema 15.6: alcune proprietà di  $\mathbb{E}[Y|\mathcal{G}]$**  <sup>47</sup> (*J-P Th 23.3 e seguenti*)

Sia  $Y \in L^1(\mathcal{A})$  e  $\mathcal{G}$  sotto- $\sigma$ -algebra di  $\mathcal{A}$ . Allora valgono le seguenti proprietà:

1.  $\mathbb{E}[\cdot|\mathcal{G}] : L^1(\mathcal{A}) \rightarrow L^1(\mathcal{G})$  è una mappa **lineare**, ovvero  $\mathbb{E}[a_1Y_1 + a_2Y_2|\mathcal{G}] = a_1\mathbb{E}[Y_1|\mathcal{G}] + a_2\mathbb{E}[Y_2|\mathcal{G}]$ .
2.  $\mathbb{E}[\cdot|\mathcal{G}] : L^1(\mathcal{A}) \rightarrow L^1(\mathcal{G})$  è una mappa **positiva**, ovvero  $Y \geq 0$  qc  $\implies \mathbb{E}[Y|\mathcal{G}] \geq 0$  qc.  
Ovviamente, ottenere delle informazioni su  $Y$  non può cambiare il fatto che  $\mathbb{E}[Y] \geq 0$ , che discende dall'ipotesi  $Y \geq 0$ .
3.  $\mathbb{E}[\mathbb{E}[Y|\mathcal{G}]] = \mathbb{E}[Y]$ .

Qui è particolarmente importante l'ipotesi  $Y \in L^1(\mathcal{A})$ : in caso contrario, uno dei due membri dell'equazione potrebbe non esistere e la formula non reggerebbe. Questa proprietà è un'utile formula di calcolo per  $\mathbb{E}[Y]$ .

4.  $Y$  è  $\mathcal{G}$ -misurabile  $\implies \mathbb{E}[Y|\mathcal{G}] \stackrel{\text{qc}}{=} Y$ .

Ricordiamo che  $Y$   $\mathcal{G}$ -misurabile significa che tutte le controimmagini di  $Y$  appartengono alla  $\sigma$ -algebra  $\mathcal{G}$ . Pertanto, la  $\mathcal{G}$ -misurabilità fa sì che scoprendo se gli eventi di  $\mathcal{G}$  sono avvenuti o meno, allora si possa risalire all'esatto valore di  $Y$  cercando tra le immagini delle controimmagini note. Esempi di casi particolari di interesse sono  $\mathbb{E}[Y|\mathcal{A}] = Y$  e  $\mathbb{E}[Y|\emptyset] = \mathbb{E}[Y]$ ,  $\forall Y$   $\mathcal{A}$ -misurabile.

5.  $Y \perp\!\!\!\perp \mathcal{G} \implies \mathbb{E}[Y|\mathcal{G}] \stackrel{\text{qc}}{=} \mathbb{E}[Y]$ .

La definizione di *indipendenza tra una VA e una  $\sigma$ -algebra* è la seguente:

$$Y \perp\!\!\!\perp \mathcal{G} \iff \mathbb{P}(Y \in B, G) = \mathbb{P}(Y \in B) \mathbb{P}(Y \in G) \quad \forall B \in \mathcal{B}, \forall G \in \mathcal{G}$$

In altre parole, l'intersezione di una qualsiasi controimmagine di  $Y$  con un qualsiasi insieme di  $\mathcal{G}$  ha probabilità fattorizzabile. La proprietà conferma l'intuizione che scoprire informazioni slegate da  $Y$  non modifica in alcun modo la previsione su  $Y$ .

6. Se  $\mathcal{H}$  è una  $\sigma$ -algebra tale che  $\mathcal{H} \subseteq \mathcal{G} \subseteq \mathcal{A}$ , allora  $\mathbb{E}[\mathbb{E}[Y|\mathcal{G}]|\mathcal{H}] \stackrel{\text{qc}}{=} \mathbb{E}[Y|\mathcal{H}]$ .  
Il membro sinistro rappresenta una situazione in cui acquisiamo informazioni in più passi: prima  $\mathcal{H}$ , poi  $\mathcal{G}$ . La proprietà evidenzia che questa acquisizione per passi corrisponde a una semplice acquisizione di informazione in un passo solo.
7. Se  $W$  è  $\mathcal{G}$ -misurabile e tale che  $WY \in L^1(\mathcal{A})$ , allora  $\mathbb{E}[WY|\mathcal{G}] \stackrel{\text{qc}}{=} W \mathbb{E}[Y|\mathcal{G}]$ .

Ovvero, le variabili  $\mathcal{G}$ -misurabili possono essere portate fuori da un'attesa

<sup>47</sup>Meglio note come 13 Reasons Why

condizionata, esattamente come le costanti possono essere portate fuori da un integrale. Un esempio di utile applicazione di questa proprietà è il calcolo di  $\mathbb{E}[X^2Y|X]$  con  $(X, Y)$  congiuntamente gaussiane:

$$\mathbb{E}[X^2Y|X] = X^2\mathbb{E}[Y|X] = X^2\left(\mu_Y + \rho\frac{\sigma_X}{\sigma_Y}(X - \mu_X)\right)$$

Il calcolo sarebbe risultato piuttosto complicato utilizzando la legge congiunta.

8. Se  $Y_1 \stackrel{qc}{=} Y_2$ , allora  $\mathbb{E}[Y_1|\mathcal{G}] \stackrel{qc}{=} \mathbb{E}[Y_2|\mathcal{G}]$ .

Le variabili uguali quasi certamente hanno la stessa legge e modellizzano lo stesso evento: è naturale che le rispettive attese condizionate siano uguali a parità di informazioni rivelate.

9. **(convergenza monotona)** Se  $0 \leq Y_n \uparrow Y$  qc, allora  $\mathbb{E}[Y_n|\mathcal{G}] \uparrow \mathbb{E}[Y|\mathcal{G}]$  qc.

10. **(lemma di Fatou)** Se  $Y_n \geq 0$  qc, allora  $\mathbb{E}\left[\liminf_n Y_n \mid \mathcal{G}\right] \leq \liminf_n \mathbb{E}[Y_n|\mathcal{G}]$ .

11. **(convergenza dominata)** Se  $Y_n \rightarrow Y$  qc e  $|Y_n| \leq V$  qc per qualche  $V \in L^1(\mathcal{A})$ , allora  $\mathbb{E}[Y_n|\mathcal{G}] \rightarrow \mathbb{E}[Y|\mathcal{G}]$  qc.

Le 3 proprietà precedenti sono generalizzazioni delle già note proprietà di  $\mathbb{E}[Y]$ , formulate dal teorema 6.12, al caso  $\mathbb{E}[Y|\mathcal{G}]$ .

12. Se  $Y \in L^p(\mathcal{A})$ , allora  $\mathbb{E}[Y|\mathcal{G}] \in L^p(\mathcal{A})$ .

13. Se  $h : \mathbb{R} \rightarrow \mathbb{R}$  è misurabile e tale che  $h(Y) \in L^1(\mathcal{A})$ , e  $\mathcal{G} = \sigma(X)$ , allora  $\mathbb{E}[h(Y)|\sigma(X)] = h(n(X))$ , con  $n(s) = \int_{\mathbb{R}} h(t)P^{Y|X}(dt|s)$ .

Anche quest'ultima proprietà è una generalizzazione delle proprietà di  $\mathbb{E}[Y]$  a  $\mathbb{E}[Y|\mathcal{G}]$ .

Molte delle proprietà sopra elencate hanno funzione di utili formule di calcolo applicabili negli esercizi. Inoltre, il teorema mostra la grande somiglianza formale tra le funzioni valore atteso e attesa condizionata da una  $\sigma$ -algebra, che hanno molte proprietà simili o in comune.

**NB:** Le proprietà reggono anche se  $Y$  è un vettore multidimensionale. È sufficiente definire, con  $Y : \Omega \rightarrow \mathbb{R}^n$ , l'attesa condizionata come:

$$\mathbb{E}[Y|\mathcal{G}] = \mathbb{E}\left[\begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix} \mid \mathcal{G}\right] := \begin{bmatrix} \mathbb{E}[Y_1|\mathcal{G}] \\ \vdots \\ \mathbb{E}[Y_n|\mathcal{G}] \end{bmatrix}$$

### 15.3 Approssimazione di variabili aleatorie

Siano  $(\Omega, \mathcal{A}, \mathbb{P})$  uno spazio di probabilità e  $Y \in L^2(\mathcal{A})$ . Poiché sappiamo che  $L^2 \subseteq L^1$ , è vero che  $\mathbb{E}[Y|\mathcal{G}] \in L^2$ .

Cerchiamo ora una  $Z \in L^2(\mathcal{G})$  per *approssimare*  $Y$ . Qual è l'errore commesso?

#### Definizione 15.7

Date  $Y, Z \in L^2$  VA e un esito  $\omega \in \Omega$ ,  $|Y(\omega) - Z(\omega)|$  si dice *errore* (semplice). La funzione  $|Y - Z|$  è detta *errore casuale*. Il suo quadrato  $|Y - Z|^2$  è detto *errore quadratico casuale*. Infine, la quantità  $\mathbb{E}[(Y - Z)^2]$  si chiama **errore quadratico medio**; esso rappresenta la bontà, giudicata *a priori* rispetto all'esperimento aleatorio, dell'approssimazione rappresentata da  $Z$ .

#### Proposizione 15.8

Data  $Y \in L^1(\mathcal{A})$  e  $\mathcal{G}$  sotto- $\sigma$ -algebra di  $\mathcal{A}$ , l'errore quadratico medio  $\mathbb{E}[(Y - Z)^2]$ , al variare di  $Z \in L^2(\mathcal{G})$ , è *minimizzato* con  $Z = \mathbb{E}[Y|\mathcal{G}]$ . In altre parole, l'attesa condizionata di  $Y$  data  $\mathcal{G}$  può anche essere definita come la migliore approssimazione di  $Y$  tra le variabili  $\mathcal{G}$ -misurabili.

#### Dimostrazione

Si ponga  $\hat{Y} = \mathbb{E}[Y|\mathcal{G}]$  per semplicità di notazione. Sfruttando la linearità del valore atteso:

$$\begin{aligned} \mathbb{E}[(Y - Z)^2] &= \mathbb{E}[(Y - Z + \hat{Y} - \hat{Y})^2] \\ &= \mathbb{E}[(Y - \hat{Y})^2] + 2\mathbb{E}[(Y - \hat{Y})(\hat{Y} - Z)] + \mathbb{E}[(\hat{Y} - Z)^2] \end{aligned}$$

Ora, applicando in sequenza alcune proprietà dal teorema 15.6 al doppio prodotto, esso diventa:

$$\begin{aligned} \mathbb{E}[(Y - \hat{Y})(\hat{Y} - Z)] &\stackrel{(3)}{=} \mathbb{E}[\mathbb{E}[(Y - \hat{Y})(\hat{Y} - Z)|\mathcal{G}]] \\ &\stackrel{(7)}{=} \mathbb{E}[(\hat{Y} - Z)\mathbb{E}[Y - \hat{Y}|\mathcal{G}]] \\ &\stackrel{(1)}{=} \mathbb{E}[(\hat{Y} - Z)(\mathbb{E}[Y|\mathcal{G}] - \mathbb{E}[\hat{Y}|\mathcal{G}])] \\ &\stackrel{(7)}{=} \mathbb{E}[(\hat{Y} - Z)(\hat{Y} - \hat{Y}\mathbb{E}[1|\mathcal{G}])] \\ &= \mathbb{E}[(\hat{Y} - Z)(\hat{Y} - \hat{Y})] = 0 \end{aligned}$$

Dunque:

$$\mathbb{E}[(Y - Z)^2] = \mathbb{E}[(Y - \hat{Y})^2] + \mathbb{E}[(\hat{Y} - Z)^2] \geq \mathbb{E}[(Y - \hat{Y})^2] \quad \forall Z \in L^2(\mathcal{G})$$

Inoltre l'uguaglianza vale se e solo se  $Z = \hat{Y}$ . □



## 15.4 Varianza condizionata

Se  $Y \in L^2$ , è noto che  $Var(Y) := \mathbb{E}[(Y - \mathbb{E}[Y])^2] = \mathbb{E}[Y^2] - \mathbb{E}^2[Y]$ , e inoltre che  $Var(Y|X = s) = \int_{\mathbb{R}} (t - \mathbb{E}[Y|X = s])^2 P^{Y|X}(dt|s) = q^2(s)$ .

### Definizione 15.9

Data  $Y \in L^2(\mathcal{A})$  e  $\mathcal{G}$  sotto- $\sigma$ -algebra di  $\mathcal{A}$ , la **varianza condizionata di  $Y$  data  $\mathcal{G}$**  è definita nel seguente modo:

$$Var(Y|\mathcal{G}) := \mathbb{E}\left[(Y - \mathbb{E}[Y|\mathcal{G}])^2 | \mathcal{G}\right]$$

Essa ha le seguenti proprietà:

1.  $Var(Y|\mathcal{G}) \geq 0$  qc, grazie alla proprietà (2) dell'attesa condizionata.
2.  $Var(Y|\mathcal{G}) = \mathbb{E}[Y^2|\mathcal{G}] - (\mathbb{E}[Y|\mathcal{G}])^2$ . Infatti, definendo come prima  $\hat{Y} = \mathbb{E}[Y|\mathcal{G}]$ , si ha:

$$\begin{aligned} Var(Y|\mathcal{G}) &= \mathbb{E}\left[(Y - \hat{Y})^2 | \mathcal{G}\right] \\ &= \mathbb{E}[Y^2|\mathcal{G}] - 2\mathbb{E}[Y\hat{Y}|\mathcal{G}] + \mathbb{E}[\hat{Y}^2|\mathcal{G}] \\ &= \mathbb{E}[Y^2|\mathcal{G}] - 2\hat{Y}^2 + \hat{Y}^2 \\ &= \mathbb{E}[Y^2|\mathcal{G}] - \hat{Y}^2 \end{aligned}$$

3. Se  $\mathcal{G} = \sigma(X)$ ,  $Var(Y|\sigma(X)) = Var(Y|X) = q^2(X)$ , grazie alla proprietà (13) dell'attesa condizionata.
4. È vera la seguente proposizione.

### Proposizione 15.10: formula di scomposizione della varianza

Sia  $Y \in L^2(\mathcal{A})$  e  $\mathcal{G} \subseteq \mathcal{A}$  sotto- $\sigma$ -algebra di  $\mathcal{A}$ . Allora:

$$Var(Y) = Var(\mathbb{E}[Y|\mathcal{G}]) + \mathbb{E}[Var(Y|\mathcal{G})]$$

Dalla formula discende immediatamente che  $Var(Y) \geq Var(\mathbb{E}[Y|\mathcal{G}])$ : ottenendo nuove informazioni (rappresentate dalla  $\sigma$ -algebra condizionante  $\mathcal{G}$ ) la varianza sui dati si riduce.

### Dimostrazione

Come al solito si ponga  $\hat{Y} = \mathbb{E}[Y|\mathcal{G}]$ . Allora:

$$\begin{aligned} Var(Y) &= \mathbb{E}\left[(Y - \mathbb{E}[Y] + \hat{Y} - \hat{Y})^2\right] \\ &= \mathbb{E}\left[(Y - \hat{Y})^2\right] + 2\mathbb{E}\left[(Y - \hat{Y})(\hat{Y} - \mathbb{E}[Y])\right] + \mathbb{E}\left[(\hat{Y} - \mathbb{E}[Y])^2\right] \end{aligned}$$

Il primo addendo, grazie alla proprietà (3) dell'attesa condizionata, è uguale a  $\mathbb{E} \left[ \mathbb{E}[(Y - \hat{Y})^2 | \mathcal{G}] \right] =: \mathbb{E} [Var(Y | \mathcal{G})]$ . Il doppio prodotto è uguale a 0, come visto nella dimostrazione precedente (prendendo  $Z = \mathbb{E}[Y]$  che in quanto costante appartiene a  $L^2$ ). L'ultimo addendo è infine:

$$\mathbb{E} \left[ (\hat{Y} - \mathbb{E}[\hat{Y}])^2 \right] = Var(\hat{Y}) =: Var(\mathbb{E}[Y | \mathcal{G}]) \quad \square$$

*Esempio.* Riprendiamo il dado e le monete delle pagine precedenti. Date le VA  $D \sim U(1, \dots, 6)$  e  $T|D = n \sim Bi(n, \frac{1}{2})$ , si può calcolare la varianza come:

$$\begin{aligned} Var(T) &= \mathbb{E}[Var(T|D)] + Var(\mathbb{E}[T|D]) \\ &= \mathbb{E} \left[ \frac{D}{4} \right] + Var \left( \frac{D}{2} \right) = \frac{1}{4} \frac{(1+6)}{2} + \frac{1}{4} \frac{(6^2 - 1^2)}{12} = \frac{71}{48} \end{aligned}$$

*Esempio.* Dato un vettore gaussiano  $(X, Y) \sim \mathcal{N}$ , la varianza risulta:

$$\begin{aligned} Var(\mathbb{E}[Y|X]) + \mathbb{E}[Var(Y|X)] &= \\ &= Var \left( \mu_Y + \rho \frac{\sigma_X}{\sigma_Y} (X - \mu_X) \right) + \mathbb{E} [\sigma_Y^2 (1 - \rho^2)] \\ &= \rho^2 \frac{\sigma_Y^2}{\sigma_X^2} \sigma_X^2 + \sigma_Y^2 (1 - \rho^2) = \sigma_Y^2 \end{aligned}$$

---

## 16 Convergenza di variabili aleatorie

Lo studio delle *successioni di variabili aleatorie* in probabilità è tanto importante quanto lo studio delle successioni numeriche e funzionali in analisi. In particolare, si vuole studiare la loro *convergenza* a un limite. Nel caso delle variabili aleatorie esistono, in realtà, diverse definizioni di limite, e conseguentemente di convergenza, alcune più “forti” di altre. In questo capitolo analizzeremo i principali tipi di convergenza, le loro proprietà, alcuni criteri e teoremi, il legame tra le varie convergenze, ed esempi e controesempi notevoli.

**NB:** I contenuti di questo capitolo e dei seguenti si estendono anche al caso in cui  $X$  sia un vettore aleatorio e non una semplice variabile aleatoria (con una sola ovvia eccezione che sarà fatta notare al momento opportuno). Si è scelto di usare il caso in  $\mathbb{R}$  per semplicità di trattazione e di notazione.

Ricordiamo dall’analisi che una successione di funzioni  $f_n : \mathbb{R}^k \rightarrow \mathbb{R}$  può convergere alla sua funzione limite  $f : \mathbb{R}^k \rightarrow \mathbb{R}$  in due modi:

- *convergenza puntuale*, se  $f_n(x) \xrightarrow{n} f(x) \forall x \in \mathbb{R}^k$ ;
- *convergenza uniforme*, se  $\forall \varepsilon > 0 \exists n_0 = n_0(\varepsilon) : |f_n(x) - f(x)| < \varepsilon \forall x \in \mathbb{R}^k \forall n \geq n_0$ .

Per tutto il capitolo, sia  $(\Omega, \mathcal{A}, \mathbb{P})$  lo spazio di probabilità su cui definiamo le variabili aleatorie  $X_n : \Omega \rightarrow \mathbb{R}$  e  $X : \Omega \rightarrow \mathbb{R}$ .

### 16.1 Convergenza certa e quasi certa

#### Definizione 16.1

La successione di VAR  $X_n$  **converge certamente** a  $X$  se  $X_n(\omega) \rightarrow X(\omega) \quad \forall \omega \in \Omega$ .

#### Definizione 16.2

La successione di VAR  $X_n$  **converge quasi certamente** a  $X$  (indicato con la notazione  $X \xrightarrow{\text{qc}} X$ ) se l’evento  $A = (X_n \rightarrow X)$  è tale che  $A \in \mathcal{A}$  e  $\mathbb{P}(A) = 1$ .

*Esempio.* Alcune convergenze certe e quasi certe:

- Siano  $f, f_n : \mathbb{R} \rightarrow \mathbb{R}$  funzioni continue, tali per cui  $f_n$  converga puntualmente a  $f$ .  
Sia inoltre lo spazio di probabilità  $\Omega = \mathbb{R}, \mathcal{A} = \mathcal{B}, \mathbb{P} = \mathcal{N}(0, 1)$ , su cui si definiscono le variabili aleatorie  $X_n(\omega) = f_n$  e  $X(\omega) = f, \forall \omega \in \Omega$ .  
Allora  $X_n \rightarrow X$  certamente, e dunque anche quasi certamente. Si presti attenzione al fatto che le  $X_n$  siano uguali alle  $f_n$ , non distribuite come le  $f_n$ .

- Sia  $X_n = \mathbb{1}_{[0, \frac{1}{n}]}$ .  
 $X_n$  converge certamente a  $X$ , VA così definita:

$$X = \begin{cases} 0 & \text{per } \omega \neq 0 \\ 1 & \text{per } \omega = 0 \end{cases}$$

Inoltre,  $X_n \xrightarrow{qc} 0$ . Infatti,  $X_n \rightarrow X$  per gli  $\omega$  appartenenti ad  $A = \mathbb{R}_n \setminus \{0\}$ .

- Sia  $X$  una VAR. Siano inoltre  $X_n = \frac{X}{n}$ . Allora  $X_n \rightarrow 0$  certamente, e dunque anche quasi certamente.
- Sia  $X \geq 0$  una VAR.  
 Abbiamo già visto che è possibile costruire una successione di VA semplici tale per cui  $X_n \uparrow X \forall \omega$ . Questo è un caso di convergenza certa.
- Sia  $X$  una VAR di legge  $\mathcal{G}(P)$ . Siano inoltre  $X_n = X \wedge n = \min(X, n)$ . Allora  $X_n \xrightarrow{qc} X$ .

Enunciamo ora alcune proprietà della convergenza quasi certa. Alcune sembreranno scontate, ma vedremo che non tutte varranno per gli altri tipi di convergenza.

**Proposizione 16.3**

Siano le VAR  $X_n, Y_n (\forall n), X$  e  $Y$  tali che  $X_n \xrightarrow{qc} X, Y_n \xrightarrow{qc} Y$ . Siano inoltre  $a \in \mathbb{R}$  ed  $h : \mathbb{R} \rightarrow \mathbb{R}$  funzione continua. Allora valgono le seguenti proprietà:

1.  $aX_n \xrightarrow{qc} aX$
2.  $h(X_n) \xrightarrow{qc} h(X)$
3.  $X_n + Y_n \xrightarrow{qc} X + Y$
4.  $X_n Y_n \xrightarrow{qc} XY$
5.  $Y, Y_n \neq 0 \forall \omega \implies \frac{X_n}{Y_n} \xrightarrow{qc} \frac{X}{Y}$

**Dimostrazione**

Sarà dimostrato solo il punto (3).

Per ipotesi,  $X_n \xrightarrow{qc} X$  e  $Y_n \xrightarrow{qc} Y$ . Siano gli eventi  $A_1 = (X_n \rightarrow X)$  e  $A_2 = (Y_n \rightarrow Y)$ . Per definizione di convergenza quasi certa,  $\mathbb{P}(A_1) = \mathbb{P}(A_2) = 1$ .

Se  $X_n$  e  $Y_n$  convergono, necessariamente converge anche la loro somma<sup>48</sup>: da questa condizione necessaria discende che  $(X_n + Y_n \rightarrow X + Y) \supseteq A_1 \cap A_2$ .

Dunque  $\mathbb{P}(X_n + Y_n \rightarrow X + Y) \geq \mathbb{P}(A_1, A_2) = 1$ . Ma allora anche  $\mathbb{P}(X_n + Y_n \rightarrow X + Y) = 1$  e vale la convergenza quasi certa.  $\square$

---

<sup>48</sup>“Perché lo dice Verri”

## 16.2 Convergenza in $L^p$

### Definizione 16.4

La successione di VAR  $X_n$  **converge in  $L^p$**  a  $X$  ( $X_n \xrightarrow{L^p} X$ ) se  $X, X_n \in L^p$  e  $\mathbb{E}[|X_n - X|^p] \rightarrow 0$ .

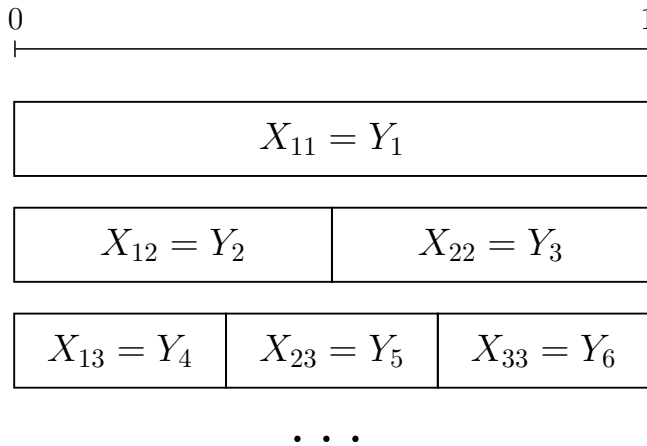
*Esempio.* Consideriamo le VA  $X_n = \mathbb{1}_{[0, \frac{1}{n}]} \sim \mathcal{B}(\frac{1}{n})$ , definite su  $\Omega = \mathbb{R}$ ,  $\mathcal{A} = \mathcal{B}$ ,  $\mathbb{P} = \mathcal{U}([0, 1])$ .

Allora  $X_n \xrightarrow{L^1} 0$ . Infatti  $X_n \in L^1, 0 \in L^1, \mathbb{E}[|X_n - 0|] = \mathbb{E}[X_n] = \frac{1}{n} \rightarrow 0$ .

*Controesempio:* macchina da scrivere. Si considerino le VAR  $X_{jk} = \mathbb{1}_{[\frac{j-1}{k}; \frac{j}{k}]}$ , con  $k \in \mathbb{N}$  e  $1 \leq j \leq k$ .

Si faccia poi corrispondere a ogni  $X_{jk}$  una  $Y_i$ , in modo da avere un solo indice:  $X_{11} = Y_1, X_{12} = Y_2, X_{22} = Y_3, X_{13} = Y_4 \dots$

$Y_n \xrightarrow{L^1} 0$ , infatti  $\mathbb{E}[|Y_n|] \rightarrow 0$ ; ma  $Y_n \not\xrightarrow{q.c.} 0$ . Anzi,  $\liminf_n Y_n = 0$  e  $\limsup_n Y_n = 1, \forall \omega \in [0, 1]$ : il limite della successione nemmeno esiste (perché i limiti inferiore e superiore non coincidono), men che meno c'è convergenza con probabilità 1!



### Proposizione 16.5

Siano le VAR  $X_n \xrightarrow{L^p} X, Y_n \xrightarrow{L^p} Y$ . Sia inoltre  $a \in \mathbb{R}$ . Allora valgono le seguenti proprietà:

1.  $aX_n \xrightarrow{L^p} aX$ ;
2.  $X_n + Y_n \xrightarrow{L^p} X + Y$ ;

3. se  $q \in [1, p]$ , allora  $\mathbb{E}[|X_n|^q] \xrightarrow{n} \mathbb{E}[|X|^q]$ .

Le due proprietà “base” di uno spazio vettoriale (somma e prodotto per scalare) sono garantite, il che è intuitivamente sensato, in quanto gli  $L^p$  sono appunto spazi vettoriali. Mancano tuttavia le proprietà di prodotto e rapporto. Sappiamo infatti che non è detto, per esempio, che il prodotto di due VA  $L^1$  sia una VA  $L^2$ : mancando l'appartenenza a priori, non si può nemmeno parlare a priori di convergenza.

I casi più interessanti di applicazione delle proprietà sono la varianza e il valore atteso:

- $p = 1 \implies \mathbb{E}[X_n] \xrightarrow{n} \mathbb{E}[X]$
- $p = 2 \implies \mathbb{E}[X_n] \xrightarrow{n} \mathbb{E}[X], \text{Var}(X_n) \xrightarrow{L^p} \text{Var}(X)$

**NB:** Si ricorda che, presi  $1 \leq p \leq q < +\infty$ , gli spazi  $L^p$  sono “inscatolati”:  
 $L^1 \supseteq L^p \supseteq L^q$ .

### 16.3 Convergenza in probabilità

#### Definizione 16.6

La successione di VAR  $X_n$  **converge in probabilità** a  $X$  ( $X_n \xrightarrow{\mathbb{P}} X$ ) se  $\forall \varepsilon > 0, \mathbb{P}(|X_n - X| > \varepsilon) \xrightarrow{n} 0$ .

*Esempio.* Riprendiamo la macchina da scrivere.  $Y_n \xrightarrow{\mathbb{P}} 0$ , poiché  $\mathbb{P}(|Y_n| > \varepsilon) = \mathbb{P}(Y_n = 1) \xrightarrow{n} 0 \forall \varepsilon > 0$ .

#### Proposizione 16.7

Siano le VAR  $X_n \xrightarrow{\mathbb{P}} X, Y_n \xrightarrow{\mathbb{P}} Y$ . Siano inoltre  $a \in \mathbb{R}$  e  $h : \mathbb{R} \rightarrow \mathbb{R}$  funzione continua. Allora valgono le seguenti proprietà:

1.  $aX_n \xrightarrow{\mathbb{P}} aX$
2.  $h(X_n) \xrightarrow{\mathbb{P}} h(X)$
3.  $X_n + Y_n \xrightarrow{\mathbb{P}} X + Y$
4.  $X_n Y_n \xrightarrow{\mathbb{P}} XY$
5.  $Y, Y_n \neq 0 \forall \omega \implies \frac{X_n}{Y_n} \xrightarrow{\mathbb{P}} \frac{X}{Y}$

Si noti che sono le stesse 5 proprietà della convergenza quasi certa.

**Teorema 16.8** (*J-P Th 17.1*)

$X_n$  converge in probabilità a  $X \iff \mathbb{E} \left[ \frac{|X_n - X|}{1 + |X_n - X|} \right] \rightarrow 0$ .

**Dimostrazione**

Sia  $\varepsilon > 0$  fissato e  $X = 0$  senza perdita di generalità (infatti grazie alla proprietà (3) si può considerare la successione  $Y_k := X_n - X \rightarrow 0$ : la tesi per  $X = 0$  equivale dunque alla tesi per  $X$  generica).

- ( $\implies$ ): ricordando che  $\frac{|x|}{1+|x|} \leq 1 \quad \forall x$  e che  $\frac{|x|}{1+|x|} \leq \varepsilon \quad \forall x : |x| \leq \varepsilon$ , si può scrivere:

$$\begin{aligned} \frac{|X_n|}{1+|X_n|} &= \frac{|X_n|}{1+|X_n|} \underbrace{(\mathbb{1}_{\{|X_n|>\varepsilon\}} + \mathbb{1}_{\{|X_n|\leq\varepsilon\}})}_{=1} \\ &\leq 1 \cdot \frac{|X_n|}{1+|X_n|} \mathbb{1}_{\{|X_n|>\varepsilon\}} + \varepsilon \cdot \mathbb{1}_{\{|X_n|\leq\varepsilon\}} \\ &\leq \mathbb{1}_{\{|X_n|>\varepsilon\}} + \varepsilon \cdot 1 \end{aligned}$$

Passando al valore atteso:

$$\mathbb{E} \left[ \frac{|X_n|}{1+|X_n|} \right] \leq \underbrace{\mathbb{P}(|X_n| > \varepsilon)}_{\rightarrow 0 \text{ per hp}} + \varepsilon \implies \limsup_n \mathbb{E} \left[ \frac{|X_n|}{1+|X_n|} \right] \leq \varepsilon \quad \forall \varepsilon > 0$$

Quindi, per definizione di convergenza, abbiamo la tesi:

$$\mathbb{E} \left[ \frac{|X_n|}{1+|X_n|} \right] \rightarrow 0$$

Infatti  $\mathbb{E} \left[ \frac{|X_n|}{1+|X_n|} \right]$  è una VA sempre non negativa.

- ( $\impliedby$ ): Si può scrivere che:

$$\frac{\varepsilon}{1+\varepsilon} \mathbb{1}_{\{|X_n|>\varepsilon\}} \leq \frac{|X_n|}{1+|X_n|} \mathbb{1}_{\{|X_n|>\varepsilon\}} \leq \frac{|X_n|}{1+|X_n|}$$

Quindi, passando al valore atteso:

$$\frac{\varepsilon}{1+\varepsilon} \mathbb{P}(|X_n| > \varepsilon) \leq \mathbb{E} \left[ \frac{|X_n|}{1+|X_n|} \right] \rightarrow 0 \implies \mathbb{P}(|X_n| > \varepsilon) \rightarrow 0 \quad \forall \varepsilon > 0$$

La tesi è dunque dimostrata.  $\square$

## 16.4 Relazioni tra tipi di convergenza (1)

**Teorema 16.9** (*J-P Th 17.2*)

1.  $X_n \xrightarrow{qc} X \implies X_n \xrightarrow{\mathbb{P}} X$
2.  $X_n \xrightarrow{L^p} X \implies X_n \xrightarrow{\mathbb{P}} X$
3.  $X_n \xrightarrow{L^q} X \implies X_n \xrightarrow{L^p} X$ , per  $1 \leq p \leq q < +\infty$

Si noti che per il punto (1) non vale l'implicazione inversa: si veda come controesempio la macchina da scrivere (pagina 189). Similmente, esistono controesempi in cui  $\int \lim \neq \lim \int$  che smentiscono il viceversa del punto (2).

**Dimostrazione**

1.  $X_n - X \rightarrow 0$ . Allora si può applicare la funzione continua  $h(x)$  così definita:

$$h(x) = \frac{|x|}{1 + |x|} \implies \frac{|X_n - X|}{1 + |X_n - X|} \xrightarrow{qc} 0$$

Si ha inoltre  $\frac{|X_n - X|}{1 + |X_n - X|} \leq 1 \in L^1$ . Allora, per convergenza dominata:

$$\mathbb{E} \left[ \frac{|X_n - X|}{1 + |X_n - X|} \right] \rightarrow 0$$

2. Fissato  $\varepsilon > 0$ , si ha:

$$\mathbb{P}(|X_n - X| > \varepsilon) = \mathbb{P}(|X_n - X|^p > \varepsilon^p) \leq \mathbb{P}(|X_n - X|^p \geq \varepsilon^p)$$

Applicando la disuguaglianza di Markov (teorema 6.22), si trova:

$$\mathbb{P}(|X_n - X|^p \geq \varepsilon^p) \leq \frac{\mathbb{E}[|X_n - X|^p]}{\varepsilon^p} \rightarrow 0$$

3. La dimostrazione di questo punto richiede conoscenze avanzate, che lo studente più motivato apprenderà in corsi successivi.  $\square$

**Teorema 16.10** (*J-P Th 17.3*)

Se  $X_n$  converge in probabilità a  $X$ , allora esiste una successione numerica  $n_k$ , cui corrisponde una sottosuccessione di VA  $X_{n_k}$ , tale che  $X_{n_k} \xrightarrow{qc} X$ .

**Dimostrazione**

Per ipotesi,  $X_n \xrightarrow{\mathbb{P}} X$ , e dunque:

$$\mathbb{E} \left[ \frac{|X_n - X|}{1 + |X_n - X|} \right] \rightarrow 0$$



Ma allora esiste una sottosuccessione  $X_{n_k}$  tale per cui:

$$\mathbb{E} \left[ \frac{|X_{n_k} - X|}{1 + |X_{n_k} - X|} \right] \leq \frac{1}{2^k}$$

Consideriamo allora il valore atteso della somma. Poiché i termini sono positivi, si ha:

$$\mathbb{E} \left[ \sum_{k=1}^{+\infty} \frac{|X_{n_k} - X|}{1 + |X_{n_k} - X|} \right] = \sum_{k=1}^{+\infty} \mathbb{E} \left[ \frac{|X_{n_k} - X|}{1 + |X_{n_k} - X|} \right] \leq \sum_{k=1}^{+\infty} \frac{1}{2^k} < +\infty$$

Poiché il valore atteso della somma converge:

$$\sum_{k=1}^{+\infty} \frac{|X_{n_k} - X|}{1 + |X_{n_k} - X|} \stackrel{\text{qc}}{<} +\infty$$

Allora  $\frac{|X_{n_k} - X|}{1 + |X_{n_k} - X|} \stackrel{\text{qc}}{\rightarrow} 0$ , da cui  $|X_{n_k} - X| \stackrel{\text{qc}}{\rightarrow} 0$ , e infine  $X_{n_k} \stackrel{\text{qc}}{\rightarrow} X$ .  $\square$

*Esercizio.* Si applichi il teorema appena dimostrato all'esempio della macchina da scrivere.

Si prenda la sottosuccessione di VA  $X_{11}, X_{22}, X_{33}, \dots$ , che corrisponde all'ultima  $X_{jk}$  per ogni riga  $k$ . Essa converge a  $\mathbb{1}_{\{1\}}(\omega)$ , ovvero a 1 se l'esito elementare  $\omega$  è proprio 1 e a 0 altrove. Inoltre si noti che  $\mathbb{1}_{\{1\}}(\omega) = 0$  quasi certamente: dunque,  $X_{nn} \stackrel{\text{qc}}{\rightarrow} 0$ . La successione in questione è  $X_{nn}$  con  $n \in \mathbb{N}$ , ovvero  $Y_{n_k}$  con  $n_k = \frac{n(n+1)}{2}$  (si può facilmente verificare che i numeri 1, 3, 6, ... di  $n_k$  sono tutti numeri triangolari).

**Teorema 16.11** (*J-P Th 17.4*)

Siano le VAR  $X_n, X, Y$  tali per cui:  $X_n \xrightarrow{\mathbb{P}} X$ ,  $|X_n| \leq Y$  qc  $\forall n$  e  $Y \in L^p$  con  $1 \leq p < +\infty$ . Allora:

$$|X| \in L^p \quad \text{e} \quad X_n \xrightarrow{L^p} X$$

**Teorema 16.12: unicità del limite**

Sia  $X_n$  una successione di VAR, e siano  $X$  e  $\tilde{X}$  VAR tali che  $X_n \rightarrow X$  e  $X_n \rightarrow \tilde{X}$ , sia essa convergenza quasi certa, in  $L^p$  o in probabilità. Allora  $\tilde{X} \stackrel{\text{qc}}{=} X$ .

**Dimostrazione**

Prendiamo il caso della convergenza quasi certa. Per definizione, dati  $A, \tilde{A} \in \mathcal{A}$  eventi quasi certi, si ha che:

$$X_n(\omega) \rightarrow X(\omega) \quad \forall \omega \in A \quad \text{e} \quad X_n(\omega) \rightarrow \tilde{X}(\omega) \quad \forall \omega \in \tilde{A}$$

Si definisca  $A_0 = A \cap \tilde{A} \in \mathcal{A}$ , che è anch'esso un evento quasi certo. Pertanto:

$$X_n(\omega) \rightarrow X(\omega) \quad \text{e} \quad X_n(\omega) \rightarrow \tilde{X}(\omega) \quad \forall \omega \in A_0$$

Per l'unicità del limite puntuale, si ha che  $X(\omega) = \tilde{X}(\omega) \quad \forall \omega \in A_0$ , il quale è un evento quasi certo.

La dimostrazione per gli altri tipi di convergenza è una semplice conseguenza del caso trattato.  $\square$

## 16.5 Convergenza in legge

Si vuole spostare l'attenzione dai valori delle VA, osservati misurando per esempio  $|X_n - X|$ , alle loro leggi: che relazione c'è tra  $P^{X_n}$  e  $P^X$ ?

### Definizione 16.13

Siano  $\mathbb{P}_n$  e  $\mathbb{P}$  probabilità su  $(\mathbb{R}, \mathcal{B})$ . Si dice che  $\mathbb{P}_n$  **converge debolmente** a  $\mathbb{P}$  ( $\mathbb{P}_n \xrightarrow{\text{deb}} \mathbb{P}$ ) se:

$$\int_{\mathbb{R}} h \, d\mathbb{P}_n \xrightarrow{n} \int_{\mathbb{R}} h \, d\mathbb{P} \quad \forall h : \mathbb{R} \rightarrow \mathbb{R} \text{ continua e limitata}$$

La condizione di limitatezza di  $h$  serve semplicemente a garantire la convergenza dell'integrale a prescindere dalla probabilità considerata. Di gran lunga più significativa è invece la richiesta di continuità della  $h$ , della cui utilità forniremo ora un esempio:

### Proposizione 16.14

Sia  $x_n$  una successione numerica reale tale che  $x_n \rightarrow x \in \mathbb{R}$ . Allora  $\delta_{x_n} \xrightarrow{n} \delta_x$  debolmente.

### Dimostrazione

Sia  $h$  funzione qualsiasi continua e limitata. Allora:

$$\lim_{n \rightarrow +\infty} \int_{\mathbb{R}} h \, d\delta_{x_n} = \lim_{n \rightarrow +\infty} h(x_n) = h\left(\lim_{n \rightarrow +\infty} x_n\right) = h(x) = \int_{\mathbb{R}} h \, d\delta_x \quad \square$$

Come si può osservare, la continuità è necessaria per permettere lo scambio di limiti (quello di  $n$  e l'integrale) che rappresenta la definizione di convergenza debole; non basta dunque  $h$  solo misurabile per garantire tale convergenza sulla più semplice delle densità.

*Controesempio.* Cerchiamo una  $x_n$  successione numerica reale e  $h$  misurabile e limitata (ma non continua) tali che  $x_n \rightarrow x$  ma  $\int_{\mathbb{R}} h \, d\delta_{x_n} \not\rightarrow \int_{\mathbb{R}} h \, d\delta_x$ . Un possibile controesempio è dato da  $x_n = \frac{1}{n} \rightarrow 0 = x$  con  $h(t) = \mathbb{1}_{(t=0)}$  che è misurabile in quanto indicatrice: infatti  $h(x_n) = 0 \quad \forall n$  e  $h(x_n) \not\rightarrow h(x) = 1$ .

**Definizione 16.15**

Siano  $X_n$  e  $X$  VAR. Si dice che  $X_n$  **converge in legge** o in distribuzione a  $X$  ( $X_n \xrightarrow{\mathcal{L}} X$ ) se  $P^{X_n} \xrightarrow{\text{deb}} P^X$ .

Si noti che la convergenza in legge è una *relazione tra leggi* e solo tra leggi: non indica alcun legame tra i valori delle variabili aleatorie, bensì solo sulla loro probabilità. Per evidenziare ciò, spesso si indica  $P^X$  al posto di  $X$  nella scrittura del limite. Per esempio, se  $X \sim \mathcal{N}(0, 1)$ , si scriverà  $X_n \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1)$ .

**Teorema 16.16** (*J-P Th 18.1*)

$X_n \rightarrow X$  in legge  $\iff \mathbb{E}_n[h(X_n)] \rightarrow \mathbb{E}[h(X)] \quad \forall h : \mathbb{R} \rightarrow \mathbb{R}$  continua e limitata

**Dimostrazione**

$$\begin{aligned} X_n \xrightarrow{\mathcal{L}} X &\iff P^{X_n} \xrightarrow{\text{deb}} P^X \iff \int_{\mathbb{R}} h \, dP^{X_n} \rightarrow \int_{\mathbb{R}} h \, dP^X \\ &\iff \mathbb{E}_n[h(X_n)] \rightarrow \mathbb{E}[h(X)] \quad \forall h \text{ cont. e lim.} \end{aligned}$$

L'ultimo passaggio vale per la regola del valore atteso. □

Si noti che la complicazione tra la scrittura con il simbolo di integrale e quella del valore atteso sembra una ridondanza, ma in realtà esse esprimono due concetti differenti. Nell'integrale la funzione integranda è sempre la stessa e cambia invece la probabilità che "agisce" su di essa; invece nel valore atteso è la funzione  $h$  a cambiare valore e a convergere.

**NB:** A differenza degli altri limiti (incluso quello della convergenza debole), il *limite in legge* non è *unico*; ovvero, se  $X_n$  converge in legge a due variabili  $X$  e  $\tilde{X}$  si può concludere che  $P^X = P^{\tilde{X}}$ , ma *non* che  $X = \tilde{X}$ , nemmeno qc. Ancora una volta, distinguere tra convergenza di leggi di VA e convergenza di valori di VA è fondamentale.

## 16.6 Criteri di convergenza in legge

### 16.6.1 Caso discreto

**Proposizione 16.17**

Siano  $X_n$  e  $X$  VAR discrete, e sia  $S = S_X \cup \left( \bigcup_{n=1}^{+\infty} S_n \right)$  l'unione dei loro supporti, al più numerabile. Allora:

$$X_n \xrightarrow{\mathcal{L}} X \iff \mathbb{P}(X_n = s) \rightarrow \mathbb{P}(X = s) \quad \forall s \in S$$

*Esempio:* binomiale  $\rightarrow$  Poisson.

Siano  $X_n \sim Bi\left(n, \frac{\lambda}{n}\right)$ , con  $n \geq \lambda > 0$ , e  $X \sim Po(\lambda)$ . Il supporto è:

$$S_{X_n} = \{0, \dots, n\} \quad \forall n, \quad S_X = \{0, 1, \dots\} \implies S = \{0, 1, \dots\}$$

Fissando  $k \in S$ , si ha che:

$$\mathbb{P}(X_n = k) = \begin{cases} 0 & \text{per } n < k \\ \binom{n}{k} p^k (1-p)^{n-k} & \text{per } n \geq k \end{cases}$$

Per  $n \rightarrow +\infty$  il primo caso non si verificherà più (ovvero si verificherà con probabilità nulla): calcolando  $\lim_{n \rightarrow +\infty} \mathbb{P}(X_n = k)$  si ottiene,  $\forall k \in S$ , che:

$$\begin{aligned} \lim_{n \rightarrow +\infty} \frac{n!}{k!(n-k)!} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} \\ = \frac{\lambda^k}{k!} \lim_{n \rightarrow +\infty} \underbrace{\frac{n(n-1)\dots(n-k+1)}{n^k}}_{\rightarrow 1} \underbrace{\left(1 - \frac{\lambda}{n}\right)^n}_{\rightarrow e^{-\lambda}} \underbrace{\left(1 - \frac{\lambda}{n}\right)^{-k}}_{\rightarrow 1} = e^{-\lambda} \frac{\lambda^k}{k!} \end{aligned}$$

Dunque  $X_n \xrightarrow{\mathcal{L}} X$ , ovvero:

$$X_n \text{ iid } Bi\left(n, \frac{\lambda}{n}\right) \implies X_n \xrightarrow{\mathcal{L}} Po(\lambda)$$

### 16.6.2 Caso continuo

**Teorema 16.18** (*J-P Th 18.5*)

Siano  $X_n, X$  VAR continue di densità continue rispettivamente  $f_n$  e  $f$ . Allora:

$$f_n \rightarrow f \text{ qo}^{49} \implies X_n \xrightarrow{\mathcal{L}} X$$

*Esempio.*  $\mathcal{N} \rightarrow \mathcal{N}$ .

Siano  $X_n \sim \mathcal{N}(\mu, \sigma_n^2)$  tali che,  $\sigma_n^2 > 0 \quad \forall n, \mu_n \rightarrow \mu$  e  $\sigma_n^2 \rightarrow \sigma^2 > 0$ . Per definizione:

$$f_n(s) = \frac{1}{\sqrt{2\pi\sigma_n^2}} \exp\left\{-\frac{(s - \mu_n)^2}{2\sigma_n^2}\right\}$$

---

<sup>49</sup>Ricordiamo che “quasi ovunque” significa “convergenza puntuale tranne eventualmente in insiemi di misura di Lebesgue nulla”. Se a pagina 196 non avete ancora capito il concetto di “quasi ovunque” potrebbe essere un problema.

Per  $n$  che tende ad infinito e data una  $f_n$  continua, allora:

$$f_n(s) \xrightarrow[n \text{ cont.}]{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{\frac{-(s-\mu)^2}{2\sigma^2}\right\} = f(s) \quad \forall s \in \mathbb{R} = S$$

Ovvero, riassumendo:

$$X_n \sim \mathcal{N}(\mu_n, \sigma_n^2) \implies X_n \xrightarrow{\mathcal{L}} \mathcal{N}(\mu, \sigma^2)$$

Da ciò consegue, per definizione di convergenza in legge, che,  $\forall h$  continua e limitata:

$$\int_{\mathbb{R}} h(s) \frac{1}{\sqrt{2\pi\sigma_n^2}} \exp\left\{\frac{-(s-\mu_n)^2}{2\sigma_n^2}\right\} ds \xrightarrow{n} \int_{\mathbb{R}} h(s) \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{\frac{-(s-\mu)^2}{2\sigma^2}\right\} ds$$

Abbiamo, di fatto, lo scambio dei limiti anche senza aver invocato la convergenza monotona o dominata; esse sono infatti condizioni solo sufficienti, e non necessarie, per lo scambio.

In generale, se  $X_n \xrightarrow{\mathcal{L}} X$ , non si può dedurre che  $\mathbb{P}(X_n \in B) \rightarrow \mathbb{P}(X \in B)$  con  $B \in \mathcal{B}$ , che  $\mathbb{E}[X_n] \rightarrow \mathbb{E}[X]$  o che  $Var(X_n) \rightarrow Var(X)$ .

*Esercizio.* Trovare controesempi per le ultime due affermazioni.

### 16.6.3 Caso generale: funzione di ripartizione

#### **Teorema 16.19** (*J-P Th 18.4*)

Siano  $X_n$  e  $X$  VAR con funzione di ripartizione  $F_n$  e  $F$ , rispettivamente.

Allora,  $\forall s \in \mathbb{R}$  in cui  $F$  è continua, ovvero esclusi eventualmente i salti di  $F$ :

$$X_n \xrightarrow{\mathcal{L}} X \iff F_n(s) \rightarrow F(s)$$

Quindi, se  $X$  è continua (e dunque anche  $F$  è continua), allora  $\mathbb{P}(X_n \in B) \rightarrow \mathbb{P}(X \in B) \quad \forall B \in \mathcal{B}$ .

Questa proprietà non è valida per i vettori aleatori in  $\mathbb{R}^n$ , visto che in spazi di dimensione maggiore di 1 la funzione di ripartizione diventa un oggetto più complicato da studiare e da utilizzare.

*Esempio.* Continua  $\rightarrow$  discreta.

Siano  $X_n \sim U\left(\left[-\frac{1}{n}, \frac{1}{n}\right]\right)$ . L'intuizione suggerisce che  $X_n$  convergerà a una legge massa (discreta).

Calcolare le funzioni di ripartizione delle VA consentirà di studiarne la convergenza in legge:

$$F_n(s) = \begin{cases} 0 & \text{per } s \leq -\frac{1}{n} \\ \frac{1}{2} - \frac{n}{2}s & \text{per } -\frac{1}{n} < s \leq \frac{1}{n} \\ 1 & \text{per } s > \frac{1}{n} \end{cases} \xrightarrow{n} \begin{cases} 0 & \text{per } s < 0 \\ \frac{1}{2} & \text{per } s = 0 \stackrel{\text{qo}}{=} \mathbf{1}_{[0, +\infty)}(s) = F_{\delta_0}(s) \\ 1 & \text{per } s > 0 \end{cases}$$

Ovvero:

$$X_n \sim U\left(\left[-\frac{1}{n}, \frac{1}{n}\right]\right) \implies X_n \xrightarrow{\mathcal{L}} \delta_0$$

Tale risultato si può anche scrivere come  $X_n \xrightarrow{\mathcal{L}} X$  con  $X \sim \delta_0$ , come  $X \stackrel{\text{qc}}{=} 0$  o come  $X_n \xrightarrow{\mathcal{L}} 0$ .

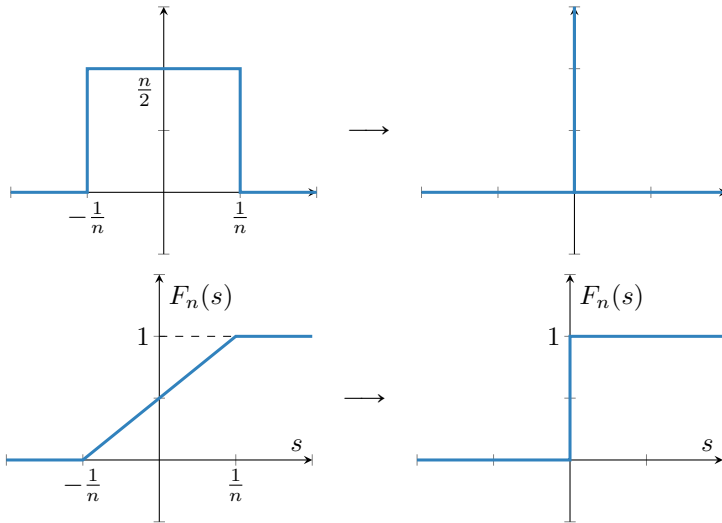


Figura 16.1: distribuzione e funzione di ripartizione da uniforme continua a delta discreta

*Esempio.* Discreta  $\rightarrow$  continua.

Siano  $X_n \sim U(A)$ , con  $A = \{\frac{1}{n}, \frac{2}{n}, \dots, 1\}$ . L'intuizione suggerisce che  $X_n$  convergerà a una legge uniforme sull'intervallo  $[0, 1]$  (continua).

Come già nell'esempio precedente, operiamo sulle funzioni di ripartizione. Si definisca quindi  $F_n(s)$ :

$$F_n(s) = \begin{cases} 0 & \text{per } s \leq \frac{1}{n} \\ \frac{k}{n} & \text{per } \frac{k}{n} < s \leq \frac{k+1}{n} \quad (k \in \mathbb{N}, k \leq n) \\ 1 & \text{per } s > 1 \end{cases}$$

Prendendo il limite su  $n$ :

$$F_n(s) \xrightarrow{n \rightarrow +\infty} \begin{cases} 0 & \text{per } s \leq 0 \\ s & \text{per } 0 < s \leq 1 \\ 1 & \text{per } s > 1 \end{cases} = F_{X \sim U([0,1])}(s) = F(s)$$

Ovvero, riassumendo:

$$X_n \sim U\left(\left\{\frac{1}{n}, \dots, 1\right\}\right) \implies X_n \xrightarrow{\mathcal{L}} U([0, 1])$$

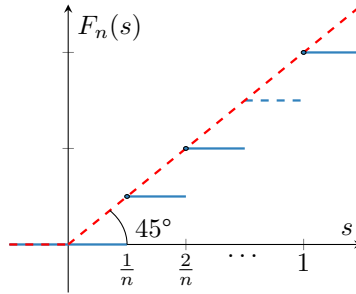


Figura 16.2: intuizione del passaggio da discreto a continuo

Per convincersi che un insieme discreto *numerabile* converga su un insieme *continuo*<sup>50</sup>, si osservi che:

$$|F_n(s) - F(s)| = s - F(s) < \frac{k+1}{n} - \frac{k}{n} = \frac{1}{n} \rightarrow 0 \quad \forall s \in \mathbb{R}$$

Ovvero, l'errore tra le due funzioni si annulla per  $n \rightarrow +\infty$ . In alternativa, si osservi che:

$$\mathbb{P}(a < X_n \leq b) = \frac{\#\{k \in \mathbb{N} | a < \frac{k}{n} \leq b\}}{n} \rightarrow b - a = \mathbb{P}(a < X \leq b)$$

Ovvero, la probabilità di un intervallo sulla uniforme discreta tende alla sua probabilità sull'uniforme continua.

#### 16.6.4 Caso generale: funzione caratteristica

È il caso più importante e utilizzato dei 4 menzionati finora.

**Teorema 16.20: continuità di Levy** (*J-P Th 19.1*)

1. Se  $X_n \xrightarrow{\mathcal{L}} X$ , allora  $\varphi_n(u) \rightarrow \varphi(u) \quad \forall u \in \mathbb{R}$ .
2. Siano  $X_n$  VAR tali che  $\varphi_n(u) \rightarrow \psi(u)$ , con  $\psi$  funzione generica ma continua in 0. Allora  $\exists X$  VAR avente  $\psi$  come funzione caratteristica e tale che  $X_n \xrightarrow{\mathcal{L}} X$ .

<sup>50</sup>No, seriamente, WTF?

Il punto (1) è una conseguenza dell'unicità della legge limite e della reciproca caratterizzazione tra una legge e la sua funzione caratteristica. L'intero teorema è sostanzialmente una coimplicazione tra convergenza in legge ("probabilistica") delle VA e convergenza puntuale ("analitica") delle rispettive funzioni caratteristiche.

Tuttavia per il punto (2) si richiedono ipotesi più semplici: non serve controllare che  $\psi$  sia effettivamente una funzione caratteristica, basta che sia continua in 0. A questo riguardo, si noti che se è verificata questa ipotesi allora necessariamente  $\psi(0) = 1$  perché  $\psi$  è la funzione limite delle  $\varphi$ , che hanno tutte valore 1 in 0.

*Esempio.* Siano  $X_n \sim \mathcal{N}(\mu_n, \sigma_n^2)$ ,  $\sigma^2 \geq 0$ ,  $\mu_n \rightarrow \mu$ ,  $\sigma_n^2 \rightarrow \sigma^2 \geq 0$ . Per definizione, la funzione caratteristica è:

$$\varphi_n(u) = \exp \left\{ iu\mu_n - \frac{1}{2}u^2\sigma_n^2 \right\}$$

Facendo tendere  $n$  all'infinito e per  $\varphi_n(u)$  continua in  $\mu_n$  e  $\sigma_n$ :

$$\varphi_n(u) \xrightarrow{n} \exp \left\{ iu\mu - \frac{1}{2}u^2\sigma^2 \right\} = \varphi_X(u) \text{ con } X \sim \mathcal{N}(\mu, \sigma^2)$$

In altre parole:

$$X_n \sim \mathcal{N}(\mu_n, \sigma_n^2) \implies X_n \xrightarrow{\mathcal{L}} \mathcal{N}(\mu, \sigma^2)$$

**NB:** Se  $\sigma^2 = 0$ , allora  $\mathcal{N}(\mu, \sigma^2) = \mathcal{N}(\mu, 0) = \delta_\mu$ , ovvero  $X_n \xrightarrow{\mathcal{L}} \mu$ .

*Esempio.* Siano  $X_n \sim Bi\left(n, \frac{\lambda}{n}\right)$  con  $n \geq \lambda > 0$ . Allora, per definizione:

$$\varphi_n(u) = \left( e^{iu\frac{\lambda}{n}} + 1 - \frac{\lambda}{n} \right)^n = \left[ 1 + \frac{1}{n}\lambda(e^{iu} - 1) \right]^n$$

Prendendo il limite su  $n$ :

$$\varphi_n(u) \xrightarrow{n} \exp\{\lambda(e^{iu} - 1)\} = \varphi_X(u) \text{ con } X \sim Po(\lambda)$$

Ovvero, riassumendo:

$$X_n \sim Bi\left(n, \frac{\lambda}{n}\right) \implies X_n \xrightarrow{\mathcal{L}} Po(\lambda)$$

*Controesempio.* La successione  $X_n \sim U([-n, n])$  non converge in legge, in quanto:

$$\varphi_n(u) = \frac{\sin(nu)}{nu} \xrightarrow{n} \begin{cases} 1 & \text{per } u = 0 \\ 0 & \text{per } u \neq 0 \end{cases} = \psi(u)$$

$\psi(u)$  non è infatti continua in  $u = 0$ .



## 16.7 Relazioni tra tipi di convergenza (2)

### Teorema 16.21 (J-P Th 18.2)

Siano  $X_n, X$  VAR tutte definite sullo stesso spazio di probabilità  $(\Omega, \mathcal{A}, \mathbb{P})$ . Allora:

$$X_n \xrightarrow{\mathbb{P}} X \implies X_n \xrightarrow{\mathcal{L}} X$$

Con questo teorema la convergenza in legge diventa la più “debole” tra i vari tipi di convergenza, perché è implicata dalla convergenza in probabilità che a sua volta era implicata da tutte le altre.

### Dimostrazione

Poiché per ipotesi  $X_n \xrightarrow{\mathbb{P}} X$ , è anche vero che  $h(X_n) \xrightarrow{\mathbb{P}} h(X)$  per tutte le  $h$  continue, comprese quelle continue e anche limitate. In quest’ultimo caso, applicando la convergenza dominata si ottiene che  $\mathbb{E}[h(X_n)] \rightarrow \mathbb{E}[h(X)] \forall h$  continua e limitata, ovvero, per definizione,  $X_n \xrightarrow{\mathcal{L}} X$ .  $\square$

Il viceversa *non* è valido perché il limite in legge non è unico, tranne che nel seguente caso specifico:

### Teorema 16.22 (J-P Th 18.3)

Siano  $X_n, X$  VAR tutte definite sullo stesso spazio di probabilità  $(\Omega, \mathcal{A}, \mathbb{P})$  e  $c \in \mathbb{R}$  una costante. Allora:

$$X_n \xrightarrow{\mathcal{L}} c \implies X_n \xrightarrow{\mathbb{P}} c$$

### Dimostrazione

$X_n \xrightarrow{\mathcal{L}} c, c \in \mathbb{R}$  se e solo se  $\mathbb{E}[h(X_n)] \rightarrow \mathbb{E}[h(c)] = h(c) \forall h$  continua e limitata. Scegliendo in particolare:

$$h(x) = \frac{|x - c|}{1 + |x - c|} \implies \mathbb{E} \left[ \frac{|X_n - c|}{1 + |X_n - c|} \right] \rightarrow h(c) = 0$$

Ovvero, per definizione,  $X_n \xrightarrow{\mathbb{P}} c$ .  $\square$

### Proposizione 16.23: proprietà della convergenza in legge

Siano  $X_n, X, Y_n$  VAR tali che  $X_n \xrightarrow{\mathcal{L}} X$  e  $Y_n \xrightarrow{\mathcal{L}} c \in \mathbb{R}, a \in \mathbb{R}$ , e  $h : \mathbb{R} \rightarrow \mathbb{R}$  funzione continua. Allora:

0.  $P^X$  è unica, come già sappiamo

1.  $aX_n \xrightarrow{\mathcal{L}} aX$

2.  $h(X_n) \xrightarrow{\mathcal{L}} h(X)$

Vale inoltre il *teorema di Slutsky*, riportato di seguito.

**Teorema 16.24: teorema di Slutsky** (*J-P Th 18.8*)

Nelle stesse ipotesi:

3.  $X_n + Y_n \xrightarrow{\mathcal{L}} X + c.$
4.  $X_n Y_n \xrightarrow{\mathcal{L}} cX.$
5.  $Y_n, c \neq 0 \implies \frac{X_n}{Y_n} \xrightarrow{\mathcal{L}} \frac{1}{c}X.$

Si noti che nella coppia di variabili aleatorie descritta dal teorema di Slutsky, almeno una delle due converge in legge a una *costante*: non potrebbe essere altrimenti, visto che l'eventuale legge di convergenza di una funzione di  $X$  e  $Y$  dipende anche dalle correlazioni tra  $X$  e  $Y$  (covarianza etc.), di cui non c'è menzione nelle ipotesi del teorema.

**16.7.1 Riepilogo delle relazioni fra convergenze**

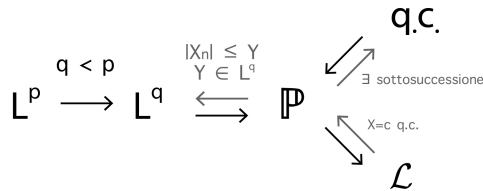


Figura 16.3: schema riassuntivo dei teoremi sulle relazioni tra i teoremi della convergenza

- La convergenza in probabilità ha una sottosuccessione che converge quasi certamente per il teorema 16.10;
- Dati  $p$  e  $q$  con  $p \geq q$ , la convergenza in  $L^p$  implica la convergenza in  $L^q$  per la proposizione 16.9;
- Dati  $|X| \in L^p$ ,  $|X_n| \leq Y$  q.c.  $\forall n$  e  $Y \in L^p$ , ovvero rispettano una convergenza dominata e sono in  $L^p$ , allora è possibile, data la convergenza in probabilità, ottenere quella in  $L^p$  come enunciato dal teorema 16.11;
- Se  $X_n$  converge ad una costante allora la convergenza in legge implica quella in probabilità, per il teorema 16.22.

---

## 17 Teoremi limite

I due *teoremi limite* sono tra i risultati più potenti dello studio della probabilità. Partendo da ipotesi minime, contribuiscono a mettere le basi per lo studio della statistica; essi poggiano su tutta la teoria affrontata finora, in particolare sulla convergenza di variabili aleatorie, ma avranno delle dimostrazioni sorprendentemente semplici a questo punto della trattazione.

La *legge dei grandi numeri* permette l'approssimazione della media di una variabile aleatoria attraverso esperimenti ripetuti; il *teorema centrale del limite* cementa definitivamente l'importanza della legge gaussiana nello studio di variabili aleatorie con legge *qualsiasi*. L'introduzione di questi teoremi consentirà l'analisi del *comportamento asintotico* di alcune successioni (in particolare della loro *velocità di convergenza*) e la *stima dei parametri* di una popolazione, tecnica frequentemente sfruttata in statistica.

### 17.1 Legge dei grandi numeri

**Teorema 17.1: legge forte dei grandi numeri (LGN)** (*J-P Th 20.1, 20.2*)

Siano  $X_n \in L^1$  una successione di variabili aleatorie iid e  $\mu \in \mathbb{R}$ . Allora:

$$\mathbb{E}[X_n] = \mu \iff \bar{X}_n \xrightarrow{\text{qc}} \mu$$

Inoltre in questo caso vale anche  $\bar{X}_n \xrightarrow{L^1} \mu$ .

Questa legge viene definita “forte” perché dà un risultato di convergenza quasi certa, che è ovviamente un risultato più potente rispetto ad altri tipi di convergenza. Si noti che la parte ( $\implies$ ) di questo teorema ha ipotesi sul codominio e dà una tesi sul dominio.<sup>51</sup>

Inoltre, nel caso in cui le VA della successione siano continue, si può notare che la loro media, per definizione un integrale e quindi una somma *continua*, possa anche essere ottenuta come un limite *numerabile*.

Abbiamo già visto una delle applicazioni (inconsapevoli) di questo enunciato all'inizio del testo, a pagina 15, per assegnare una probabilità in modo frequentista.

*Esempio.* Se  $X_n \sim Be(p)$  iid, allora  $\mathbb{E}[X_n] = \mathbb{P}(X_n = 1) = p \iff \bar{X}_n \xrightarrow{\text{qc}} p$ .

*Esempio.* Se  $X_n \sim P^X$  iid,  $\mathbb{P}(X_n \in B) = P^X(B) = p \iff \frac{\sum_{k=1}^n \mathbb{1}_B(X_k)}{n} \xrightarrow{\text{qc}} p$ .

*Esempio.* Definiamo  $\Omega = \{0, 1\}^{\mathbb{N}}$ ,  $\mathcal{A} = \sigma(X_n = 1 : n \in \mathbb{N})$  e  $X_n(\omega) = \omega_n$ . Allora si ha  $\bar{X}_n(\omega) = \bar{\omega}_n \xrightarrow{\text{qc}} p$ . In questo teorema vengono definiti solo oggetti che

---

<sup>51</sup>“Teoremi come questi non li avete visti in analisi”

dipendono dalla legge: in quest'ultimo caso, alterando il parametro  $p$  la probabilità si sposta, senza modificarsi, andandosi a concentrare su certi  $\omega_n$  che fanno tendere la media campionaria a  $p$ . Per questo il teorema funziona per ogni valore ammissibile di  $p$ .

### Proposizione 17.2

Siano  $X_n$  VA iid in  $L^2$ . Allora:

$$\mu = \mathbb{E}[X_n] \text{ e } \sigma^2 = \text{Var}(X_n) \implies S_n^2 \xrightarrow{\text{qc}} \sigma^2$$

### Dimostrazione

Possiamo riscrivere  $S_n^2$  a partire dalla sua definizione:

$$S_n^2 = \frac{1}{n-1} \sum_{k=1}^n (X_k - \bar{X}_n)^2 = \frac{n}{n-1} \left[ \frac{1}{n} \cdot \sum_{k=1}^n X_k^2 - \bar{X}_n^2 \right]$$

Analizziamo ora la convergenza di ciascun elemento:

$$\frac{n}{n-1} \xrightarrow{n} 1, \quad \frac{1}{n} \sum_{k=1}^n X_k^2 \xrightarrow{\text{qc}} \mathbb{E}[X^2] = \mu^2 + \sigma^2, \quad \bar{X}_n^2 \xrightarrow{\text{qc}} \mu^2$$

Per le proprietà della convergenza quasi certa,  $S_n^2 \xrightarrow{\text{qc}} 1(\mu^2 + \sigma^2 - \mu^2) = \sigma^2$ .  $\square$

### Dimostrazione della LGN

Dimostreremo solo ( $\implies$ ) nel caso  $L^2$ .

Siano  $X_n \in L^2$  iid e chiamiamo  $\mu = \mathbb{E}[X_n]$  e  $\sigma^2 = \text{Var}(X_n)$  per semplicità di notazione.

1. Dimostriamo che  $\bar{X}_n \xrightarrow{L^2} \mu$ . Per linearità del valore atteso possiamo affermare che  $\mathbb{E}[\bar{X}_n] = \mu$  e che:

$$\text{Var}(\bar{X}_n) = \mathbb{E}[(\bar{X}_n - \mu)^2] = \frac{\sigma^2}{n} \xrightarrow{n} 0$$

Da ciò discende immediatamente l'affermazione di sopra.

Possiamo quindi imporre  $\mu = 0$  senza perdita di generalità definendo la VA  $Y_n := X_n - \mu$ , infatti:

$$\mathbb{E}[X_n] = \mu, \quad \bar{X}_n \xrightarrow{\text{qc}} \mu \iff \mathbb{E}[Y_n] = 0, \quad Y_n \xrightarrow{\text{qc}} 0$$

2. Mostriamo ora che la sottosuccessione  $\bar{X}_{n^2}$ , ovvero la media campionaria degli  $X_k$  tali che  $k = n^2$  per un qualche  $n$ , converge quasi certamente a 0.

$$\begin{aligned} \mathbb{E} \left[ \sum_{n=1}^{+\infty} (\bar{X}_{n^2})^2 \right] &= \sum_{n=1}^{+\infty} \mathbb{E} [(\bar{X}_{n^2})^2] \quad (\text{scambio di limiti 6.26 con VA positive}) \\ &= \sum_{n=1}^{+\infty} \text{Var}(\bar{X}_{n^2}) \quad (\text{perché } \mathbb{E}[X_n] = 0) \\ &= \sum_{n=1}^{+\infty} \frac{\sigma^2}{n^2} < +\infty \quad (\text{la serie di } \frac{1}{n^2} \text{ converge}) \end{aligned}$$

Condizione necessaria per la convergenza della serie è che il limite del termine generale della successione valga zero:

$$\sum_{n=1}^{+\infty} (\bar{X}_{n^2})^2 < +\infty \text{ qc} \implies (\bar{X}_{n^2})^2 \xrightarrow{n} 0 \text{ qc} \iff \bar{X}_{n^2} \xrightarrow{\text{qc}} 0$$

3. Infine proviamo che  $\bar{X}_n \xrightarrow{\text{qc}} 0$ .

Sia  $p(n) \in \mathbb{N}$  il numero<sup>52</sup> tale che:

$$p(n)^2 \leq n < (p(n) + 1)^2, \quad \text{ovvero} \quad \frac{(\sqrt{n} - 1)^2}{n} < \frac{p(n)^2}{n} \leq 1$$

Per il teorema dei due carabinieri, vale che  $\frac{p(n)^2}{n} \xrightarrow{n} 1$ .

Sfruttiamo  $p(n)$  per definire una nuova VA:

$$\bar{X}_n - \frac{p(n)^2}{n} \bar{X}_{p(n)^2} = \frac{1}{n} \sum_{k=p(n)^2+1}^n X_k$$

Passando alle varianze:

$$\text{Var} \left( \bar{X}_n - \frac{p(n)^2}{n} \bar{X}_{p(n)^2} \right) = \text{Var} \left( \frac{1}{n} \sum_{k=p(n)^2+1}^n X_k \right) = \frac{1}{n^2} \sigma^2 (n - p(n)^2)$$

Con una catena di minorazioni, troviamo:

$$\frac{1}{n^2} \sigma^2 (n - p(n)^2) \leq \frac{\sigma^2}{n^2} (2p(n) + 1) \leq \frac{\sigma^2}{n^2} (2\sqrt{n} + 1) \leq \frac{3\sigma^2}{n^{\frac{3}{2}}}$$

---

<sup>52</sup>Formalmente si può definire questo numero come  $p(n) := \lfloor \sqrt{n} \rfloor$ , che è un valore ben definito per ogni  $n$ . Tale definizione non è comunque rilevante ai fini della dimostrazione.

Sviluppando la varianza tramite la definizione e sfruttando la linearità del valore atteso, abbiamo:

$$\begin{aligned} \mathbb{E} \left[ \sum_{n=1}^{+\infty} \left( \bar{X}_n - \frac{p(n)^2}{n} \bar{X}_{p(n)^2} \right)^2 \right] &= \sum_{n=1}^{+\infty} \mathbb{E} \left[ \left( \bar{X}_n - \frac{p(n)^2}{n} \bar{X}_{p(n)^2} \right)^2 \right] \\ &\leq \sum_{n=1}^{+\infty} \frac{3\sigma^2}{n^{\frac{3}{2}}} < +\infty \end{aligned}$$

Si ottiene così una serie convergente; quindi, come abbiamo già visto, il termine generale tende a 0. Rimuoviamo allora il quadrato:

$$\left( \bar{X}_n - \frac{p(n)^2}{n} \bar{X}_{p(n)^2} \right)^2 \xrightarrow{qc} 0 \implies \bar{X}_n - \frac{p(n)^2}{n} \bar{X}_{p(n)^2} \xrightarrow{qc} 0$$

Sfruttando i risultati precedenti vediamo che:

$$\bar{X}_n = \underbrace{\bar{X}_n - \frac{p(n)^2}{n} \bar{X}_{p(n)^2}}_{\xrightarrow{qc} 0} + \underbrace{\frac{p(n)^2}{n}}_{\rightarrow 1} \cdot \underbrace{\bar{X}_{p(n)^2}}_{\xrightarrow{qc} 0} \xrightarrow{qc} 0 \quad \square$$

### 17.1.1 Applicazioni notevoli

*Metodo Monte Carlo.* Vogliamo stimare numericamente il seguente integrale:

$$c = \int_0^1 h(x) dx, \quad \text{con } h \in L^1$$

Per farlo definiamo  $X_n \sim U([0, 1])$  iid. Notiamo che  $h(X_n) \in L^1(\mathbb{P})$  iid  $\iff h \in L^1(P^{X_n})$ . Allora:

$$\frac{1}{n} \sum_{k=1}^n h(X_k) \xrightarrow{qc} \mathbb{E}[h(X_1)] = \int_0^1 h(s) ds \implies \frac{1}{n} \sum_{k=1}^n h(X_k) \simeq c \quad (\text{se } n \text{ è grande})$$

In generale, il metodo Monte Carlo è una tecnica che calcola integrali approssimandone il valore con la media campionaria, anche se la funzione integranda non ammette integrale elementare.

*Statistica.* Tramite la consistenza è possibile stimare  $\mu$  e  $\sigma^2$ :

$$\hat{\mu} = \bar{X}_n \xrightarrow{n} \mu \quad \text{e} \quad \hat{\sigma}^2 = S_n^2 = \frac{1}{n-1} \sum_{k=1}^n (X_k - \bar{X}_n)^2$$

## 17.2 Teorema centrale del limite

**Teorema 17.3: teorema centrale del limite (TCL)** (*J-P Th 21.1*)

Siano  $X_n$  iid in  $L^2$ , con  $\mu = \mathbb{E}[X_n]$  e  $\sigma^2 = \text{Var}(X_n) > 0$ . Allora:

$$\frac{\bar{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1)$$

Se  $\sigma^2 = 0$  il teorema perde di significato, ma questo è comunque un caso di banale trattazione.

Nella pratica statistica questo teorema risolve il problema dell'impossibilità di fare infiniti esperimenti: considerando un campione sufficientemente numeroso, la distribuzione può essere approssimata con una gaussiana.<sup>53</sup>

**NB:** Dati  $X_n \in L^2$  iid, abbiamo visto che, per la legge dei grandi numeri,  $\bar{X}_n \xrightarrow{qc} \mu$  e che la varianza della media campionaria si riduce fino a diventare nulla:

$$\text{Var}(\bar{X}_n) = \frac{\sigma^2}{n} \longrightarrow 0$$

Dunque la distribuzione si stringe e converge in legge a  $\delta_\mu$ , a prescindere dalla legge originaria delle VA. Normalizzando si impedisce che essa degeneri in una delta, permettendo l'osservazione della distribuzione per  $n$  che tende ad infinito:

$$\frac{\bar{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} \implies \mathbb{E} \left[ \frac{\bar{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} \right] = 0, \quad \text{Var} \left( \frac{\bar{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} \right) = 1$$

Senza il teorema, il limite del rapporto nella tesi sarebbe un'irrisolvibile forma indeterminata  $\frac{0}{0}$ .

Le ipotesi del TCL (o delle sue varianti) possono, in alcuni casi, non richiedere che le VA siano iid, e ciò spiega perché le distribuzioni gaussiane siano così comuni: essendo molte VA somme di diversi componenti indipendenti, gli effetti del TCL si ripercuotono su tali casi. Un esempio può essere l'altezza, dovuta a molti fattori diversi che la influenzano in modo non identicamente distribuito; infatti, preso un numero sufficientemente elevato di individui, essa risulta gaussiana.

### Dimostrazione del TCL

Questa dimostrazione fa uso del logaritmo complesso, argomento tipicamente sviscerato nel corso di analisi 3. Per agevolare la comprensione si possono trovare alcuni cenni alle sue proprietà nell'appendice A, a pagina 257.

<sup>53</sup>“La statistica, di colpo, serve a qualcosa”

Come per la LGN, è possibile porre  $\mu = 0$  senza perdita di generalità mediante la traslazione  $Y_n := X_n - \mu$ . Infatti questo implica che  $Y_n$  iid,  $Y_n \in L^2$ ,  $\mathbb{E}[Y_n] = 0$ ,  $\bar{Y}_n = \bar{X}_n - \mu$ ,  $Var(Y_n) = \sigma^2$ , e che pertanto:

$$\frac{\bar{Y}}{\frac{\sigma}{\sqrt{n}}} = \frac{\bar{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1)$$

Avendo una convergenza in legge è opportuno usare la funzione caratteristica: sia dunque  $\varphi$  la funzione caratteristica di  $Y_n$ . Si nota che  $\varphi(u) = \mathbb{E}[e^{iuY_n}] \implies \varphi \in C^2$  perché  $Y_n \in L^2$  e perché  $\varphi'(0) = 0$  e  $\varphi''(0) = -\sigma^2$ , valori che si possono facilmente verificare. Pertanto, lo sviluppo in serie di Taylor arrestato al second'ordine di  $\varphi$  è:

$$\varphi(u) = 1 - \frac{1}{2}\sigma^2 u^2 + o(u^2), \quad \text{con } o(u^2) \in \mathbb{C}$$

Si può ora scrivere la  $\varphi$  del rapporto cercato e modificarla nel modo che segue:

$$\begin{aligned} \varphi_{\frac{\bar{Y}_n}{\frac{\sigma}{\sqrt{n}}}}(u) &= \varphi_{\frac{\sum_{k=1}^n Y_k}{\sigma\sqrt{n}}}(u) = \varphi_{\sum_{k=1}^n Y_k}\left(\frac{u}{\sigma\sqrt{n}}\right) && \text{(per le proprietà di } \varphi) \\ &= \left(\varphi\left(\frac{u}{\sigma\sqrt{n}}\right)\right)^n && \text{(le VA sono iid)} \\ &= \exp\left\{\log\left(\varphi\left(\frac{u}{\sigma\sqrt{n}}\right)\right)^n\right\} && \text{(introducendo un exp)} \\ &= \exp\left\{n \log\left(\varphi\left(\frac{u}{\sigma\sqrt{n}}\right)\right)\right\} && \text{(proprietà dei logaritmi)} \\ &= \exp\left\{n \log\left(1 - \frac{1}{2}\sigma^2 \frac{u^2}{n\sigma^2} + o\left(\frac{1}{n}\right)\right)\right\} && \text{(sviluppando in serie } \varphi) \\ &= \exp\left\{n\left(-\frac{1}{2}\frac{\sigma^2 u^2}{n\sigma^2} + o\left(\frac{1}{n}\right)\right)\right\} && \text{(sviluppando in serie il log)} \\ &= \exp\left\{\left(-\frac{1}{2}u^2 + o(1)\right)\right\} \xrightarrow{n \rightarrow +\infty} e^{-\frac{u^2}{2}} = \varphi_{\mathcal{N}(0,1)}(u) \quad \forall u \in \mathbb{R} \end{aligned}$$

Questo conclude la dimostrazione. Si noti ancora come essa regga per qualsiasi distribuzione, dipendendo solo da media e varianza della legge.  $\square$

### 17.3 Comportamento asintotico di alcune distribuzioni

L'introduzione del TCL permette uno studio dell'andamento delle distribuzioni di una successione di VA  $X_1, \dots, X_n$  quando  $n$  è sufficientemente grande. Forniamo di seguito qualche esempio notevole.



### 17.3.1 Chi-quadro

Data una successione di VA  $X_n \sim \chi^2(n)$ , si ha  $X_n \sim \sum_{k=1}^n Z_k^2$ , dove  $Z_k \sim \mathcal{N}(0, 1)$  iid. Notiamo che anche  $Z_k^2 \sim \chi^2(1)$  iid.

I momenti della  $Z_k^2$  sono  $\mathbb{E}[Z_k^2] = 1$  e  $Var(Z_k^2) = 2 \ \forall k$ , quindi possiamo procedere con la normalizzazione, come richiesto dal TCL:

$$\frac{X_n - n}{\sqrt{2n}} \sim \frac{\sum_{k=1}^n Z_k^2 - n}{\sqrt{2n}} = \frac{(\overline{Z_n^2}) - 1}{\sqrt{\frac{2}{n}}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1)$$

Per  $n$  grande, si ha:

$$\frac{X_n - n}{\sqrt{2n}} \approx \mathcal{N}(0, 1) \implies X_n \sim \chi^2(n) \approx \mathcal{N}(n, 2n)$$

**NB:** Si nota che, come sempre parlando di leggi, non è detto che  $X_n$  sia effettivamente una somma di normali standard, ma semplicemente ha la stessa distribuzione di una somma di normali standard.

### 17.3.2 $t$ di Student

Data una successione di VA  $X_n \sim t(n)$ , si ha  $X_n \sim \frac{Z}{\sqrt{Q/n}}$ , dove  $Z \sim \mathcal{N}(0, 1)$ ,  $Q \sim \chi^2$  e  $Z \perp\!\!\!\perp Q$ .

Grazie alla LGN possiamo mostrare che:

$$\frac{Q}{n} \sim \frac{\sum_{k=1}^n Z_k^2}{n} \xrightarrow{qc} 1 \implies \frac{Q}{n} \xrightarrow{\mathcal{L}} 1 \implies X_n \sim \frac{Z}{\sqrt{\frac{Q}{n}}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1)$$

Per l'ultima implicazione abbiamo invocato Slutsky<sup>54</sup>, affermando che  $Q/n$  tende ad una costante. In altri termini, si è ottenuta la seguente relazione tra leggi:

$$t(n) \xrightarrow{\text{deb}} \mathcal{N}(0, 1)$$

### 17.3.3 Binomiale

Data una successione di  $X_n \sim Bi(n, p)$ , si ha  $X_n \sim \sum_{k=1}^n Y_k$ , dove  $Y_k \sim Be(p)$  iid.

Ne consegue che  $X_n \sim n\bar{Y}_n$ . Per il TCL si scrive che:

$$\frac{\bar{Y}_n - p}{\sqrt{np(1-p)/n}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1) \implies \frac{X_n - np}{\sqrt{np(1-p)}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1)$$

<sup>54</sup>Che è il nuovo Fubini-Tonelli.

Ovvero, dato  $n$  grande, la generica relazione tra le due leggi è la seguente:

$$Bi(n, p) \approx \mathcal{N}(np, np(1-p))$$

## 17.4 Velocità di convergenza

Data una successione di VAR  $X_k$  iid in  $L^2$ , indicata con  $\bar{X}_n$  la loro media campionaria per  $n \rightarrow +\infty$  si ha:

- $\bar{X}_n \xrightarrow{qc} \mu$  per la legge dei grandi numeri;
- $\sqrt{n}(\bar{X}_n - \mu) \xrightarrow{qc} \mathcal{N}(0, \sigma^2)$  per il teorema centrale del limite.

L'ultimo criterio viene utilizzato per misurare la *velocità di convergenza* del limite.

Riprendiamo una definizione dall'analisi per agevolare la spiegazione:

### Definizione 17.4

La *velocità di convergenza* della funzione  $f(x)$  è l'esponente  $\alpha$  tale per cui:

$$\lim_{x \rightarrow +\infty} x^\alpha f(x) = L \in \mathbb{R} \setminus \{0\}$$

*Esempio.* Data la funzione  $\cos\left(\frac{1}{n}\right) \xrightarrow{n} 1$ , l'espressione  $n^\alpha \left(\cos\left(\frac{1}{n}\right) - 1\right)$  tende ad un limite finito diverso da zero per  $\alpha = 2$ .

### Definizione 17.5

Sia  $X_n$  una successione di VAR,  $a \in \mathbb{R}$  una costante, e  $v_n$  una successione reale con  $v_n \rightarrow +\infty$  per  $n \rightarrow +\infty$ .

Se  $\exists T$  VAR, con  $T \not\sim \delta_0$ , tale che  $v_n(X_n - a) \xrightarrow{\mathcal{L}} T$ , allora si dice che  $X_n$  converge ad  $a$  con **velocità di convergenza**  $v_n$ .

**NB:** La distribuzione  $\delta_0$  prevede tutta la probabilità concentrata nel punto 0:  $T \not\sim \delta_0 \iff \mathbb{P}(T = 0) < 1$ . La distribuzione limite  $\delta_0$  è dunque l'analogo del limite nullo nel caso presentato nella definizione 17.4.

**NB:**  $X_n \xrightarrow{\mathcal{L}} a \implies X_n \xrightarrow{\mathbb{P}} a$ . Infatti  $X_n - a = \underbrace{v_n(X_n - a)}_{\rightarrow T} \underbrace{1/v_n}_{\rightarrow 0} \xrightarrow{n} 0$  per il teorema di Slutsky.

*Esempio.* Sia  $X_k$  una successione di VA iid in  $L^2$  con  $\sigma_k^2 > 0 \forall k$ ; allora la media campionaria  $\bar{X}_n$  converge a  $\mu$  con velocità  $v_n = \sqrt{n}$ , a prescindere dalla legge delle

$X_k$ . Infatti, applicando il TCL, si ha che  $\sqrt{n}(\bar{X}_n - \mu) \longrightarrow \mathcal{N}(0, \sigma^2)$ .

Questo risultato, data la sua generalità, viene spesso utilizzato negli esercizi per indovinare la velocità di convergenza di successioni più complesse.

**NB:**  $v_n$  non è univocamente determinata. Infatti, data una successione  $X_n \rightarrow \mu$  con  $v_n = \sqrt{n}$ , questa converge a un valore reale non nullo anche, per esempio, per  $v'_n = \frac{n}{n+1}\sqrt{n}$  grazie al teorema di Slutsky.

*Esempio.* Si consideri una successione di VA  $X_k \sim U([0, \lambda])$  iid con  $\lambda > 0$  fissato. Sia poi  $X_{(1)} = \min\{X_1, \dots, X_n\} = W_n$ .<sup>55</sup> Dimostriamo che  $W_n$  tende in legge a 0.

Si utilizza, come spesso succede quando sono coinvolti dei min, la funzione di ripartizione di  $W_n$ : essa è  $F_{W_n}(t) = \mathbb{P}(W_n \leq t) = 1 - \mathbb{P}(W_n > t)$ . Tuttavia, essendo  $W_n$  un minimo di VA iid,  $F_{W_n}(t) = 1 - \mathbb{P}(X_1 > t, \dots, X_n > t) = 1 - \mathbb{P}(X_1 > t)^n$ . Distinguiamo quindi i casi per la variabile  $t$ :

$$F_{W_n}(t) = \begin{cases} 0 & \text{per } t \leq 0 \\ 1 - \left(\frac{\lambda-t}{\lambda}\right)^n & \text{per } 0 < t \leq \lambda \\ 1 & \text{per } t > \lambda \end{cases} \xrightarrow{n \rightarrow +\infty} \begin{cases} 0 & \text{per } t \leq 0 \\ 1 & \text{per } 0 < t \leq \lambda \\ 1 & \text{per } t > \lambda \end{cases} = \mathbf{1}_{(0, +\infty)}(t)$$

Si può quindi affermare che  $F_{W_n}(t) \xrightarrow{qo} \mathbf{1}_{[0, +\infty)}(t) \quad \forall t$  (per la precisione, ovunque tranne che in  $t = 0$ ). Ma questa è la funzione di ripartizione della delta di Dirac  $\delta_0$ , quindi  $W_n \xrightarrow{\mathcal{L}} 0$ .

In questo caso la velocità di convergenza è  $v_n = n$ : infatti, si può dimostrare che  $nW_n = T_n \xrightarrow{\mathcal{L}} \varepsilon(\lambda)$ .

### 17.4.1 Asintotica normalità

#### Definizione 17.6

Una successione di VA è detta **asintoticamente normale** di parametri  $a$  e  $\frac{q}{v_n^2}$  (con  $q > 0$  e  $v_n \rightarrow +\infty$ ) se:

$$v_n(X_n - a) \xrightarrow{\mathcal{L}} \mathcal{N}(0, q), \quad \text{e si scriverà } X_n \sim AN\left(a, \frac{q}{v_n^2}\right)$$

*Esempio.* Riscriviamo il TCL con la notazione appena introdotta:

$$X_k \text{ iid in } L^2, \text{ con } \sigma^2 > 0 \implies \bar{X}_n \sim AN\left(\mu, \frac{\sigma^2}{n}\right)$$

<sup>55</sup>Il pedice (1) è una notazione alternativa per il minimo frequentemente usata in statistica. Più in generale, dato un insieme di valori  $\{X_1, \dots, X_n\}$ ,  $X_{(i)}$  è l' $i$ -esimo valore più piccolo dell'insieme (e quindi  $X_{(n)}$  è il valore massimo tra essi).

**NB:** Nonostante  $X_n \rightarrow a$ , non è necessariamente vero che  $\mathbb{E}[X_n] \rightarrow a$  o che  $\text{Var}(X_n) \sim \frac{q}{v_n^2}$ .

*Controesempio.* Siano  $X_n, Y_n$  successioni di VAR tali che  $X_k \perp\!\!\!\perp Y_k \forall k$ . Siano inoltre  $p_n, \mu_n, v_n$  e  $\sigma_n$  successioni numeriche tali che  $v_n \rightarrow +\infty$ , che  $Y_n \sim Be(p_n)$  con  $p_n \xrightarrow{n} 1$ , e che le  $X_n$  abbiano, per ogni  $n$ , la seguente legge condizionata:

$$X_n | Y_n = k \sim \mathcal{N} \left( ak + \mu_n(1 - k), k \frac{q}{v_n^2} + (1 - k) \frac{\sigma_n^2}{v_n^2} \right)$$

Dimostriamo che  $X_n$  è asintoticamente normale, ovvero che  $Z_n = v_n(X_n - a) \xrightarrow{\mathcal{L}} \mathcal{N}(0, q)$ . Mediante il teorema 11.10 sulle trasformazioni affini delle funzioni caratteristiche si ottiene:

$$\varphi_{Z_n}(u) = e^{-iav_n u} \varphi_{X_n}(v_n u)$$

Ricordando che  $\varphi_{X_n}$  è un valore atteso e che pertanto vale la formula di pagina 179, si può continuare scrivendo:

$$\begin{aligned} e^{-iav_n u} \varphi_{X_n}(v_n u) &= e^{-iav_n u} \mathbb{E} [ \varphi_{X_n}(v_n u) ] \\ &= e^{-iav_n u} \left[ p \mathbb{E} [ \varphi_{X_n}(v_n u) | Y_n = 1 ] + (1 - p) \mathbb{E} [ \varphi_{X_n}(v_n u) | Y_n = 0 ] \right] \\ &= e^{-iav_n u} \left[ p \varphi_{\mathcal{N}\left(a, \frac{q}{v_n^2}\right)}(v_n u) + (1 - p) \varphi_{\mathcal{N}\left(\mu, \frac{\sigma_n^2}{v_n^2}\right)}(v_n u) \right] \end{aligned}$$

Infatti  $k = 0$  o  $k = 1$  cancellano l'uno o l'altro dei due addendi nei parametri della legge di  $X_n | Y_n = k$ . Ora, la funzione caratteristica  $\varphi_{\mathcal{N}\left(\mu, \frac{\sigma_n^2}{v_n^2}\right)}(v_n u)$  è limitata  $\forall u$  e  $1 - p \rightarrow 0$ : l'intero secondo addendo è un  $o(1)$ . Sostituendo la formula per la  $\varphi_{\mathcal{N}}$  nell'addendo rimasto e ricordando che  $v_n \rightarrow +\infty$ , risulta che:

$$\begin{aligned} \varphi_{Z_n}(u) &= e^{-iav_n u} \left[ p e^{iav_n u - \frac{1}{2} \frac{q}{v_n^2} v_n^2 u^2} + o(1) \right] \\ &= p e^{-\frac{1}{2} q u^2} + e^{-iav_n u} o(1) \xrightarrow{n} e^{-\frac{1}{2} q u^2} \end{aligned}$$

La funzione ottenuta è la funzione caratteristica di una normale con media nulla e varianza  $q$ . Dunque effettivamente  $X_n \sim \mathcal{AN}\left(a, \frac{q}{v_n^2}\right)$ . Infine, calcoliamo media e varianza di  $X_n$  usando le definizioni di attesa e varianza condizionata:

$$\begin{aligned} \mathbb{E}[X_n] &= \mathbb{E}[\mathbb{E}[X_n | Y_n]] \\ &= \mathbb{E}[aY_n + \mu_n(1 - Y_n)] = a p_n + (1 - p_n)\mu_n \end{aligned}$$

$$\text{Var}(X_n) = \text{Var}(\mathbb{E}[X_n | Y_n]) + \mathbb{E}[\text{Var}(X_n | Y_n)] \quad (\text{proprietà 15.10})$$

Quindi per le proprietà di varianza e valore atteso si ha:

$$\begin{aligned} \text{Var}(X_n) &= \text{Var}(aY_n + \mu_n(1 - Y_n)) + \mathbb{E} \left[ Y_n \frac{q}{v_n^2} + (1 - Y_n) \frac{\sigma_n^2}{v_n^2} \right] \\ &= (a - \mu_n)^2 p(1 - p) + p \frac{q}{v_n^2} + (1 - p) \frac{\sigma_n^2}{v_n^2} \end{aligned}$$

Come si può osservare, media e varianza non sono sempre uguali, rispettivamente, ad  $a$  e a  $\frac{q}{v_n^2}$ ; anzi, a seconda delle successioni scelte nelle ipotesi si può farle convergere a qualsiasi valore desiderato o anche divergere a  $+\infty$ .

Data una funzione  $h : \mathbb{R} \rightarrow \mathbb{R}$  continua e  $X_n$  che converge in legge, probabilità o quasi certa<sup>56</sup>  $X_n \rightarrow a$  allora è noto che  $Y_n = h(X_n) \implies Y_n \rightarrow h(a)$ , ma per scoprire la sua velocità di convergenza e se è asintoticamente normale introduciamo il metodo delta 1, molto utile negli esercizi.

### Teorema 17.7: metodo delta 1<sup>57</sup>

Siano:

- $X_n$  successione di VA reali a valori in un boreliano  $B \subseteq \mathbb{R}$ ;
- $a \in \overset{\circ}{B}$ , ovvero punto interno a  $B$ ;
- $v_n \rightarrow +\infty$ , velocità di convergenza;
- $T$  VAR.

Se  $v_n(X_n - a) \xrightarrow{\mathcal{L}} T$  e  $h : B \rightarrow \mathbb{R}$  è una funzione misurabile su  $B$  e differenziabile in  $a$ , allora:

$$v_n(h(X_n) - h(a)) \xrightarrow{\mathcal{L}} h'(a)T$$

Si noti che la funzione  $h$  non deve dipendere da  $n$ , altrimenti  $h$  non sarebbe una funzione, bensì una successione di funzioni  $h_n$  e il metodo delta 1 non sarebbe valido.

### Dimostrazione

- *Caso generale.* Si scriva lo sviluppo di Taylor-Peano di  $h(x)$  centrato in  $x = a$ :

$$h(x) = h(a) + h'(a)(x - a) + o(x - a) \quad \forall x \in B$$

Si definisca ora il resto  $r$  come  $r(x - a) := h(x) - h(a) - h'(a)(x - a) = o(x - a)$ , con  $x \in B$ , che è tale che  $\frac{r(x - a)}{x - a} \rightarrow 0$  per  $x \rightarrow a$ . Definiamo quindi anche la

<sup>56</sup>“Nel modo che preferite”, disse Gregoratti.

<sup>57</sup>Nel caso ve lo steste chiedendo: sì, esiste un metodo delta 2, ma non fa parte del programma di questo corso.

funzione  $g : \mathbb{R} \rightarrow \mathbb{R}$ :

$$g(y) := \begin{cases} \frac{r(y)}{y} & \text{con } y + a \in B, y \neq 0 \\ 0 & \text{con } y + a \notin B, y \neq 0 \end{cases}$$

Essa è continua in  $y = 0$  e misurabile. Dunque  $h(x) - h(a) = h'(a)(x - a) + (x - a)g(x - a) \forall x \in B$ . Applichiamo la velocità di convergenza:

$$\begin{aligned} v_n(h(X_n) - h(a)) &= v_n(h'(a)(X_n - a) + (X_n - a)g(X_n - a)) \\ &= v_n(X_n - a)h'(a) + v_n(X_n - a)g(X_n - a) \end{aligned}$$

I termini  $v_n(X_n - a)$  convergono in legge a  $T$ ; inoltre, poiché  $X_n - a \xrightarrow{\mathcal{L}} 0$ , per la continuità di  $g$  anche  $g(X_n - a) \xrightarrow{\mathcal{L}} g(0) = 0$ . Per il teorema di Slutsky la somma converge dunque in legge a  $T \cdot h'(a) + T \cdot 0 = h'(a)T$ .

- *Caso*  $B = \mathbb{R}$ .

$$g(s) := \begin{cases} \frac{h(s) - h(a)}{s - a} - h'(a) & \text{per } s \neq a \\ 0 & \text{per } s = a \end{cases}$$

Mediante uno sviluppo di Taylor al prim'ordine si può scrivere che:

$$v_n(h(X_n) - h(a)) = v_n(h'(a)(X_n - a) + o(X_n - a))$$

Portando  $v_n h'(a)(X_n - a)$  al membro sinistro:

$$\begin{aligned} v_n o(X_n - a) &= v_n(h(X_n) - h(a) - h'(a)(X_n - a)) \\ &= v_n(X_n - a) \left( \frac{h(X_n) - h(a)}{X_n - a} - h'(a) \right) \\ &= \underbrace{v_n(X_n - a)}_{\xrightarrow{\mathcal{L}} T} \underbrace{g(X_n)}_{\xrightarrow{\mathcal{L}} g(a)} \xrightarrow{\mathcal{L}} 0T = 0 \quad (\text{per Slutsky}) \end{aligned}$$

Pertanto, ricordando la definizione di  $g$ :

$$\begin{aligned} v_n(X_n - a) \left( \frac{h(X_n) - h(a)}{X_n - a} - h'(a) \right) &\xrightarrow{\mathcal{L}} 0 \\ \implies v_n(h(X_n) - h(a)) - h'(a)v_n(X_n - a) &\xrightarrow{\mathcal{L}} 0 \\ \implies v_n(h(X_n) - h(a)) &\xrightarrow{\mathcal{L}} h'(a)T \quad \square \end{aligned}$$

### Corollario 17.8

Se  $T \neq \delta_0$  e  $h'(a) \neq 0$ ,  $v_n$  è la velocità di convergenza sia di  $X_n$  che di  $h(X_n)$ ; in altri termini, *quasi* tutte le trasformazioni *non* modificano la velocità di convergenza.

**Corollario 17.9: metodo delta 1, caso gaussiano**

Siano  $X_n : \Omega \rightarrow B$  tali che  $X_n \sim AN\left(a, \frac{q}{v_n^2}\right)$  e  $h : B \rightarrow \mathbb{R}$  funzione misurabile tale che  $h'(a) \neq 0$ . Allora:

$$Y_n = h(X_n) \sim AN\left(h(a), (h'(a))^2 \frac{q}{v_n^2}\right)$$

Vediamo qualche esempio di applicazione del corollario enunciato poc'anzi.

*Esempio.* Siano  $X_n \sim \mathcal{E}(\lambda)$  iid con  $\lambda > 0$ .

È noto che  $\bar{X}_n \sim AN\left(\frac{1}{\lambda}, \frac{1}{n\lambda^2}\right)$  per il TCL, e che  $\bar{X}_n \xrightarrow{\text{qc}} \frac{1}{\lambda}$  per la LGN. Da qui si nota, peraltro, come l'asintotica normalità non sia in senso stretto una convergenza di  $\bar{X}_n$ , bensì di una *funzione* di  $\bar{X}_n$ , per la precisione di  $(\sqrt{n}\bar{X}_n - a)$ : in caso contrario sarebbe violata l'unicità del limite.

Si supponga ora che il valore di  $\lambda$  sia incognito, e che lo si voglia stimare con la successione  $Y_n := \frac{1}{\bar{X}_n}$ . Dunque  $Y_n = h(\bar{X}_n)$ , con  $h : B = (0, +\infty) \rightarrow \mathbb{R}$ ,  $h(t) = \frac{1}{t}$ , che è tale che  $h'(t) = -\frac{1}{t^2} \neq 0 \forall t \in (0, +\infty)$ . Si osservi che in questo caso la restrizione del dominio di  $h$  a un boreliano è stata fondamentale per garantire regolarità  $C^1$  in ogni punto della funzione, che altrimenti non sarebbe stata derivabile in  $t = 0$ .

Per le proprietà di convergenza quasi certa e funzioni continue,  $Y_n \xrightarrow{\text{qc}} \lambda = a$  e quindi c'è anche convergenza in legge a  $\lambda$ . È dunque possibile riscrivere il limite con la notazione dell'asintotica normalità:

$$Y_n \sim AN\left(\lambda, (h'(a)) \frac{1}{n\lambda^2}\right) = AN\left(\lambda, \frac{\lambda^4}{n\lambda^2}\right) = AN\left(\lambda, \frac{\lambda^2}{n}\right)$$

Inoltre, ricordando la formula per i momenti della gamma, la quale è appunto la distribuzione di una somma di VA esponenziali, si ottiene che:

$$\mathbb{E}[Y_n] = n\mathbb{E}\left[\frac{1}{\sum_{k=1}^n X_k}\right] = n \frac{1}{\lambda^{-1}} \frac{\Gamma(n-1)}{\Gamma(n)} = n\lambda \frac{(n-2)!}{(n-1)!} = \frac{n}{n-1} \lambda \xrightarrow{n} \lambda$$

Si noti che il valore atteso di  $Y_n$  tende asintoticamente a  $\lambda$ , ma non è *esattamente*  $\lambda$  per ogni  $n$  fissato. Per questa ragione lo stimatore  $Y_n$  viene chiamato *distorto*, definizione che preciseremo meglio nelle prossime righe.

Introduciamo, infine, un'ultima successione per perfezionare ulteriormente la stima di  $\lambda$ :  $T_n := \frac{n-1}{n} \cdot Y_n = \frac{n-1}{\sum_{k=1}^n X_k}$ , che ha valore atteso  $\mathbb{E}[T_n] \equiv \lambda \forall n$  e viene per questo definito *stimatore corretto* o *non distorto*. Tutti gli stimatori che non sono corretti sono detti, appunto, *distorti*. Si può inoltre mostrare che  $Var(T_n) = \frac{\lambda^2}{n-2}$ .

Non rimane che studiare la convergenza: poiché  $\frac{n-1}{n} \xrightarrow{n} 1$  e  $Y_n \xrightarrow{\text{qc}} \lambda$ , si ottiene che  $T_n = \frac{n-1}{n} Y_n \xrightarrow{\text{qc}} \lambda$ , ovvero  $T_n$  è asintoticamente normale. L'intuizione suggerisce che il fattore  $\frac{n-1}{n}$  lasci inalterata la media asintotica  $\lambda$  e la velocità di

convergenza  $\sqrt{n}$  di  $Y_n$ . In altre parole, l'aspettativa è che  $\sqrt{n}(T_n - \lambda) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \lambda^2)$ . Verifichiamo la correttezza della nostra *Ansatz*<sup>58</sup>:

$$\sqrt{n}(T_n - \lambda) = \sqrt{n} \left( \frac{n-1}{n} \cdot Y_n - \lambda \right) = \frac{n-1}{n} \sqrt{n}(Y_n - \lambda) + \frac{n-1}{n} \sqrt{n}\lambda - \sqrt{n}\lambda$$

Poiché  $\frac{n-1}{n} \rightarrow 1$  (quindi anche in legge) e  $\sqrt{n}(Y_n - \lambda) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \lambda^2)$  per ipotesi, applicando il teorema di Slutsky il primo addendo converge in legge a  $\mathcal{N}(0, \lambda^2)$ ; il secondo e il terzo addendo sommati danno  $\sqrt{n}\lambda \frac{n-1}{n} \xrightarrow{n} 0$ ; il tutto converge dunque in legge a  $\mathcal{N}(0, \lambda^2)$  e dunque effettivamente:

$$T_n \sim AN \left( \lambda, \frac{\lambda^2}{n} \right)$$

## 17.5 Stima dei parametri di una distribuzione

Un obiettivo frequente in statistica è ottenere più informazioni possibili su una popolazione  $X_n$ , ovvero una successione di VA iid in  $L^2$ , avente legge, parametri e funzione di ripartizione  $F$  ignoti. Applicando la teoria della probabilità introdurremo degli *stimatori puntuali* per ciascuna delle grandezze interessate e per ciascuno stimatore elencheremo le proprietà che lo rendono efficace.

### 17.5.1 Media

Lo stimatore di  $\mu = \mathbb{E}[X_n]$  è la già nota **media campionaria**:

$$\bar{X}_n = \frac{1}{n} \sum_{k=1}^n X_k$$

Infatti, a prescindere da  $F$  è vero che:

1.  $\mathbb{E}[\bar{X}_n] = \mu \quad \forall n$  (stimatore corretto).
2.  $Var(\bar{X}_n) = \frac{\sigma^2}{n} \xrightarrow{n} 0$  (ovvero  $\bar{X}_n \xrightarrow{L^2} \mu$ ). In termini statistici si dice che il MSE<sup>59</sup>  $\xrightarrow{n} 0$ .
3.  $\bar{X}_n \xrightarrow{qc} \mu$  per la LGN (**consistenza**).
4.  $\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1)$  per il TCL, ovvero  $\bar{X}_n \sim AN \left( \lambda, \frac{\sigma^2}{n} \right)$ .

Da quest'ultima proprietà discende la teoria della statistica su intervalli di confidenza e test d'ipotesi.

<sup>58</sup>Come ci ricorda il buon professor Micheletti, un'*Ansatz* è un'ipotesi realistica, in tedesco.

<sup>59</sup>Mean Square Error, errore quadratico medio.



Si possono visualizzare e riassumere questi risultati nel seguente grafico:

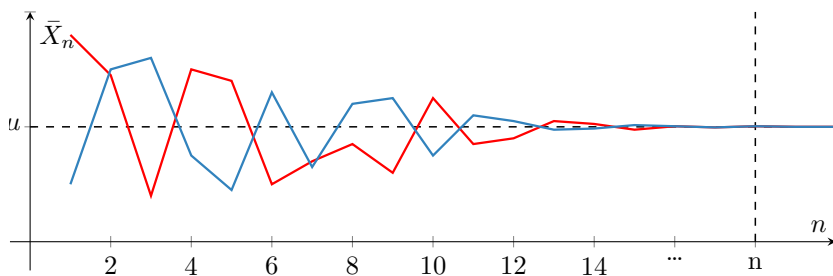


Figura 17.1: grafico della convergenza della media campionaria al valore atteso

Ogni spezzata rappresenta un diverso osservatore (supponiamo che siano in numero molto elevato) che effettua le proprie estrazioni separatamente dagli altri, aggiornando costantemente, all'aumentare del numero  $n$  di campioni raccolti, la propria media campionaria.

Le 4 osservazioni precedenti hanno ciascuna la propria rappresentazione visuale:

1. Correttezza: per ogni  $n$  fissato, la media di tutti valori di  $\bar{X}_n$  corrispondenti a quell' $n$  (detta anche, non a caso, *media verticale*) è esattamente  $\mu$ .
2. Convergenza in  $L^2$ : lo sparpagliamento dei valori di  $\bar{X}_n$  intorno a  $\mu$  per uno stesso  $n$  diminuisce verso lo zero all'aumentare di  $n$ .
3. Convergenza quasi certa: ciascuna linea spezzata rappresentante la successione  $\bar{X}_n$  di un osservatore tende al valore  $\mu$  all'aumentare di  $n$ ; questa quantità è anche detta *media orizzontale*.
4. Convergenza in legge: un eventuale grafico della frequenza di valori di  $\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}$  per uno stesso  $n$  fissato tenderebbe, all'aumentare di  $n$ , ad avere una forma gaussiana.

### 17.5.2 Varianza

Lo stimatore di  $\sigma^2 = \text{Var}(X_n)$  è la già nota **varianza campionaria**:

$$S_n^2 = \frac{1}{n-1} \sum_{k=1}^n (X_k - \bar{X}_n)^2$$

Per ogni  $F$ :

1.  $\mathbb{E}[S_n^2] = \sigma^2 \quad \forall n$  (corretto).

2. “Non si dice cosa succede alla varianza della varianza campionaria”<sup>60</sup>
3.  $S_n^2 \xrightarrow{qc} \sigma^2$  (consistente).
4.  $\frac{S_n^2 - \sigma^2}{\sqrt{\frac{\mu_4 - \sigma^4}{n}}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1)$ ,  
ovvero  $S_n^2 \sim AN\left(\sigma^2, \frac{\mu_4 - \sigma^4}{n}\right)$ , dove  $\mu_4 = \mathbb{E}[(X_n - \mu)^4] \neq \sigma^4$ .

### 17.5.3 Stima di una proporzione

Lo stimatore della proporzione  $p = \mathbb{P}(X_n \in B)$  è la **proporzione campionaria**:

$$\hat{p}_n = \frac{1}{n} \sum_{k=1}^n \mathbb{1}_B(X_k)$$

Essa non è altro che il numero dei campioni caduti effettivamente in  $B$  divisi per il numero totale dei campioni. Si noti inoltre che:

$$\mathbb{1}_B(X_k) = Y_k \sim Be(p), \quad \text{con } p = \mathbb{P}(X_k \in B) = \mathbb{P}(Y_k = 1)$$

Per ogni  $F$ , è vero che:

1.  $\mathbb{E}[\hat{p}_n] = p \quad \forall n$  (corretto).
2.  $Var(\hat{p}_n) = n \frac{p(1-p)}{n^2} = \frac{p(1-p)}{n} \xrightarrow{n} 0$ .
3.  $\hat{p}_n \xrightarrow{qc} p$  per la LGN.
4.  $\frac{\hat{p}_n - p}{\sqrt{\frac{p(1-p)}{n}}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1)$  per il TCL. Da questa proprietà discende la teoria della statistica per i test d'ipotesi.
5.  $\frac{\hat{p}_n - p}{\sqrt{\frac{\hat{p}_n(1-\hat{p}_n)}{n}}} = \frac{\hat{p}_n - p}{\sqrt{\frac{p(1-p)}{n}}} \cdot \frac{\sqrt{\frac{p(1-p)}{n}}}{\sqrt{\frac{\hat{p}_n(1-\hat{p}_n)}{n}}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1)$  per il teorema di Slutsky.

Infatti la prima frazione converge in legge a  $\mathcal{N}(0, 1)$  per il punto precedente e la seconda frazione, funzione continua di  $\hat{p}_n$ , converge qc a 1 per la LGN.

### 17.5.4 Funzione di ripartizione

Lo stimatore di  $F(t) = \mathbb{P}(X_n \leq t)$ , con  $t \in \mathbb{R}$ , è la **funzione di ripartizione empirica**:

$$F_n(t) = \frac{1}{n} \sum_{k=1}^n \mathbb{1}_{(-\infty, t)}(X_k) = \frac{1}{n} \sum_{k=1}^n \mathbb{1}_{[X_k, +\infty)}(t)$$

<sup>60</sup>Leggasi come “Non lo vuoi davvero sapere, fidati”

L'ultima scrittura equivalente è giustificata dal fatto che  $X_k \leq t \iff t \geq X_k$ . Ogni  $\omega \in \Omega$  è mappato, attraverso questa trasformazione, in una funzione di ripartizione: la  $F_n(t)$  sarebbe dunque, più precisamente, una  $F_n(t, \omega)$ , perché  $X_n = X_n(\omega)$  varia al variare dell'esito elementare, ovvero dell'osservatore, in quanto  $X_n$  può essere incluso oppure no nelle varie indicatrici. Supponendo che un osservatore (corrispondente a un preciso valore di  $t$ ) abbia trovato  $n$  esiti  $X_k$  tutti distinti, e disponendo gli  $X_k$  in ordine di valore crescente (in quanto esperimenti casuali non saranno necessariamente già in ordine) sull'asse delle ascisse, la funzione avrà  $n$  discontinuità, ciascuna in un  $X_k$ , e i salti saranno tutti pari a  $\frac{1}{n}$ . In caso di  $X_k$  sovrapposte, per esempio in quantità  $m$ , ci saranno  $n - m$  salti, e quelli in corrispondenza dei valori ottenuti più volte saranno più alti ( $\frac{m}{n}$ , per la precisione).

Le varie  $F_n$  hanno grafici che dipendono dall'esito realizzatosi e dall' $n$  con cui è stato effettuato l'esperimento: sono dunque a tutti gli effetti una successione di funzioni aleatorie. Anche a  $F_n$  sono estendibili le seguenti proprietà:

1.  $\mathbb{E}[F_n(t)] = F(t) \quad \forall t$  (corretto).
2.  $Var(F_n(t)) = \frac{F(t)(1-F(t))}{n} \quad \forall t$ .
3.  $F_n(t) \rightarrow F(t)$  qc  $\forall t$  (consistente).
4.  $F_n(t) \sim AN\left(F(t), \frac{F(t)(1-F(t))}{n}\right)$ .

La proprietà (3) può essere ampliata:

- 3'.  $F_n(t) \rightarrow F(t) \quad \forall t$ , qc (**convergenza puntuale quasi certa**). È una affermazione leggermente più forte della precedente: prima la convergenza era quasi certa per ogni singolo  $t$ , ovvero, se un osservatore trovasse l'evento improbabile (la non-“convergenza quasi certa”) l'intera  $F(t)$  rappresentata dall'insieme degli osservatori (ciascuno con la propria  $t$ ) non convergerebbe, a causa di quell'unico osservatore. Invece, scambiando “qc” e “ $\forall t$ ” si ottiene che l'intera  $F(t)$  ha convergenza quasi certa garantita, non solo le singole  $F_n(t)$  con  $t$  fissato (che non sono altro che semplici successioni numeriche). Con quest'ultima affermazione la convergenza è più stabile rispetto alle fluttuazioni aleatorie del singolo sperimentatore.

3". Vale il seguente risultato.

**Teorema 17.10: teorema di Glivenko-Cantelli**

$$\sup_{t \in \mathbb{R}} |F_n(t) - F(t)| \xrightarrow[n]{qc} 0$$

Il teorema rappresenta la **convergenza uniforme quasi certa** di  $F_n(t)$  (l'estremo superiore dell'errore tende a 0), ovvero, la convergenza avviene alla stessa velocità in tutti i punti del dominio.

## 17.6 Un controesempio: la legge di Cauchy

Talvolta nella vita non tutti i tipi di convergenza sono verificati contemporaneamente.

Siano  $X_n$  iid tali che  $X_n \sim \mathcal{C}(a, \sigma)$  (distribuzione di Cauchy), ovvero con le seguenti densità e funzione caratteristica, rispettivamente:

$$f(t) = \frac{1}{\pi\sigma} \frac{1}{1 + \left(\frac{t-a}{\sigma}\right)^2} \quad \text{e} \quad \varphi(u) = \exp\{iau - \sigma|u|\}.$$

Poiché  $\mathbb{E}[|X_n|] = \int_{-\infty}^{+\infty} |t|f(t)dt$  diverge, in quanto l'integranda è asintotica a  $\frac{1}{t}$ ,  $X_n \notin L^1$ . Visto che  $X_n$  non può avere media,  $a$ , che è il valore centrale della distribuzione, può essere solo una mediana. Ma per la LGN (che ricordiamo essere una condizione necessaria e sufficiente) non c'è nemmeno convergenza quasi certa della media campionaria  $\bar{X}_n$  ad  $a$ : infatti  $\bar{X}_n$  è funzione di variabili non  $L^1$  e quindi non è  $L^1$  lei stessa. Ovvero,  $\bar{X}_n \not\xrightarrow{q.c.} a$ .

Non resta che controllare se è almeno verificata la convergenza in legge di  $\bar{X}_n$ :

$$\begin{aligned} \varphi_{\bar{X}_n}(u) &= \varphi_{\frac{1}{n} \sum X_k}(u) = \varphi_{\sum X_k} \left( \frac{u}{n} \right) \\ &= \left( \varphi \left( \frac{u}{n} \right) \right)^n = \left( \exp \left\{ ia \frac{u}{n} - \sigma \frac{|u|}{n} \right\} \right)^n \quad (\text{perché VA iid}) \\ &= \exp\{iau - \sigma|u|\} = \varphi(u) \end{aligned}$$

Tutte le medie campionarie sono ancora variabili Cauchy, e quindi c'è convergenza in legge Cauchy:

$$\bar{X}_n \sim \mathcal{C}(a, \sigma) \quad \forall n \quad \implies \quad \bar{X}_n \xrightarrow{\mathcal{L}} \mathcal{C}(a, \sigma)$$

---

## 18 Catene di Markov

In questo capitolo finale si vuole descrivere l'*evoluzione di una grandezza* distribuita in maniera casuale e descritta da una famiglia di variabili aleatorie, studiando quelli che verranno definiti come *processi stocastici*. In questo testo sarà trattata solo una sottoclasse di processi, quella delle *catene di Markov a tempo discreto*, nelle quali il valore futuro della grandezza dipende solamente dal valore attuale e non dalla storia dell'evoluzione. Si studieranno quindi le leggi secondo cui una grandezza evolve da uno *stato* (cioè un valore) ad un altro e le caratteristiche dei suddetti stati.

### 18.1 Definizioni base

#### Definizione 18.1

Un **processo stocastico** è una famiglia di VA  $(X_t)_{t \in T}$ ,  $X_t : \Omega \rightarrow E$ , tutte definite sullo stesso spazio di probabilità  $(\Omega, \mathcal{A}, \mathbb{P})$ , a valori nello stesso spazio misurabile  $(E, \mathcal{E})$ , e dipendenti da un parametro  $t \in T \subseteq \mathbb{R}^+ = [0, +\infty)$ . L'insieme  $E$  è detto **spazio degli stati** e i suoi elementi **stati**.

Il processo stocastico è una nozione generalizzata rispetto alle *successioni* di VA trattate nei capitoli precedenti. La principale differenza risiede nel fatto che in un processo stocastico il parametro  $t$  appartiene a un generico insieme  $T$  e quindi può anche essere *continuo*, a differenza dell'indice  $n \in \mathbb{N}$  delle successioni (anche se, in questo capitolo,  $T$  sarà spesso discreto per semplicità di trattazione). Inoltre, il codominio delle VA è un generico spazio misurabile  $(E, \mathcal{E})$  e non necessariamente  $\mathbb{R}$ . Tipicamente, infine, il parametro  $t$  rappresenta il *tempo* della misurazione o dell'osservazione.

*Esempio.* Un esempio frequente di processo stocastico è una grandezza misurata ad istanti  $t$  di tempo diversi. Consideriamo il caso del *tempo discreto*, che descrive una *evoluzione a scatti*:

$$T \subseteq \mathbb{Z}^+ = \{0, 1, 2, \dots\}$$

La  $\sigma$ -algebra del codominio è l'insieme delle parti  $\mathcal{E} = 2^E$ , con  $E$  discreto.

#### Definizione 18.2

Un processo stocastico si dice **catena** se il suo spazio degli stati  $E$  è discreto.

*Esempio.* L'esempio precedente è una catena, così come ogni *successione* di VA della forma  $X_n : \Omega \rightarrow E$ , con  $E$  discreto e  $n \in \mathbb{Z}^+$ , che d'ora in poi scriveremo come  $n \geq 0$  per semplicità di notazione.

*Esempio.* Sono catene a tempo discreto valide:

- il risultato dell' $n$ -esimo lancio di una moneta che viene lanciata infinite volte, dove ciascun lancio avviene ad un preciso istante di tempo:  $X_n : \Omega \rightarrow E = \{T, C\}$ ;
- il tempo metereologico a Milano dell' $n$ -esimo giorno del 2019:  $X_n : \Omega \rightarrow \{\text{sole, pioggia, } \dots\}$ ;
- il valore di un bit dopo la trasmissione attraverso l' $n$ -esimo canale rumoroso:  $X_n : \Omega \rightarrow \{0, 1\}$ ;
- il capitale di un giocatore d'azzardo dopo l' $n$ -esimo lancio di una moneta, che gli fa guadagnare se esce testa e gli fa perdere se esce croce:  $X_n : \Omega \rightarrow \mathbb{R}$ .

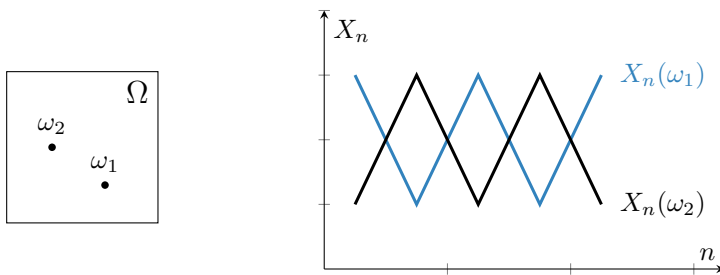


Figura 18.1: catene  $X_n$  rispetto a diversi stati iniziali  $\omega_1$  e  $\omega_2$

Fissando l'esito elementare  $\omega \in \Omega$ , si ottiene una *traiettoria* nello spazio degli stati  $E$ , ovvero una successione di valori  $\{x_n\}_n = \{X_n(\omega)\}_n \subseteq E$ . Chiaramente, fissando invece l'indice  $n \geq 0$ , si ottiene una singola VA  $X_n : \Omega \rightarrow E$ .

### Definizione 18.3

Una matrice  $P = [p_{ij}]$  con  $i, j \in E$  è detta **matrice di transizione** o **matrice stocastica** se:

$$p_{ij} \geq 0 \quad \forall i, j \in E \quad \text{e} \quad \sum_{j \in E} p_{ij} = 1 \quad \forall i \in E$$

In altre parole, la  $i$ -esima riga  $p_i$  è una *densità discreta* di probabilità su  $E$ ,  $\forall i \in E$ . Ogni elemento  $p_{ij}$  di questa matrice è detto **probabilità di transizione** da  $i$  a  $j$ .

L'intuizione suggerisce che  $p_{ij}$  sia effettivamente la probabilità, partendo dallo stato  $i$  di spostarsi nello stato  $j$  in un unico passo. Questo significato sarà verificato formalmente più avanti. Nel frattempo, rafforzeremo l'intuizione sul modello di catena a tempo discreto mediante alcuni esempi.

*Esempio.*

1. Sia  $E = \{1, 2\}$ . Dati  $p, q \in [0, 1]$ , un esempio di matrice di transizione è la seguente:

$$P = \begin{bmatrix} 1-p & p \\ q & 1-q \end{bmatrix}$$

Intuitivamente, questo significa che il sistema descritto da  $P$  ha, in ogni istante di tempo  $n \geq 0$ , due stati possibili, detti 1 e 2. Le probabilità sono così distribuite:

- \*  $p$  indica la probabilità che lo stato passi da 1 a 2 in un determinato istante di tempo;
- \*  $q$  che lo stato passi da 2 a 1;
- \*  $1-p$  che lo stato rimanga in 1;
- \*  $1-q$  che lo stato rimanga in 2.

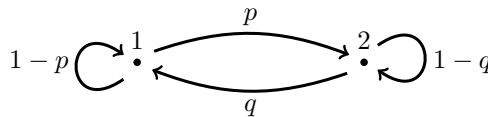


Figura 18.2: illustrazione della catena dell'esempio 1

2. Dato un sistema a tre stati  $E = \{1, 2, 3\}$ , che varia secondo un tempo discreto  $n \geq 0$ , si può avere la seguente matrice di transizione:

$$P = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 1/2 & 1/2 \\ 1/2 & 0 & 1/2 \end{bmatrix}$$

Il sistema non può compiere tutte le transizioni possibili. Per esempio, non può ripercorrere al contrario il "ciclo"  $1 \rightarrow 2 \rightarrow 3$ : può solo andare avanti o, nel caso si trovi in 2 o in 3, rimanere nello stesso stato. Questo perché la probabilità di spostarsi da 3 a 2 (in un unico passo) è nulla:  $p_{32} = 0$ . Si noti, comunque, che questo *non* impedisce il ritorno da 3 a 2 in  $n \geq 2$  passi mediante passaggi intermedi in altri stati (in questo caso, 1).

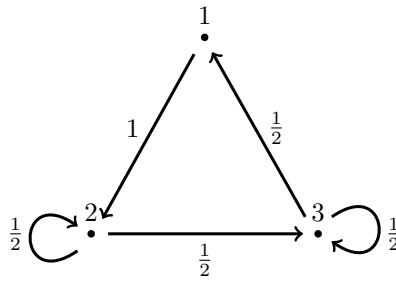


Figura 18.3: illustrazione della catena dell'esempio 2

3. Considerando ancora  $E = \{1, 2, 3\}$  un'altra matrice può essere:

$$P = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Questo sistema possiede due “sottosistemi” che non comunicano: partendo da 1 si può solo passare a 2 e viceversa, mentre partendo da 3 non si può che rimanere per sempre in 3. L'evoluzione, in questo caso, è quindi parzialmente “deterministica”, cioè si possono effettuare alcune previsioni conoscendo lo stato di partenza.

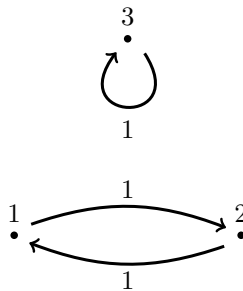


Figura 18.4: illustrazione della catena dell'esempio 3

**Definizione 18.4**

Dati uno spazio di probabilità  $(\Omega, \mathcal{A}, \mathbb{P})$  e uno spazio misurabile  $(E, 2^E)$ , una famiglia di VA  $X_n : \Omega \rightarrow E$ , con  $n \geq 0$ , è detta **catena di Markov omogenea** a valori in  $E$  di **legge iniziale**  $v$  e **matrice di transizione**  $P = [p_{ij}]$ , con  $i, j \in E$ , se:



1.  $X_0 \sim v$ ;
2.  $\forall n \geq 0$  e  $\forall i, j \in E$ , si ha la **proprietà di Markov**:

$$\mathbb{P}(X_{n+1} = j | X_n = i, X_{n-1} = i_{n-1}, \dots, X_0 = i_0) = \mathbb{P}(X_{n+1} = j | X_n = i)$$

qualora le suddette probabilità condizionate esistano;

3. vale la proprietà di **omogeneità**:

$$\mathbb{P}(X_{n+1} = j | X_n = i) = p_{ij} \quad \forall i, j \in E$$

ovvero la probabilità di transizione *non* dipende da  $n$ .

Per indicare una catena di Markov omogenea definita come sopra useremo la notazione compatta  $(X_n)_{n \geq 0}$  Markov( $v, P$ ).

### 18.1.1 Interpretazione

La proprietà di Markov afferma che, quando si conosce il valore *presente* di una VA ( $X_n = i$ ) e si vuole fare una stima sul valore *futuro* (nell'istante di tempo successivo) di quella VA ( $X_{n+1} = j$ ), conoscere i valori *passati* di quella VA ( $X_{n-1} = i_{n-1}, \dots, X_0 = i_0$ ) non fornisce alcuna informazione aggiuntiva rispetto al conoscere il valore presente. In altre parole, noti lo stato attuale e l'istante di osservazione di una catena di Markov, **il prossimo valore** assunto dalla catena **dipende solo dal valore attuale**. Tale proprietà è confermata dall'intuizione per alcuni tipi di fenomeno: per esempio, se un processo stocastico indica la posizione nello spazio di una particella e la sua velocità, è naturale pensare che i dati attuali siano sufficienti per cercare di prevedere la sua posizione nel successivo istante di tempo, senza bisogno di conoscere la sua storia.

Si noti, comunque, che questa proprietà *non* significa che i valori futuri sono indipendenti dai valori passati, perché queste caratteristiche di “quasi-indipendenza” richiedono la conoscenza dello stato attuale<sup>61</sup>, che a seconda del modello preso in esame può anche non essere accessibile. È bene infatti rimarcare che le variabili  $X_n$  della catena *non sono indipendenti* tra loro né tantomeno identicamente distribuite.

Si noti, infine, che in una catena di Markov è cruciale la scelta di due elementi: la VA osservata  $X_n$  e lo spazio degli stati  $E$ . Nell'esempio precedente della particella, la sola posizione non è un'informazione sufficiente per formulare la previsione, e servirebbero informazioni aggiuntive dal suo passato.

<sup>61</sup> Vedremo infatti che la conoscenza dell'istante di osservazione non è necessaria per formulare previsioni sull'evoluzione di una catena di Markov omogenea.

### 18.2 Esempi notevoli

Introduciamo ora due esempi particolarmente significativi di catene di Markov, che saranno ripresi più volte nel corso della trattazione dell'argomento: la rovina del giocatore e l'urna di Ehrenfest.

*Esempio:* rovina del giocatore. Immaginiamo due giocatori, Aron e Bruno hanno ciascuno un capitale iniziale, rispettivamente  $a, b \in \mathbb{N}$ . Ad ogni istante di tempo viene lanciata una moneta, che ha probabilità  $p$  di risultare testa, e possono accadere due eventi: se esce testa, Aron riceve 1 da Bruno; se esce croce invece Bruno riceve 1 da Aron. I lanci di moneta continuano fino a che uno dei due giocatori non entra in rovina esaurendo il suo capitale.

Il numero totale di lanci effettuati prima della rovina di uno dei due giocatori è casuale e incognito a priori. Una formulazione equivalente del gioco è che i lanci proseguano anche dopo la rovina di uno dei due giocatori, ma da quel momento in poi non ci siano più scambi di denaro. Questo permette di operare con le catene di Markov, in quanto si ha un numero infinito di lanci, ovvero di VA.

Sia dunque  $X_n$  il capitale di Aron dopo l' $n$ -esimo lancio. Grazie al cambio di formulazione sopra discusso si ha  $T = \mathbb{Z}^+$ . Inoltre  $E = \{0, 1, \dots, a + b\}$ , in quanto un giocatore può possedere al massimo la somma dei due capitali, e  $v = \delta_a$ , in quanto si suppone deterministicamente che Aron inizi sempre il gioco con capitale  $a$ . La matrice di transizione è tridiagonale, eccetto che per la prima e l'ultima riga, in quanto una volta raggiunta la rovina si rimane per sempre in quello stato con probabilità 1:

$$P = \begin{bmatrix} 1 & 0 & 0 & 0 & \dots \\ 1-p & 0 & p & 0 & \dots \\ 0 & 1-p & 0 & p & \dots \\ & & \ddots & \ddots & \ddots \\ & \dots & 0 & 1-p & 0 & p \\ & \dots & 0 & 0 & 0 & 1 \end{bmatrix}$$

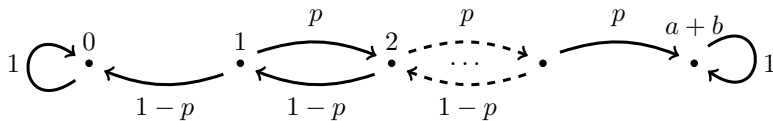


Figura 18.5: rovina del giocatore

*Esempio:* urna di Ehrenfest. Forniamo due formulazioni equivalenti del problema:

- si ha un'urna contenente  $N$  palline, di colore bianco oppure nero. Ad ogni istante di tempo si estrae una pallina a caso (con legge uniforme), la si sostituisce con una del colore opposto e la si rimette nell'urna. Qui  $X_n$  è il numero di palline nere dopo l' $n$ -esima estrazione.
- si ha un sistema chiuso formato da 2 recipienti comunicanti,  $A$  e  $B$ , che contiene in totale  $N$  molecole di gas. Ad ogni istante di tempo, una molecola a caso (con legge uniforme) nel sistema si trasferisce dal recipiente in cui si trova, nell'altro recipiente. Qui  $X_n$  è il numero di molecole nel recipiente  $A$  dopo l' $n$ -esimo istante di tempo.

In entrambi i casi  $E = \{0, 1, \dots, N\}$ ,  $T = \mathbb{Z}^+$  (similmente all'esempio precedente), e la proprietà di Markov è soddisfatta. Anche qui abbiamo una matrice tridiagonale, questa volta incluse la prima e l'ultima riga, ma con probabilità variabile a seconda dello stato. Per esempio, la probabilità di passare da 2 a 1 palline nere (o molecole nel recipiente  $A$ ) è pari alla probabilità che venga pescata una delle 2 su un totale di  $N$ ; ovvero,  $p_{21} = \frac{2}{N}$ , e così via per tutte le altre probabilità di transizione. La matrice di transizione risultante è la seguente:

$$P = \begin{bmatrix} 0 & 1 & 0 & 0 & \dots \\ \frac{1}{N} & 0 & \frac{N-1}{N} & 0 & \dots \\ 0 & \frac{2}{N} & 0 & \frac{N-2}{N} & \dots \\ & & \ddots & \ddots & \ddots \\ \dots & 0 & \frac{N-1}{N} & 0 & \frac{1}{N} \\ \dots & 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

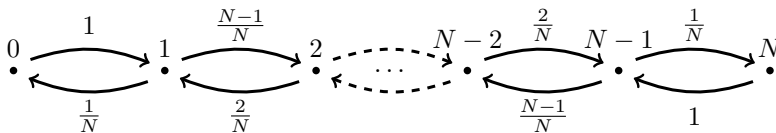


Figura 18.6: urna di Ehrenfest

### 18.3 Leggi congiunte

Data una catena  $(X_n)_{n \geq 0}$  Markov( $v, P$ ), per definizione  $X_0 \sim v$  e quindi si ha che:

$$X_1 | (X_0 = i) \sim p_i \text{ in quanto } \mathbb{P}(X_0 = i) = v_i \text{ e } \mathbb{P}(X_1 = j | X_0 = i) = p_{ij}$$

Pertanto, grazie alle proprietà della legge condizionata (capitolo 13):

$$\mathbb{P}(X_0 = i, X_1 = j) = \mathbb{P}(X_0 = i) \mathbb{P}(X_1 = j | X_0 = i) = v_i p_{ij}$$

Si può dunque constatare che conoscendo la legge marginale di  $X_0$  e la legge condizionata di  $X_1|X_0$ , è nota anche la legge congiunta di  $(X_0, X_1)$ . Questa “concatenazione” di leggi mediante un prodotto<sup>62</sup> funziona anche per ottenere la legge congiunta di un numero qualsiasi di VA della catena:

**Teorema 18.5**

Sia  $(X_n)_{n \geq 0}$  Markov( $v, P$ ). Allora,  $\forall n \geq 0$  e  $\forall i_0, \dots, i_n \in E$ , si ha:

$$\mathbb{P}(X_0 = i_0, X_1 = i_1, \dots, X_n = i_n) = v_{i_0} p_{i_0 i_1} p_{i_1 i_2} \cdots p_{i_{n-1} i_n}$$

**Dimostrazione**

1. Si supponga  $\mathbb{P}(X_0 = i_0, \dots, X_n = i_n) > 0$ .

Usando al contrario la definizione di probabilità condizionata sugli eventi  $(X_n = i_n)$  e  $(X_0 = i_0, \dots, X_{n-1} = i_{n-1})$  si ha che:

$$\begin{aligned} &\mathbb{P}(X_0 = i_0, \dots, X_n = i_n) \\ &= \mathbb{P}(X_n = i_n | X_{n-1} = i_{n-1}, \dots, X_0 = i_0) \mathbb{P}(X_{n-1} = i_{n-1}, \dots, X_0 = i_0) \\ &= p_{i_{n-1} i_n} \mathbb{P}(X_{n-1} = i_{n-1}, \dots, X_0 = i_0) \quad (\text{per la pr. Markov}) \\ &= p_{i_{n-1} i_n} \mathbb{P}(X_{n-1} = i_{n-1} | X_{n-2} = i_{n-2}, \dots, X_0 = i_0) \cdot \\ &\quad \cdot \mathbb{P}(X_{n-2} = i_{n-2}, \dots, X_0 = i_0) \\ &= \dots \quad (\text{ripetendo } n \text{ volte}) \\ &= p_{i_{n-1} i_n} \cdots p_{i_1 i_0} v_{i_0} \end{aligned}$$

2. Il caso  $\mathbb{P}(X_0 = i_0, \dots, X_n = i_n) = 0$  è un utile esercizio per il lettore.<sup>63</sup>  $\square$

Questo teorema mostra come l'evoluzione del fenomeno dipenda fortemente dalla legge iniziale. È dunque impensabile cercare di studiare l'evoluzione di una grandezza al variare di  $v$  e tenendo fissa la matrice di transizione  $P$ , perché cambierebbero tutte le leggi congiunte creando quindi altre catene di Markov diverse e slegate tra loro. Si dice dunque che la catena di Markov è definita dai due *parametri*  $P$  e  $v$ : ciò è evidenziato dalla notazione compatta scelta per indicarle.

Il caso  $v \neq \delta$  indica una situazione iniziale non deterministica e non nota a priori, per esempio il caso in cui l'inizio della catena sia il prodotto di un altro esperimento aleatorio avvenuto precedentemente.

Introduciamo ora una notazione largamente utilizzata per la sua brevità: si indichi con  $\mathbb{P}_i(\cdot)$  la probabilità  $\mathbb{P}(\cdot | X_0 = i)$ , quindi sottintendendo che è noto lo stato iniziale  $i$  della catena.

<sup>62</sup>Questa proprietà e altre simili (anche viste in altri capitoli di questo testo) che permettono di ricavare una legge congiunta dal prodotto di altre leggi sono incluse nella categoria delle cosiddette *leggi della catena*.

<sup>63</sup>Dimostrazione del lettore quotata a 1.1

**Corollario 18.6**

Sia  $(X_n)_{n \geq 0}$  Markov( $v, P$ ) sullo spazio di probabilità  $(\Omega, \mathcal{A}, \mathbb{P})$ . Siano inoltre  $v_i = \mathbb{P}(X_0 = i) > 0 \forall i$  e  $\mathbb{P}_i(A) = \mathbb{P}(A|X_0 = i) \forall A \in \mathcal{A}$ . Allora  $(X_n)_{n \geq 0}$  è Markov( $v, P$ ) anche sullo spazio di probabilità  $(\Omega, \mathcal{A}, \mathbb{P}_i)$ .

Ciò non è scontato, perché in generale sostituendo  $\mathbb{P}$  con una legge qualsiasi non si ha alcuna garanzia sul fatto che si ottenga ancora una catena di Markov. Il risultato permette di semplificare i conti e la notazione quando è nota la legge iniziale, di fatto “incorporando” la situazione iniziale direttamente nella legge di probabilità della catena. Vedremo che un risultato simile vale anche se  $v \sim \delta$  (quindi anche con componenti nulle di  $v$ ).

**18.3.1 Notazione matriciale**

Si identifichi ora ogni densità  $\pi$  di probabilità discreta su  $E$  con un *vettore riga* di componenti  $\pi_i, i \in E$ , con  $\pi_i \geq 0 \forall i$  e  $\sum_i \pi_i = 1$ . Si introduca il vettore riga  $\pi P$ , tale che:

$$(\pi P)_j = \sum_i \pi_i p_{ij} \quad \forall j \in E$$

Infine, si introducano anche la matrice  $P^2$ , tale che:

$$(P^2)_{ij} = \sum_k p_{ik} p_{kj} \quad \forall i, j \in E$$

Analogamente le altre potenze  $P^n$  di  $P$ , con  $n \geq 0$ .

Indicheremo con  $p_{ij}^{(n)}$  l'elemento in posizione  $(i, j)$  della matrice  $P^n$ ; dunque:

$$P^n = \left[ p_{ij}^{(n)} \right]_{i, j \in E}$$

Similmente denoteremo con  $v^{(n)}$  la legge della VA  $X_n$ : si ha dunque  $X_n \sim v^{(n)}$  e  $v^{(0)} = v$ .

**Teorema 18.7**

Sia  $(X_n)_{n \geq 0}$  Markov( $v, P$ ). Allora:

(i)  $v^{(n)} = v^{(0)} P^n \quad \forall n \geq 0$ , ovvero in componenti:

$$\mathbb{P}(X_n = j) = (v P^n)_j \quad \forall n \geq 0, j \in E$$

(ii) vale la seguente uguaglianza, ammesso che le probabilità condizionate ivi utilizzate esistano:

$$p_{ij}^{(n)} = \mathbb{P}(X_{m+n} = j | X_m = i) = \mathbb{P}(X_n = j | X_0 = i) \quad \forall m, n \geq 0, i, j \in E$$

(iii) per ogni  $k, n_1, \dots, n_k \geq 0, i_1, \dots, i_k \in E$ ,

$$\mathbb{P}(X_{n_1} = i_1, \dots, X_{n_k} = i_k) = \sum_{i_0 \in E} v_{i_0} p_{i_0 i_1}^{(n_1)} p_{i_1 i_2}^{(n_2 - n_1)} \dots p_{i_{k-1} i_k}^{(n_k - n_{k-1})}$$

Nel punto (ii),  $p_{ij}^{(n)}$  è la probabilità di transizione di  $n$  passi partendo dal tempo  $m$ ; in questo caso quindi la matrice di transizione del passaggio è proprio  $P^{(n)}$ , e tale passaggio *non dipende dall'istante temporale  $m$  di partenza!* Questo è un effetto dell'omogeneità (temporale, cioè in  $n$ ) della catena di Markov: le previsioni sul valore di  $X_n$  non dipendono dal tempo di osservazione, ma solo dallo stato in cui la catena si trova attualmente.

Inoltre, il punto (i) rappresenta una *formula delle probabilità totali* rispetto ai valori di  $X_0$ :

$$\begin{aligned} \mathbb{P}(X_n = j) &= \sum_{i \in E} \mathbb{P}(X_n = j, X_0 = i) = \sum_{i \in E} \mathbb{P}(X_n = j | X_0 = i) \mathbb{P}(X_0 = i) \\ &\implies \mathbb{P}(X_n = j) = \sum_{i \in E} v_i p_{ij}^{(n)} \end{aligned}$$

dove  $p_{ij}^{(n)}$  è la probabilità di passare da  $i$  a  $j$  in esattamente  $n$  passi.

### Dimostrazione

(i) Fattorizzando le probabilità e grazie alla definizione di  $P^n$ :

$$\begin{aligned} \mathbb{P}(X_n = j) &= \sum_{i_0, \dots, i_{n-1}} \mathbb{P}(X_n = j, X_{n-1} = i_{n-1}, \dots, X_0 = i_0) \\ &\stackrel{18.5}{=} \sum_{i_0, \dots, i_{n-1}} v_{i_0} p_{i_0 i_1} \dots p_{i_{n-2} i_{n-1}} p_{i_{n-1} j} \\ &= (vP^n)_j \end{aligned}$$

(ii) In caso  $m = 0$  e  $v = \delta_i$ , l'uguaglianza  $p_{ij}^{(n)} = \mathbb{P}(X_n = j | X_0 = i)$  è automaticamente soddisfatta per il punto (i).

(iii) Si dimostra, a titolo di esempio, un caso particolare con sottovettore appar-

tenente a un vettore di 9 componenti:

$$\begin{aligned}
 \mathbb{P}(X_3 = i_3, X_5 = i_5, X_9 = i_9) &= \sum_{\substack{i_0, i_1, i_2, i_4, \\ i_6, i_7, i_8}} \mathbb{P}(X_0 = i_0, \dots, X_9 = i_9) \\
 &= \sum_{i_0} v_{i_0} p_{i_0 i_1} \dots p_{i_8 i_9} \\
 &= \sum_{i_0} v_{i_0} \sum_{i_1, i_2} p_{i_0 i_1} p_{i_1 i_2} p_{i_2 i_3} \sum_{i_4} p_{i_3 i_4} p_{i_4 i_5} \sum_{i_6, i_7, i_8} p_{i_5 i_6} p_{i_6 i_7} p_{i_7 i_8} p_{i_8 i_9} \\
 &= \sum_{i_0} v_{i_0} p_{i_0 i_3}^3 p_{i_3 i_5}^2 p_{i_5 i_9}^4 \stackrel{(ii)}{=} \sum_{i_0} v_{i_0} p_{i_0 i_3}^{(3)} p_{i_3 i_5}^{(2)} p_{i_5 i_9}^{(4)}
 \end{aligned}$$

(ii) Siano ora  $m \geq 1$  e  $\mathbb{P}(X_m = i) > 0$ :

$$\begin{aligned}
 \mathbb{P}(X_{m+n} = j | X_m = i) &= \frac{\mathbb{P}(X_m = i, X_{m+n} = j)}{\mathbb{P}(X_m = i)} \stackrel{(iii)}{=} \frac{\sum_{i_0} v_{i_0} p_{i_0 i}^{(m)} p_{ij}^{(n)}}{\sum_{i_0} v_{i_0} p_{i_0 i}^{(m)}} \\
 &= \frac{p_{ij}^{(n)} \sum_{i_0} v_{i_0} p_{i_0 i}^{(m)}}{\sum_{i_0} v_{i_0} p_{i_0 i}^{(m)}} = p_{ij}^{(n)} \quad \square
 \end{aligned}$$

**Teorema 18.8: equazione di Chapman-Kolmogorov**

$$\mathbb{P}(X_{n+m} = j | X_0 = i) = \sum_{l \in E} \mathbb{P}(X_{n+m} = j | X_m = l) \mathbb{P}(X_m = l | X_0 = i)$$

Scelto un istante  $m$ , si possono dunque elencare tutte gli stati  $l$  che la catena di Markov può assumere in  $m$ , ovvero tutti i possibili “percorsi” osservati all’istante  $m$ , ed usarli per calcolare la probabilità di arrivare in  $j$ . Mediante un’accorta scelta di  $m$ , spesso questo calcolo si semplifica enormemente grazie ad eventuali percorsi che hanno probabilità nulla e ad altre peculiarità della catena analizzata.

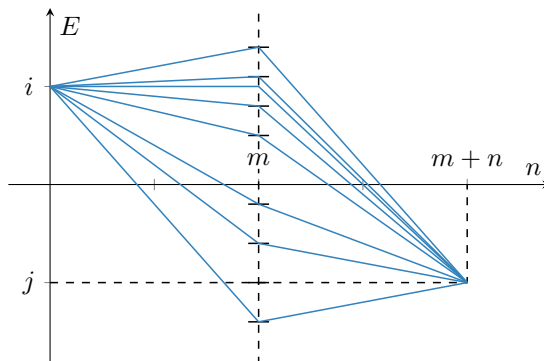


Figura 18.7: Chapman-Kolmogorov: si calcolano tutti gli stati intermedi nell’istante  $m$ , sfruttandoli per calcolare la probabilità di arrivare in  $j$ .

**Dimostrazione**

Dall'algebra delle matrici ricordiamo che  $P^{m+n} = P^m P^n$ , e in particolare:

$$p_{ij}^{(n+m)} = \sum_{l \in E} p_{il}^{(m)} p_{lj}^{(n)} \quad \forall i, j \in E$$

L'equazione non è altro che la riscrittura di quest'ultima formula, tenuto conto del punto (ii) del teorema 18.7.  $\square$

**Teorema 18.9**

Sia  $(X_n)_{n \geq 0}$  Markov( $v, P$ ) su  $(\Omega, \mathcal{A}, \mathbb{P})$ , e sia  $v_i^{(m)} = \mathbb{P}(X_m = i) > 0$  con  $m \geq 0$  fissato ed  $i \in E$ . La successione di VA:

$$(Y_n)_{n \geq 0} := (X_{n+m})_{n \geq 0}$$

è anch'essa Markov( $\delta_i, P$ ) su  $(\Omega, \mathcal{A}, \mathbb{P}_{i,m})$ , dove  $\mathbb{P}_{i,m}(A) = \mathbb{P}(A|X_m = i)$ .

Questo teorema conferma le osservazioni al teorema 18.7: definendo una nuova catena stiamo scegliendo un nuovo istante zero e “riscaldando” il tempo all'indietro per farlo diventare un istante zero “effettivo”. Ciò è possibile grazie all'omogeneità della catena, ovvero al fatto che una probabilità di transizione è identica se si è all'istante zero o se si è in qualsiasi altro istante. Mediante questa operazione, si ottiene una nuova catena che mantiene la stessa evoluzione  $P$  della precedente e ha una legge iniziale deterministica nel valore  $i$ . In altre parole, sapere di partire da una situazione del tipo  $\delta_i$  e sapere di essere in  $i$  con uno stato iniziale casuale sono due situazioni equivalenti, cosa che è confermata dall'intuizione sul significato della proprietà di Markov.

**18.4 Classificazione degli stati**

I “salti” da uno stato all'altro che la catena di Markov può compiere, seguendo le probabilità  $p_{ij}$ , fanno sì che nella catena si crei una diversificazione degli stati: possono esserci stati mai raggiunti, stati raggiunti più volte, gruppi di stati i cui raggiungimenti si escludono a vicenda, e così via. Iniziamo col dare l'importante definizione di stati e catene *periodici* e *aperiodici*:

**Definizione 18.10**

Il **periodo** di uno stato  $i \in E$  è definito nel seguente modo:

$$d(i) := \text{MCD} \left\{ n \geq 1 : p_{ii}^{(n)} > 0 \right\}$$



La definizione prende tutti i possibili percorsi (con probabilità non nulla) che vanno dallo stato  $i$  a sé stesso, ne prende le durate (cioè il numero di passi effettuati), e calcola il massimo comune denominatore di tutte queste durate.<sup>64</sup>

### Definizione 18.11

Uno stato  $i \in E$  è detto **periodico** se  $d(i) \geq 2$  e **aperiodico** se  $d(i) = 1$ .

Similmente, se tutti gli stati di una catena di Markov hanno lo stesso periodo  $d(i)$ , la **catena** si dice **aperiodica** o **di periodo**  $d(i)$ .

*Esempio.* Siano:

$$v = \delta_i \quad \text{e} \quad P = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

In questo caso la catena è deterministica: da 1 si passa sempre in 2 e viceversa. È evidente che la catena è periodica di periodo  $d(i) = 2$ . Verificando con la definizione, si ha infatti che  $P^2 = I$ ,  $P^3 = P$ , e così via: si sta dunque facendo l'MCD di tutti i numeri pari.

*Esempio.* Siano:

$$v = \delta_i \quad \text{e} \quad P = \begin{bmatrix} 0 & 1 \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix}$$

In questo caso  $d(1) = 1$ , in quanto partendo da 1 c'è una probabilità non nulla di tornare in 1 in 2, 3, 4, 5, ... passi, e l'MCD di questi valori è 1.

**NB:** Se  $p_{ii} > 0$ , allora necessariamente  $d(i) = 1$ .

### Proposizione 18.12

$$p_{ij}^{(n)} > 0 \quad \iff \quad \exists i_1, \dots, i_{n-1} \text{ tale che } p_{ii_1} p_{i_1 i_2} \cdots p_{i_{n-1} j} > 0$$

### Dimostrazione

Per definizione di  $p_{ij}^{(n)}$ :

$$p_{ij}^{(n)} = \sum_{j_1, \dots, j_{n-1}} p_{ij_1} p_{j_1 j_2} \cdots p_{j_{n-1} j} \quad \square$$

### Definizione 18.13

Si dice che lo stato  $i \in E$  **conduce** allo stato  $j \in E$ , scritto come  $i \rightarrow j$ , se esiste  $n \geq 0$  tale che  $p_{ij}^{(n)} > 0$ .

<sup>64</sup>Questa operazione è del tutto analoga all'effettuare l'MCD delle frequenze di un pacchetto di onde per determinare il periodo dell'onda globale. La frequenza infatti non è altro che l'inverso del tempo che un'onda impiega per tornare uguale a sé stessa.

<sup>65</sup>Abbiamo visto dimostrazioni decisamente peggiori.

**Proposizione 18.14**

$$i \rightarrow j \iff \mathbb{P} \left( \bigcup_{n \geq 1} (X_{n+m} = j) | X_m = i \right) > 0 \quad \forall m \geq 0 \quad (a)$$

$$\iff \exists i_1, \dots, i_{n-1} \text{ tale che } p_{ii_1} p_{i_1 i_2} \dots p_{i_{n-1} j} > 0 \quad (b)$$

Ovvero, esiste almeno un percorso con probabilità non nulla che porta da  $i$  a  $j$ , e questo evento accadrà con probabilità non nulla.

**Dimostrazione**

Se  $i \rightarrow j$ , per definizione esiste  $n_0$  tale che  $p_{ij}^{(n_0)} > 0$ . Pertanto:

$$\mathbb{P} \left( \bigcup_{n \geq 1} (X_{n+m} = j | X_m = i) \right) \geq \mathbb{P}(X_{n_0+m} = j | X_m = i) = p_{ij}^{(n_0)} > 0$$

Inoltre, grazie alla subadditività:

$$0 < \mathbb{P} \left( \bigcup_{n \geq 1} (X_n = j | X_0 = i) \right) \leq \sum_{n=1}^{+\infty} \mathbb{P}(X_n = j | X_0 = i)$$

$$\implies \exists n_0 : \mathbb{P}(X_{n_0} = j | X_0 = i) = p_{ij}^{(n_0)} > 0 \quad \square$$

**Corollario 18.15: transitività**

$$i \rightarrow j \text{ e } j \rightarrow k \implies i \rightarrow k$$

*Esempio.* Riprendiamo l'esempio di pagina 223 delle due catene "triangolari". Nel primo esempio, da ogni stato ci si può spostare in ogni altro:  $i \rightarrow j \forall i, j \in E = \{a, b, c\}$ . Nel secondo caso invece ci sono due "sottosistemi" isolati; si ha dunque  $a \rightarrow b$  e  $b \rightarrow a$ , ma  $a \not\rightarrow c$  e  $b \not\rightarrow c$ .

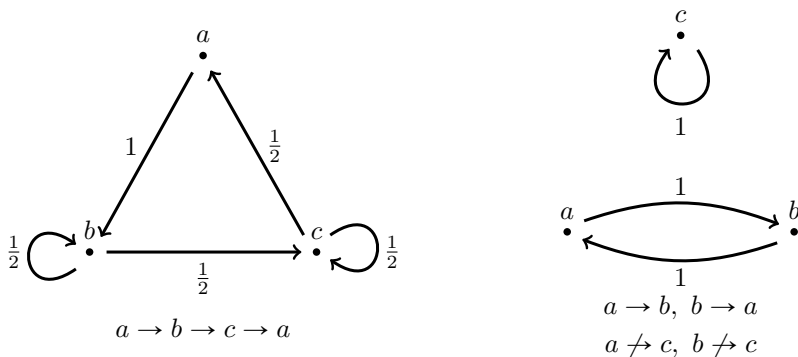


Figura 18.8: diagramma degli stati nei due esempi

**Definizione 18.16**

L'insieme  $C \subseteq E$  si dice **classe chiusa** se ogni  $i \in C$  non conduce a un qualsiasi  $j \notin C$ ; in altre parole,  $i \in C, i \rightarrow j \implies j \in C$ .

Le classi chiuse sono quelle cui finora ci siamo riferiti con il termine “sottosistema”. Nel primo dei due ultimi esempi, l'unica classe chiusa è ovviamente  $E$  stesso; nel secondo esempio, le classi chiuse sono gli insiemi  $\{a, b\}$  e  $\{c\}$ .

**Definizione 18.17**

Dati gli stati  $i, j \in E$ , si dice che  $i$  **comunica con**  $j$  se  $i \rightarrow j$  e  $j \rightarrow i$ , e ciò si rappresenta con la notazione  $i \leftrightarrow j$ .

**Definizione 18.18**

L'insieme  $C \subseteq E$  si dice **classe irriducibile** se ogni stato  $i \in C$  comunica con ogni altro stato  $j \in C$ , ovvero se non esistono sottoclassi chiuse in  $C$ .

In particolare, una catena di Markov  $(X_n)_{n \geq 0}$ , o equivalentemente la sua matrice di transizione  $P$ , si dice **irriducibile** se il suo spazio degli stati  $E$  è irriducibile.

**Definizione 18.19**

$i \in E$  è uno **stato assorbente** se l'insieme  $\{i\}$  è una classe chiusa, ovvero se  $p_{ii} = 1$ .

*Esempio:* catene irriducibili.

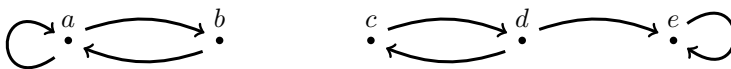
- Questa catena è irriducibile, con periodo 2 per ogni stato  $i \in \{a, b\}$ :



- Questa catena è irriducibile, con periodo 1 per ogni stato  $i \in \{a, b\}$ :



- Questa catena non è irriducibile:



Abbiamo infatti le classi chiuse  $E, \{a, b\}, \{c, d, e\}, \{e\}$ , le classi irriducibili  $\{a, b\}$  ed  $\{e\}$  e lo stato assorbente  $e$ .

#### 18.4.1 Primo arrivo in uno stato

##### Definizione 18.20

L'istante di primo arrivo nello stato  $j \in E$  è la VA così definita:

$$\tau_j : \Omega \rightarrow \mathbb{N} \cup \{+\infty\}, \quad \tau_j := \min\{n \geq 1 : X_n = j\}$$

**NB:**  $\tau_j = +\infty$  se  $X_n(\omega) \neq j \forall n \geq 1$ , ovvero se  $\{n \geq 1 : X_n = j\} = \emptyset$ . In tal caso, la catena non raggiungerà mai lo stato  $j$ .

Si può facilmente verificare che  $\tau_j$  sia effettivamente una VA: l'evento  $(\tau_j = m) = (X_1 \neq j, \dots, X_{m-1} \neq j, X_m = j) \in \mathcal{A}$  per ogni  $m \geq 0$ . Pertanto, anche l'evento  $(\bigcup_{m=1}^{+\infty} (\tau_j = m))^C$  è in  $\mathcal{A}$  per le proprietà delle  $\sigma$ -algebre. Possiamo dunque esprimere tutte le possibili controimmagini delle  $\tau_j$  nel seguente modo:

$$\begin{aligned} (\tau_j = +\infty) &= \left( \bigcup_{m=1}^{+\infty} (\tau_j = m) \right)^C = \bigcap_{m=1}^{+\infty} (\tau_j \neq m) = \bigcap_{n=1}^{+\infty} (X_n \neq j) \in \mathcal{A} \\ (\tau_j < +\infty) &= \bigcup_{m=1}^{+\infty} (\tau_j = m) = \bigcup_{n=1}^{+\infty} (X_n = j) \in \mathcal{A} \end{aligned}$$

Concludiamo che  $\tau_j$  rispetta la definizione di VA.

##### Definizione 18.21

Dati gli stati  $i, j \in E$ , la probabilità di arrivare in  $j$  partendo da  $i$  è:

$$\rho_{ij} := \mathbb{P}(\tau_j < +\infty | X_0 = i) = \mathbb{P}_i(\tau_j < +\infty)$$

In questo caso non è importante il tempo necessario per la transizione da  $i$  a  $j$ , ma semplicemente che questa transizione sia possibile o meno.

##### Proposizione 18.22

Ovviamente:

$$i \rightarrow j \iff \rho_{ij} > 0.$$

##### Definizione 18.23

$i \in E$  è uno **stato ricorrente** se  $\rho_{ii} = 1$ , o è uno **stato transitorio** se  $\rho_{ii} < 1$ .

Gli stati ricorrenti ricompariranno quasi certamente almeno una volta nel corso dell'evoluzione, mentre gli stati transitori potrebbero non ricomparire più.

##### Teorema 18.24

Sia  $E$  spazio degli stati *finito* ed  $i, j \in E$ . Allora:

1.  $i$  assorbente  $\implies i$  ricorrente;
2.  $i$  transitorio  $\iff$  esiste  $j$  tale che  $i \rightarrow j$  e  $j \not\rightarrow i$ <sup>66</sup>;
3.  $i \rightarrow j$ ,  $j$  transitorio  $\implies i$  transitorio;
4.  $i \rightarrow j$ ,  $i$  ricorrente  $\implies j$  ricorrente e  $j \rightarrow i$ ;
5.  $i \leftrightarrow j \implies i$  e  $j$  sono o entrambi ricorrenti con  $d(i) = d(j)$ , o entrambi transitori;
6.  $j$  transitorio  $\implies p_{ij}^{(n)} \xrightarrow{n} 0 \quad \forall i \in E$ ;
7. esiste almeno uno stato ricorrente  $i \in E$ ;
8.  $E$  irriducibile  $\implies$  tutti gli stati sono ricorrenti con lo stesso periodo;
9. vale la seguente partizione:  $E = E_T \cup C_1 \cup \dots \cup C_k$ , con  $E_T$  classe contenente tutti gli stati transitori di  $E$  e  $C_1, \dots, C_k$  classi chiuse, disgiunte e irriducibili. Partendo da una delle classi  $C_m$  si rimane al suo interno per sempre, e se si parte dalla classe  $E_T$  si finisce inevitabilmente in una delle classi  $C_m$ .

**Dimostrazione**

I punti 5, 6 e 9 sono lasciati al lettore come esercizio.

1.  $i$  assorbente  $\implies p_{ii} = 1 \implies \mathbb{P}_i(\tau_i = 1) = 1 \implies \mathbb{P}_i(\tau_i < +\infty) = 1$ , cioè  $i$  ricorrente.
2. (solo  $\iff$ ) Sia  $i \rightarrow j$  e  $j \not\rightarrow i$ . Sia  $m$  il più piccolo elemento di  $\mathbb{N}$  tale che  $p_{ij}^{(m)} \geq 0$ . Sia inoltre  $\{i_1, \dots, i_{m-1}\} \subseteq E$  un possibile percorso di transizione da  $i$  a  $j$ , ovvero tale che l'evento  $A = (X_0 = i, X_1 = i_1, \dots, X_{m-1} = i_{m-1}, X_m = j)$  abbia probabilità non nulla. Per costruzione  $i_k \neq j \quad \forall k = 1, \dots, m-1$ .

Calcoliamo ora la seguente probabilità:

$$\begin{aligned} \mathbb{P}_i(A, \tau_i < +\infty) &= \sum_{n=1}^{+\infty} \mathbb{P}_i(A, \tau_i = n) \\ &= \sum_{n=m+1}^{+\infty} \mathbb{P}_i(A, \tau_i = n) \quad (\text{i primi } m \text{ sono } 0 \text{ per def. di } A) \\ &= \sum_{n=1}^{+\infty} \mathbb{P}_i(A, \tau_i = n+m) \quad (\text{cambio di indice}) \end{aligned}$$

Studiamo i singoli addendi.  $(\tau_i = n+m) \subseteq (X_{n+m} = i)$  perché il secondo evento ( $X$  si trova nello stato  $i$ ) include il primo ( $X$  si trova nello stato  $i$  per

---

<sup>66</sup>È bene evidenziare che questa proprietà *non* vale nel caso in cui  $E$  sia numerabile.

la prima volta). Pertanto:

$$\begin{aligned}
 \mathbb{P}_i(A, \tau_i = m+n) &\leq \mathbb{P}_i(A, X_{n+m} = i) \\
 &= \mathbb{P}_i(X_{n+m} = i|A) \mathbb{P}_i(A) && \text{(def. di prob. cond.)} \\
 &= \mathbb{P}_i(X_{n+m} = i|X_n = j) \mathbb{P}_i(A) && \text{(per la pr. di Markov)} \\
 &= p_{ji}^{(n)} \mathbb{P}_i(A) \\
 &= 0 \mathbb{P}_i(A) = 0 && \text{(perché } j \not\rightarrow i)
 \end{aligned}$$

Quindi, l'intera sommatoria è nulla:  $\mathbb{P}_i(A, \tau_i < +\infty) = 0$ . Infine:

$$\begin{aligned}
 \rho_{ii} &= \mathbb{P}_i(\tau_i < +\infty) \\
 &= \mathbb{P}_i(\tau_i < +\infty, A) + \mathbb{P}_i(\tau_i < +\infty, A^C) \\
 &= 0 + \mathbb{P}_i(\tau_i < +\infty, A^C) && \text{(come appena dimostrato)} \\
 &\leq \mathbb{P}(A^C) && \text{(perché è un sottoevento)} \\
 &= 1 - \mathbb{P}(A) < 1 && \text{(per costruzione)}
 \end{aligned}$$

$i$  è dunque transitorio, come da tesi.

3. Sia  $i \rightarrow j$  con  $j$  transitorio. Per il punto 2 esiste  $k$  tale che  $i \rightarrow j$ ,  $j \rightarrow k$ ,  $k \not\rightarrow j$ . Pertanto  $i \rightarrow k$  e necessariamente  $k \not\rightarrow i$ , altrimenti si avrebbe l'assurdo  $k \rightarrow i \rightarrow j$ . Quindi anche  $i$  è transitorio grazie all'inverso del punto 2.
4. Sia  $i \rightarrow j$  con  $i$  ricorrente. Sia  $k$  tale che  $j \rightarrow k$ : dunque  $i \rightarrow j \rightarrow k$ ; ma  $i$  è ricorrente, quindi anche  $k \rightarrow i$ . Ne segue che  $k \rightarrow i \rightarrow j$ , e in particolare  $k \rightarrow j$ , che dimostra che  $j$  è ricorrente. Si noti che dev'essere  $j \rightarrow i$ , altrimenti  $i$  sarebbe transitorio.
5. La dimostrazione è una banale conseguenza del punto 3.
7. Sia  $E = E_T \cup E_R$ , con  $E_R = C_1 \cup \dots \cup C_k$ . Per il punto 6:

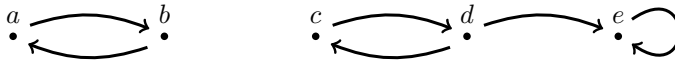
$$\mathbb{P}_i(X_n \in E_T) = \sum_{j \in E_T} \mathbb{P}_i(X_n = j) = \sum_{j \in E_T} p_{ij}^{(n)} \xrightarrow{n} 0$$

In quanto il numero di addendi è finito. Quindi  $\mathbb{P}_i(X_n \in E_R) = 1 - \mathbb{P}_i(X_n \in E_T) \xrightarrow{n} 1$ . Segue che  $E_R \neq \emptyset$ : esiste uno stato  $i$  ricorrente, come da tesi.

8. Sia  $E$  irriducibile. Per il punto 5, i suoi stati sono o tutti ricorrenti o tutti transitori, ma il punto 7 afferma che esiste almeno un  $i$  ricorrente: gli stati sono dunque tutti ricorrenti con lo stesso periodo.  $\square$

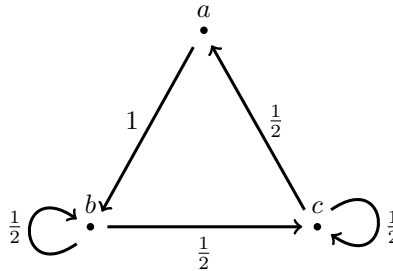
*Esempio:* stati transitori e ricorrenti.

- Nel primo caso abbiamo gli stati ricorrenti  $a, b, e$  e i transitori  $c, d$ . Inoltre,  $e$  è anche uno stato assorbente.



Il periodo è 1 per  $e$ , ed è 2 per ogni altro stato.

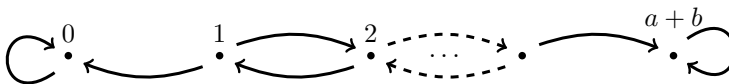
- In questo caso abbiamo una catena irriducibile con solo stati ricorrenti di periodo 1, pertanto la catena è aperiodica.



- Entrambe le catene sono irriducibili a stati ricorrenti, ma la prima è periodica di periodo  $d = 2$ , mentre la seconda è aperiodica.



- In questo esempio, la rovina del giocatore (già affrontato a pagina 226), la catena *non* è irriducibile e contiene quattro classi chiuse:  $E, \{0\}, \{a + b\}, \{0, a + b\}$ .



Gli stati transitori sono  $1, \dots, a + b - 1$ ; quelli ricorrenti  $0, a + b$ , ed essi sono anche assorbenti; i periodi sono  $d(i) = 2 \forall i = 1, \dots, a + b - 1$  e  $d(0) = d(a + b) = 1$ .

- Nell'urna di Ehrenfest (già affrontata a pagina 226) la catena è irriducibile, ricorrente, e periodica di periodo 2.



### 18.5 Problemi di assorbimento

Sia  $C \subseteq E$  una classe chiusa (irriducibile o meno<sup>67</sup>).

**Definizione 18.25**

$\tau_C : \Omega \rightarrow \mathbb{N} \cup \{+\infty\}$ ,  $\tau_C := \min\{n \geq 1 : X_n \in C\}$  è l'**istante di primo arrivo** nella classe  $C$ .

La definizione è del tutto simile a quella per l'istante di primo arrivo in un singolo stato, e si può verificare che anche in questo caso la funzione è una VA.

**Definizione 18.26**

La probabilità di arrivare in  $C$  partendo da  $i$ , o **probabilità di assorbimento** nella classe  $C$ , è definita come:

$$\lambda_i := \mathbb{P}_i(X_n \in C \text{ per almeno un } n \geq 1) = \mathbb{P}_i(\tau_C < +\infty)$$

Inoltre,  $\tau_C^i := \mathbb{E}_i[\tau_C]$  è il **tempo atteso di assorbimento** (o di arrivo) in  $C$  partendo da  $i$ , dove la notazione  $\mathbb{E}_i[\cdot]$  indica il valore atteso calcolato sulla legge condizionata  $\mathbb{P}_i(\cdot)$ .

*Esempio.* Riprendiamo nuovamente l'esempio della rovina del giocatore di pagina 226. Ricordiamo che  $X_n$  è il capitale di Aron al tempo  $n$ . Prendendo  $C = \{0\}$ ,  $\lambda_a$  è la probabilità di rovina di Aron. Prendendo invece  $C = \{0, a + b\}$ ,  $\lambda_a$  è la probabilità che il gioco finisca, o con la rovina di Aron o con quella di Bruno. In tal caso,  $\tau_C^a$  è la *durata attesa* del gioco.

**NB:** Sia  $(X_n)_{n \geq 0}$  Markov( $v, P$ ). Per le probabilità totali si ha:

$$\mathbb{P}(\tau_C < +\infty) = \sum_{i \in E} \mathbb{P}(\tau_C < +\infty | X_0 = i) \mathbb{P}(X_0 = i) = \sum_{i \in E} \lambda_i v_i$$

Per le regole dell'attesa condizionata (pagina 179) si ha inoltre:

$$\mathbb{E}[\tau_C] = \mathbb{E}[\mathbb{E}[\tau_C | X_0]] = \sum_{i \in E} \mathbb{E}[\tau_C | X_0 = i] \mathbb{P}(X_0 = i) = \sum_{i \in E} \tau_C^i v_i$$

**NB:**  $\lambda_i$  può assumere diversi valori:

- è 1 se  $i \in C$ ;
- è 0 se  $i \in \tilde{C}$ , con  $\tilde{C}$  classe chiusa e disgiunta da  $C$ ;

---

<sup>67</sup>Quindi  $C$  può anche essere l'unione di più classi chiuse irriducibili o contenere stati transitori.



- assume un altro valore, ignoto a priori, in  $(0, 1)$  se  $i \in D = E_T \setminus C$ .

Cerchiamo ora di calcolare i valori di  $\lambda_i$  in quest'ultimo caso.

**Proposizione 18.27**

I valori  $\lambda_i$ , con  $i$  che varia in  $D = E_T \setminus C$ , sono soluzioni del seguente sistema lineare:

$$\lambda_i = \sum_{k \in C} p_{ik} + \sum_{j \in D} p_{ij} \lambda_j \quad \forall i \in D$$

Si noti che se  $E$  è *finito*, questo sistema ammette *sempre* soluzione.

**Dimostrazione**

Grazie alle probabilità totali si ha che:

$$\lambda_i = \mathbb{P}_i(\tau_C < +\infty) = \sum_{j \in E} \mathbb{P}_i(\tau_C < +\infty | X_1 = j) \mathbb{P}_i(X_1 = j)$$

Suddividiamo ora la sommatoria in tre mediante una partizione:  $E = C \cup D \cup (C \cup E_T)^C$ . Grazie al punto 9 del teorema 18.24, sappiamo che l'ultimo di questi tre insiemi è interamente composto da classi ricorrenti disgiunte. Per quanto ricordato poco sopra, le probabilità condizionate nelle sommatorie sono unitarie se  $j \in C$  e nulle se  $j \in (C \cup E_T)^C$ ; inoltre, per definizione  $\mathbb{P}_i(X_1 = j) = p_{ij}$  in tutte e tre le sommatorie.

Ricapitolando, si ha dunque:

$$\lambda_i = \sum_{j \notin C \cup E_T} 0 \cdot p_{ij} + \sum_{j \in C} 1 \cdot p_{ij} + \sum_{j \in D} \mathbb{P}_i(\tau_C < +\infty | X_1 = j) p_{ij}$$

Svolgiamo i calcoli per il singolo addendo di probabilità condizionata:

$$\begin{aligned} \mathbb{P}_i(\tau_C < +\infty | X_1 = j) &= \sum_{n=1}^{+\infty} \mathbb{P}_i(\tau_C = n | X_1 = j) \\ &= \sum_{n=2}^{+\infty} \mathbb{P}_i(\tau_C = n | X_1 = j) && (j \notin C) \\ &= \sum_{n=2}^{+\infty} \mathbb{P}_i(X_2 \notin C, \dots, X_{n-1} \notin C, X_n \in C | X_1 = j) \\ &= \sum_{n=2}^{+\infty} \frac{\mathbb{P}_i(X_1 = j, X_2 \notin C, \dots, X_{n-1} \notin C, X_n \in C)}{\mathbb{P}_i(X_1 = j)} \\ &= \sum_{n=2}^{+\infty} \frac{\sum_{j_n \in C} \sum_{j_2, \dots, j_{n-1} \notin C} \cancel{p_{ij}} p_{jj_2} p_{j_2 j_3} \dots p_{j_{n-1} j_n}}{\cancel{p_{ij}}} && (\text{th. 18.7}) \\ &= \sum_{n=1}^{+\infty} \sum_{\substack{j_1, \dots, j_{n-1} \notin C \\ j_n \in C}} p_{jj_1} \dots p_{j_{n-1} j_n} = \lambda_j \end{aligned}$$

Negli ultimi passaggi riscaldiamo il tempo da  $n = 2$  a  $n = 1$  grazie alla proprietà di Markov; inoltre, sommando le probabilità di tutti i possibili percorsi per arrivare in  $C$  si ottiene la probabilità di assorbimento  $\lambda_j$ .

Inserendo quest'ultimo risultato nell'espressione precedente, si ha la tesi.  $\square$

Enunciamo ora l'esistenza di un sistema lineare simile per determinare i tempi attesi di assorbimento  $\tau_C^i$ .

**Proposizione 18.28**

Sia  $E = E_R \cup E_T$ , con  $C = E_R$  insieme degli stati ricorrenti ed  $E_T$  insieme degli stati transitori. Allora:

$$\tau_C^i < +\infty \quad \forall i \in E \quad \text{e} \quad \tau_C^i = 1 + \sum_{j \in E_T} p_{ij} \tau_C^j \quad \forall i \in E_T$$

**NB:** La soluzione di questo sistema *esiste ed è unica*.

*Esempio.* Riprendiamo ancora la rovina del giocatore di pagina 226.  $C = \{0\}$  è la classe in cui avviene la rovina di Aron, mentre  $\tau = \min\{n \geq 1 : X_n = 0\}$  è l'istante di rovina. Inoltre, avendo  $D = E_T \setminus C = E_T = \{1, \dots, a + b - 1\}$ , dalla proposizione 18.27 si ottiene:

$$\lambda_i = \mathbb{P}_i(\tau < +\infty) \quad \text{e} \quad \lambda_i = \sum_{k \in C} p_{ik} + \sum_{j \in D} p_{ij} \lambda_j \quad \forall i \in D$$

Il valore d'interesse è  $\lambda_a$ , la probabilità di rovina di Aron partendo da un capitale pari ad  $a$  (come nelle ipotesi del modello). Scrivendo il sistema per esteso si ha:

$$\begin{cases} \lambda_1 = p_{10} + p_{12} \lambda_2 = q + p \lambda_2 \\ \lambda_2 = p_{21} \lambda_1 + p_{23} \lambda_3 = q \lambda_1 + p \lambda_3 \\ \vdots \\ \lambda_{a+b-2} = q \lambda_{a+b-3} + p \lambda_{a+b-1} \\ \lambda_{a+b-1} = q \lambda_{a+b-2} \end{cases}$$

Risolviamo il sistema in alcuni casi esplicativi:

- Se  $p = q = \frac{1}{2}$ , si ottiene:

$$\lambda_i = \frac{a + b - i}{a + b} \quad \text{e} \quad \lambda_a = \frac{b}{a + b}$$

- Se  $\frac{q}{p} = \gamma \neq 1$ , si ottiene:

$$\lambda_i = \frac{\gamma^i - \gamma^{a+b}}{1 - \gamma^{a+b}} \quad \text{e} \quad \lambda_a = \frac{\gamma^a - \gamma^{a+b}}{1 - \gamma^{a+b}}$$

Per quanto riguarda il tempo medio di assorbimento, se  $\tau_{E_R}$  ha valore atteso finito, vuol dire che  $\tau_{E_R}$  stesso è finito quasi certamente.

Si prenda ora  $C = \{0, a + b\} = E_R$  (classe di fine del gioco) e  $\tau_C^i = \mathbb{E}_i[\tau_C]$  (durata attesa del gioco). Si imposta il sistema della proposizione 18.28:

$$\begin{cases} \tau_C^1 = 1 + p_{12} \tau_C^2 \\ \tau_C^2 = 1 + p_{21} \tau_C^1 + p_{23} \tau_C^3 \\ \vdots \\ \tau_C^{a+b-1} = 1 + p_{a+b-1, a+b-2} \tau_C^{a+b-2} \end{cases}$$

Fissati  $p = q = \frac{1}{2}$ , si ha  $\tau_C^i = i(a + b - i)$  e  $\tau_C^a = ab$ . Ponendo che i capitali iniziali siano  $a = 100$  e  $b = 1$ , si ottiene  $\tau_C^a = 100$ , che è la durata attesa del gioco in lanci.

## 18.6 Comportamento asintotico

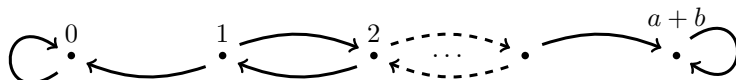
### Definizione 18.29

Data una matrice di transizione  $P$  su uno spazio degli stati  $E$ , una densità di probabilità  $\pi = (\pi_i)_{i \in E}$ , si dice **invariante** o **stazionaria** per  $P$  se:

$$\pi = \pi P.$$

Dalla definizione discende che data una catena di Markov  $(X_n)_{n \geq 0}$  con matrice di transizione  $P$  e legge iniziale  $v = \pi$ , allora  $X_n \sim v^{(n)} = \pi \forall n$ , ovvero la catena ha probabilità, appunto, *invariante* nel tempo.

*Esempio.* La rovina del giocatore (pagina 226) può essere così rappresentata:



In tal caso  $\pi = [\alpha \ 0 \ 0 \ \dots \ 0 \ 1 - \alpha]$  con  $\alpha \in [0, 1]$ : infatti, intuitivamente, gli stati transitori avranno probabilità invariante nulla. La questione verrà meglio sviscerata tra qualche pagina.

### Definizione 18.30

$P$  è una matrice di transizione **bistocastica** se  $\sum_{i \in E} p_{ij} = 1 \ \forall j \in E$ , ovvero se anche le colonne, oltre alle righe, sommano a 1.

In particolare:

### Proposizione 18.31

Per qualsiasi matrice di transizione bistocastica  $P$ , la distribuzione *uniforme*  $\pi = [\frac{1}{N} \ \dots \ \frac{1}{N}]$  con  $N = \#E$  è sempre una densità stazionaria.

**Dimostrazione**

$$(\pi P)_j = \sum_{i \in E} \pi_i p_{ij} = \frac{1}{N} \sum_{i \in E} p_{ij} = \frac{1}{N} = \pi_j \quad \forall j \in E \quad \square$$

*Esempio.* La matrice di transizione di questo esempio è  $P = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$  ed è bistocastica. Pertanto, la densità  $\pi = [\frac{1}{2} \ \frac{1}{2}]$  è stazionaria.

**Teorema 18.32: Markov-Kakutani**

Per ogni matrice di transizione  $P$  su uno spazio degli stati  $E$  finito, esiste  $\pi$  vettore di probabilità tale che  $\pi = \pi P$ , ovvero tale che sia *autovettore sinistro* di  $P$  (con autovalore 1).

Ricordiamo che  $\pi$  non può essere un autovettore qualsiasi ma deve rispettare i vincoli delle densità di probabilità, ovvero avere componenti non negative e di somma unitaria.

**Dimostrazione**

Sia  $v$  un generico vettore di probabilità. Si definisca il vettore:

$$v^n = \frac{1}{n} \sum_{j=0}^{n-1} v P^j$$

È evidente che esso abbia componenti non negative, e si può mostrare che:

$$\sum_{i \in E} v_i^n = \frac{1}{n} \sum_{j=0}^{n-1} \sum_{i \in E} (v P^j)_i = \frac{1}{n} \sum_{j=0}^{n-1} 1 = 1.$$

Dunque,  $v^n$  è una densità di probabilità. Più precisamente,  $\{v^n\}_{n \geq 0}$  è una successione di vettori di probabilità, cioè è contenuta nello spazio vettoriale  $S = \{\pi \in \mathbb{R}^N : \pi \geq 0, \sum_i \pi_i = 1\}$ , dove  $N = \#E < +\infty$ .  $S$  è un insieme chiuso e limitato e pertanto compatto.<sup>68</sup> Esiste dunque una sottosuccessione  $\{n_k\}_{k \geq 0}$

<sup>68</sup>Come ben noto, il teorema di Heine-Borel che garantisce questa equivalenza funziona solo negli spazi euclidei finito-dimensionali; da qui la richiesta di  $E$  finito.

tale che  $v^{n_k} \xrightarrow{k} u \in S$ . Calcoliamo ora la distanza  $u - uP$ :

$$\begin{aligned}
 u - uP &= \lim_{k \rightarrow +\infty} (v^{n_k} - v^{n_k}P) \\
 &= \lim_{k \rightarrow +\infty} \frac{1}{n_k} \left( \sum_{j=0}^{n_k-1} vP^j - \sum_{j=0}^{n_k-1} vP^{j+1} \right) \\
 &= \lim_{k \rightarrow +\infty} \frac{1}{n_k} (vP^0 - vP^{n_k-1+1}) \quad (\text{elisioni}) \\
 &= \lim_{k \rightarrow +\infty} \underbrace{\frac{1}{n_k}}_{\rightarrow 0} (v - vP^{n_k}) = 0 \quad ((v - vP^{n_k})_{k \geq 0} \subseteq S \text{ limitata})
 \end{aligned}$$

Pertanto  $u$  è una probabilità invariante, come da tesi.  $\square$

### Osservazione 18.33

Se  $\pi$  invariante non è unica, allora ne esistono infinite; infatti, date  $\pi_1$  e  $\pi_2$  probabilità invarianti e  $\alpha \in [0, 1]$ , la combinazione convessa  $\alpha\pi_1 + (1 - \alpha)\pi_2$  è ancora invariante.

#### 18.6.1 Ricerca dell'invariante

In primo luogo, la transitorietà o meno di uno stato fornisce una preziosa informazione sulla forma dell'invariante:

#### Proposizione 18.34

Sia  $\pi$  una probabilità invariante per  $P$ . Allora:

1.  $j$  transitorio  $\implies \pi_j = 0$ ;
2.  $E$  irriducibile (i.e. non ci sono stati transitori)  $\implies \pi_i > 0 \forall i$ .

#### Dimostrazione

1. Per definizione:

$$\pi_j = (\pi P)_j = (\pi P^2)_j = \dots = (\pi P^n)_j = \sum_i \pi_i p_{ij}^{(n)} \xrightarrow{n} 0$$

Considerando che tutti gli addendi tendono a 0 in quanto probabilità di stati transitori.

2. Poiché le componenti di  $\pi$  sono non negative e sommano a 1, esiste  $\pi_i > 0$  per un certo  $i \in E$ . Sia inoltre  $j \in E$ . Allora  $i \rightarrow j$ , essendo la catena irriducibile, e quindi esiste  $n \geq 0$  tale che  $p_{ij}^{(n)} > 0$ .

Di conseguenza, in quanto i singoli fattori degli addendi sono strettamente positivi, vale:

$$\pi_j = (\pi P^n)_j = \sum_k \pi_k p_{kj}^{(n)} \geq \pi_i p_{ij}^{(n)} > 0 \quad \square$$

*Esempio:* ancora la rovina del giocatore<sup>69</sup>. I vettori della forma  $[\alpha \ 0 \ \dots \ 0 \ 1 - \alpha]$  con  $\alpha \in [0, 1]$  sono tutte e sole le invarianti della catena per via del punto 1.

Cerchiamo ora dei metodi per calcolare questi invarianti, aggiungendo una sottoclasse per semplificare lo studio.

**Definizione 18.35**

Una densità di probabilità  $\pi$  si dice **reversibile** se soddisfa la cosiddetta **equazione del bilancio dettagliato**:

$$\pi_i p_{ij} = \pi_j p_{ji} \quad \forall i, j \in E$$

Questo equivale ad avere un'eguale probabilità di arrivare da  $i$  a  $j$  o viceversa.

**Teorema 18.36**

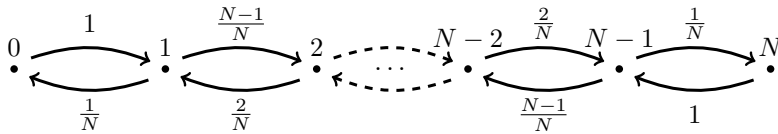
Se  $\pi$  è reversibile, allora è anche invariante.

**Dimostrazione**

Sfruttando la definizione di reversibilità:

$$(\pi P)_j = \sum_{i \in E} \pi_i p_{ij} = \sum_{i \in E} \pi_j p_{ji} = \pi_j \cdot \sum_{i \in E} p_{ji} = \pi_j \quad \forall j \in E \quad \square$$

*Esempio:* urna di Ehrenfest (pagina 226). Si scriva il sistema discendente dalla definizione di probabilità invariante:



$$\pi = \pi P \iff \begin{cases} \pi_0 = \pi_1 p_{10} = \pi_1 \cdot \frac{1}{N} \\ \pi_1 = \pi_0 \cdot 1 + \pi_2 \cdot \frac{2}{N} \\ \vdots \\ \pi_{N-1} = \pi_{N-2} \cdot \frac{2}{N} + \pi_N \cdot 1 \\ \pi_N = \pi_{N-1} \cdot \frac{1}{N} \end{cases} \iff \begin{cases} \pi_1 = N\pi_0 \\ \pi_2 = \frac{N-1}{2}\pi_1 \\ \pi_3 = \frac{N-2}{3}\pi_3 \\ \vdots \\ \pi_N = \frac{1}{N}\pi_{N_1} \end{cases}$$

La soluzione generale del sistema è  $\pi_i = \binom{N}{i} \pi_0 \ \forall i = 0, \dots, N$ . Imponiamo ora le opportune condizioni per un vettore di probabilità: da  $\pi_i \geq 0 \ \forall i$  segue  $\pi_0 > 0$ , e da  $\sum_i \pi_i = 1$  si ha:

$$\pi_0 \sum_i \binom{N}{i} = 1 \implies \pi_0 2^N = 1 \implies \pi_i = \binom{N}{i} 2^{-N}$$

<sup>69</sup>che ormai sapete a che pagina trovare

Questa è la definizione di una distribuzione binomiale! In particolare,  $\pi \sim Bi(N, \frac{1}{2})$ . Si può inoltre verificare che  $\pi$  è unica e reversibile.

Vediamo ora il caso in cui la distribuzione di partenza *non* sia invariante. Sia  $(X_n)_{n \geq 0}$  Markov( $v, P$ ). Esistono i limiti in legge e in probabilità di  $X_n$ ? In caso affermativo, quali sono? Vedremo che il limite in legge potrebbe non esistere.

*Esempio.* Siano  $P = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$  e  $v = [\alpha \ 1 - \alpha]$ , con  $\alpha \in [0, 1]$ .



Allora:

$$v^{(n)} = vP^{(n)} = \begin{cases} [\alpha \ 1 - \alpha] & \text{per } n \text{ pari} \\ [1 - \alpha \ \alpha] & \text{per } n \text{ dispari} \end{cases}$$

Pertanto  $v^{(n)}$  ha limite puntuale se e solo se  $\alpha = \frac{1}{2}$ , ovvero se la catena è stazionaria.

**Osservazione 18.37**

Si supponga che  $v^{(n)}$  ammetta limite:  $\lim_{n \rightarrow +\infty} v^{(n)} = v^{(\infty)}$ . Allora  $v^{(n)}$  è una probabilità invariante. Infatti:

$$v^{(\infty)}P = \lim_{n \rightarrow +\infty} v^{(n)}P = \lim_{n \rightarrow +\infty} v^{(n+1)} = v^{(\infty)}$$

**Definizione 18.38**

Data una matrice di transizione  $P$ , una densità di probabilità  $\pi$  si dice **distribuzione asintotica** di  $P$  se  $\pi$  è la sua unica distribuzione invariante e se:

$$v^{(n)} \xrightarrow{n} \pi \quad \forall v^{(0)} = v$$

**Teorema 18.39**

Sia  $(X_n)_{n \geq 0}$  Markov( $v, P$ ) *irriducibile*. Allora:

- a)  $\exists!$   $\pi$  invariante;
- b)  $\pi_i = \frac{1}{\mathbb{E}_i[\tau_i]} \quad \forall i \in E$ ;
- c) vale il **teorema ergodico**:

$$\frac{1}{n+1} \sum_{k=0}^n h(X_k) \xrightarrow{qc} \sum_{i \in E} h(i) \cdot \pi_i \quad \forall v, \forall h : E \rightarrow \mathbb{R}$$

c') un importante caso particolare del punto precedente, prendendo  $h = \mathbb{1}_{\{i\}}$ , è:

$$\frac{1}{n+1} \sum_{k=0}^n \mathbb{1}_{\{i\}}(X_k) \xrightarrow{\text{qc}} \pi_i \quad \forall v, \forall i \in E$$

d) se la catena è anche aperiodica, allora  $p_{ij}^{(n)} \xrightarrow{n} \pi_j \quad \forall i, j \in E$ ;

e) esiste limite in legge:  $v^{(n)} \xrightarrow{\mathcal{L}} \pi \quad \forall v$ .

Il teorema ergodico del punto (c) effettua la composizione della catena mediante una generica funzione  $h$  e definendo le nuove VA  $Y_k = h(X_k)$ , che in generale non saranno iid, proprio come le  $X_n$ .

La media campionaria di queste  $Y_k$ , detta *media orizzontale*, tende al valore atteso (rispetto a  $\pi$ ) della funzione stessa, calcolata nella media campionaria delle  $X_k$ , la quale viene chiamata *media verticale* e ha appunto distribuzione  $\pi$ . Il teorema afferma dunque che asintoticamente la media verticale è una buona approssimazione della media orizzontale.

Il caso particolare (c') afferma che la frazione di tempo in cui una catena è in un certo stato  $i$  tende proprio al valore  $\pi_i$ . Questo illustra un nuovo significato della distribuzione invariante, ovvero quello di essere una *frequenza media asintotica* di passaggio nei vari stati. Questa intuizione è ulteriormente confermata dal punto (b) del teorema, che spiega che tale frequenza media  $\pi_i$  è il reciproco dell'*attesa media*  $\mathbb{E}_i[\tau_i]$  di primo ingresso in  $i$ .

*Esempio.* Prendiamo l'ormai familiare esempio  $P = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ .



Sappiamo che le potenze pari di  $P$  sono uguali all'identità e quelle dispari sono uguali a  $P$  stessa. Sia  $v = \delta_a$ . Si ha dunque  $v^{(n)} = \delta_a$  per  $n$  pari e  $v^{(n)} = \delta_b$  per  $n$  dispari. Non esiste limite in legge in questo caso!

Poiché  $\tau_a \sim \delta_2$ , si ha che  $\mathbb{E}_a[\tau_a] = 2$ . Sappiamo che  $\pi = [\frac{1}{2} \quad \frac{1}{2}]$ , che è coerente con il risultato (b), e che la sua unicità è garantita dal punto (a). Grazie al teorema ergodico si può inoltre affermare che:

$$\frac{1}{n+1} \sum_{k=0}^n \mathbb{1}_{\{a\}}(X_k) \xrightarrow{\text{qc}} \frac{1}{2} = \pi_a \quad \forall v$$

*Esempio.* Riprendiamo l'esempio dell'urna di Ehrenfest di pagina 226. A pagina 246 abbiamo invece appurato che l'invariante ha distribuzione binomiale  $Bi(N, \frac{1}{2})$ .



La catena è irriducibile e aperiodica, e infatti ammette invariante unico. Ancora, dal punto (b) si deduce che:

$$\mathbb{E}_i[\tau_i] = \frac{2^N}{\binom{N}{i}} \quad \forall i = 0, \dots, N$$

Si scelga come  $h$  la funzione identità:  $h : E \rightarrow \mathbb{R}, h(i) = i$ ; per il teorema ergodico si ha dunque:

$$\frac{1}{n+1} \sum_{k=0}^n X_k \xrightarrow[n]{\text{qc}} \sum_{i=0}^N i \binom{N}{i} 2^{-N} = \frac{N}{2} \quad (\text{media di } Bi(N, \frac{1}{2}))$$

Usando invece le indicatrici in 0:

$$\frac{1}{n+1} \sum_{k=0}^n \mathbf{1}_{\{0\}}(X_k) \xrightarrow{\text{qc}} \pi_0 = \frac{1}{2^N}$$

Quest'ultima formula, nell'esempio del gas in due urne comunicanti, indica che l'urna  $A$  rimarrà vuota mediamente per  $2^N$  istanti di tempo.

### 18.6.2 Criteri di irriducibilità e aperiodicità

#### Definizione 18.40

Una matrice di transizione  $P$  è detta **regolare** se esiste  $n \geq 0$  tale che  $p_{ij}^{(n)} > 0 \quad \forall i, j \in E$ .

#### Proposizione 18.41

1.  $P$  regolare  $\implies P$  irriducibile e aperiodica;
2. se  $P$  è irriducibile ed esiste uno stato  $i$  tale che  $p_{ii} > 0$ , allora  $P$  è regolare e aperiodica.

*Esempio.* Sia  $P = \begin{bmatrix} 0 & 1 \\ 1-\alpha & \alpha \end{bmatrix}$  con  $\alpha \in (0, 1)$ .



La catena è irriducibile e aperiodica. Bisogna ora cercare il vettore di probabilità  $\pi = \pi P$ . Impostiamo il sistema (che avrà tre equazioni e due incognite, quindi con un'equazione superflua) con l'aggiunta della condizione  $\pi_i \geq 0$ :

$$\begin{cases} (1-\alpha)\pi_2 = \pi_1 \\ \pi_2 + \pi_1 = 1 \\ \pi_2 = \pi_1 + \alpha\pi_2 \end{cases} \implies \pi = \begin{bmatrix} 1-\alpha & 1 \\ 2-\alpha & 2-\alpha \end{bmatrix}$$

Si ha dunque, per il teorema ergodico:

$$\frac{1}{n+1} \sum_{k=0}^n \mathbb{1}_{\{1\}}(X_k) \xrightarrow{\text{qc}} \pi_1 = \frac{1-\alpha}{2-\alpha} \stackrel{\text{(b)}}{\implies} \mathbb{E}_1[\tau_1] = \frac{2-\alpha}{1-\alpha}$$

Abbiamo quindi convergenza in legge:

$$X_n \xrightarrow[n]{\mathcal{L}} \pi \quad \text{cioè} \quad \mathbb{P}(X_n = 1) \simeq \frac{1-\alpha}{2-\alpha} \quad \text{e} \quad \mathbb{P}(X_n = 2) \simeq \frac{1}{2-\alpha} \quad \text{per } n \text{ grande}$$

Dal fatto che  $\tau_1 < +\infty$  quasi certamente discende che:

$$\mathbb{E}_1[\tau_1] = \sum_{n=1}^{+\infty} n \mathbb{P}_1(\tau_1 = n) = \sum_{n=2}^{+\infty} n \alpha^{n-2} (1-\alpha)$$

Anzitutto, il termine per  $n = 1$  è nullo perché il primo passo è sempre nello stato 2. Inoltre, affinché da 1 si torni in 1 per la prima volta in esattamente  $n$  passi, si deve avere un primo passaggio da 1 a 2 (che ha probabilità 1), poi bisogna rimanere in 2 per  $(n-2)$  passi (ciascuno dei quali ha probabilità  $\alpha$ ), e poi tornare in 1 (probabilità  $1-\alpha$ ); si effettua poi il prodotto di queste probabilità grazie al teorema 18.5 sulle leggi congiunte.

Si vuole ora sfruttare la seguente serie notevole, che converge se  $|\alpha| < 1$ :

$$\sum_{n=1}^{+\infty} n \alpha^{n-1} = \frac{1}{(1-\alpha)^2}$$

Per farlo, aggiungiamo e togliamo il termine per  $n = 1$  per ottenere la serie completa:

$$\begin{aligned} \mathbb{E}_1[\tau_1] &= \sum_{n=1}^{+\infty} n \alpha^{n-2} (1-\alpha) - \frac{1-\alpha}{\alpha} \\ &= \frac{(1-\alpha)}{\alpha} \sum_{n=1}^{+\infty} n \alpha^{n-1} - \frac{1-\alpha}{\alpha} \quad (\text{moltiplicando e dividendo per } \alpha) \\ &= \frac{(1-\alpha)}{\alpha(1-\alpha)^2} - \frac{1-\alpha}{\alpha} \quad (\text{sfruttando la serie notevole}) \\ &= \frac{2-\alpha}{1-\alpha} \quad (\text{svolgendo i conti}) \end{aligned}$$

Il risultato di (b) è così confermato anche dal computo manuale del valore atteso.

---

## Appendice A. Dispense integrative

### A.1 Richiami di analisi e insiemistica

#### A.1.1 Limite inferiore e superiore

##### Definizione A.1

Data una successione  $\{x_n\}_{n \in \mathbb{N}} \subseteq \mathbb{R}$ , sono definiti il **limite inferiore** e il **limite superiore** di tale successione:

$$\liminf x_n := \sup_{n \in \mathbb{N}} \inf_{k \geq n} x_k = \lim_{n \rightarrow +\infty} \inf_{k \geq n} x_k$$

$$\limsup x_n := \inf_{n \in \mathbb{N}} \sup_{k \geq n} x_k = \lim_{n \rightarrow +\infty} \sup_{k \geq n} x_k$$

L'intervallo  $[\liminf x_n, \limsup x_n]$  è detto **classe limite** della successione.

Una successione ammette limite se e solo se  $\liminf x_n = \limsup x_n$ , ovvero se la classe limite è composta da un solo elemento. In particolare vale il seguente teorema.

##### Teorema A.2

Per qualsiasi successione  $\{x_n\}_{n \in \mathbb{N}}$ :

1. il limite inferiore e il limite superiore sono entrambi *unic*;
2.  $\liminf x_n \leq \limsup x_n$ ;
3.  $x_n$  converge o diverge  $\iff \liminf x_n = \limsup x_n$ .

*Esempio.*

- $x_n = (-1)^n$  ha  $\liminf x_n = -1$  e  $\limsup x_n = 1$ , pertanto non ammette limite.
- $x_n = \begin{cases} 1 & \text{se } n \text{ è primo} \\ 0 & \text{altrimenti} \end{cases}$  ha  $\liminf x_n = 0$  e  $\limsup x_n = 1$ , pertanto non ammette limite.
- $x_n = (-n)^n$  ha  $\liminf x_n = -\infty$  e  $\limsup x_n = +\infty$ , pertanto non ammette limite.
- $x_n = (-\frac{1}{n})^n$  ha  $\liminf x_n = \limsup x_n = 0$  e di conseguenza  $\lim_n x_n = 0$ .

### A.1.2 Operazioni insiemistiche

#### Teorema A.3: leggi di De Morgan

Sia  $I$  un insieme di indici, anche non numerabile, e sia  $\{A_\alpha\}_{\alpha \in I} \subseteq \Omega$  una collezione di insiemi. Allora:

$$1. \left( \bigcup_{\alpha \in I} A_\alpha \right)^C = \bigcap_{\alpha \in I} A_\alpha^C \quad 2. \left( \bigcap_{\alpha \in I} A_\alpha \right)^C = \bigcup_{\alpha \in I} A_\alpha^C$$

#### Dimostrazione

Per la prima uguaglianza:

$$\begin{aligned} \omega \in \left( \bigcup_{\alpha \in I} A_\alpha \right)^C &\iff \omega \notin \bigcup_{\alpha \in I} A_\alpha \\ &\iff \omega \notin A_\alpha \quad \forall \alpha \in I \\ &\iff \omega \in A_\alpha^C \quad \forall \alpha \in I \quad (\text{Complementarità}) \\ &\iff \omega \in \bigcap_{\alpha} A_\alpha^C \end{aligned}$$

Per la seconda uguaglianza si procede in maniera analoga. □

Si esploreranno ora i concetti di immagine e controimmagine.

#### Definizione A.4

Data una funzione  $X : \Omega \rightarrow F$  e un insieme  $A \subseteq \Omega$ , si dice **immagine** di  $A$  l'insieme  $X(A) := \{x \in F : x = X(\omega), \text{ per qualche } \omega \in A\} \subseteq F$ .

#### Definizione A.5

Data una funzione  $X : \Omega \rightarrow F$  e un insieme  $B \subseteq F$ , si dice **controimmagine** di  $B$  l'insieme  $X^{-1}(B) = (X \in B) := \{\omega \in \Omega : X(\omega) \in B\} \subseteq \Omega$ .

#### Teorema A.6

Data una funzione  $X : \Omega \rightarrow F$ , la controimmagine ha le seguenti proprietà:

1.  $X(X^{-1}(B)) \subseteq B \quad \forall B \subseteq F$ ;
2.  $X^{-1}(X(A)) \supseteq A \quad \forall A \subseteq \Omega$ ;
3.  $X^{-1}(B^C) = (X^{-1}(B))^C \quad \forall B \subseteq F$ ;
4.  $X^{-1}(\bigcup_{\alpha} B_\alpha) = \bigcup_{\alpha} X^{-1}(B_\alpha) \quad \forall \{B_\alpha\}_{\alpha \in I} \subseteq F$ ;
5.  $X^{-1}(\bigcap_{\alpha} B_\alpha) = \bigcap_{\alpha} X^{-1}(B_\alpha) \quad \forall \{B_\alpha\}_{\alpha \in I} \subseteq F$ .

#### Teorema A.7: convergenza puntuale

Sia  $X_n : \Omega \rightarrow \mathbb{R}$  una successione di funzioni, con  $n \in \mathbb{N}$ . Allora:

1.  $(\liminf X_n)(\omega) = \liminf (X_n(\omega))$ ;
2.  $(\limsup X_n)(\omega) = \limsup (X_n(\omega))$ ;
3. Se per ogni  $\omega \in \Omega$  fissato  $X_n(\omega)$  converge per  $n \rightarrow +\infty$ , allora  $(\lim_n X_n)(\omega) = \lim_n (X_n(\omega))$ .

I membri sinistri delle 3 uguaglianze sono da intendersi come una singola funzione, nella quale  $n$  è solo un indice muto, valutata nel punto  $\omega$ , mentre i membri destri come i limiti per  $n \rightarrow +\infty$  di una successione numerica (cioè contenuta in  $\mathbb{R}$ ) il cui termine generale  $X_n(\omega)$  dipende da  $n$ .

### A.1.3 Limite inferiore e superiore per insiemi

Similmente a quanto fatto per le successioni numeriche, si possono definire il limite inferiore e superiore anche per una successione di insiemi.

#### Definizione A.8

Data una successione di insiemi  $\{A_n\}_{n \in \mathbb{N}} \subseteq \Omega$ , sono definiti il **limite inferiore** e il **limite superiore** di tale successione:

$$\limsup A_n := \bigcap_{n \in \mathbb{N}} \bigcup_{k \geq n} A_k \quad \text{e} \quad \liminf A_n := \bigcup_{n \in \mathbb{N}} \bigcap_{k \geq n} A_k$$

#### Proposizione A.9

Il limite inferiore e superiore hanno un'ulteriore interpretazione insiemistica:

1.  $\liminf A_n = \{\omega \in \Omega : \omega \in A_n \text{ definitivamente } (\forall n \geq \bar{n})\}$
2.  $\limsup A_n = \{\omega \in \Omega : \omega \in A_n \text{ per infiniti } A_n\}$

#### Dimostrazione

Per dimostrare il primo punto:

$$\begin{aligned} \omega \in \liminf A_n &\iff \omega \in \bigcap_{n \in \mathbb{N}} \bigcup_{k \geq n} A_k \\ &\iff \exists \bar{n} \in \mathbb{N} : \omega \in \bigcap_{k \geq \bar{n}} A_k \\ &\iff \exists \bar{n} \in \mathbb{N} : \omega \in A_k \quad \forall k \geq \bar{n} \end{aligned}$$

Per dimostrare il secondo punto:

$$\begin{aligned}
 \omega \in \limsup A_n &\iff \omega \in \bigcap_{n \in \mathbb{N}} \bigcup_{k \geq n} A_k \\
 &\iff \omega \in \bigcup_{k \geq n} A_k \quad \forall n \in \mathbb{N} \\
 &\iff \forall n \in \mathbb{N} \exists k \geq n : \omega \in A_k \\
 &\iff \exists k_j : \omega \in A_{k_j} \quad \forall j \in \mathbb{N} \\
 &\iff \omega \in A_n \text{ per infiniti } A_n \quad \square
 \end{aligned}$$

#### A.1.4 Funzione indicatrice

##### Definizione A.10

Dato un insieme  $A \in \Omega$ , la sua **funzione indicatrice**  $\mathbb{1}_A : \Omega \rightarrow \mathbb{R}$  è definita come segue:

$$\mathbb{1}_A(\omega) = \begin{cases} 1 & \omega \in A \\ 0 & \omega \notin A \end{cases}$$

##### Proposizione A.11

Siano  $A_1, \dots, A_n \subseteq \Omega$ . Allora:

$$\mathbb{1}_{\{\bigcap_{k=1}^n A_k\}} = \prod_{k=1}^n \mathbb{1}_{A_k}$$

##### Dimostrazione

$X = Y$  significa che  $X(\omega) = Y(\omega) \quad \forall \omega \in \Omega$ . Visto che  $X$  e  $Y$  possono assumere solo i valori 0 e 1, è sufficiente dimostrare che, fissato  $\omega \in \Omega$ , si ha che:

$$\mathbb{1}_{\bigcap_k A_k}(\omega) = 1 \iff \prod_{k=1}^n \mathbb{1}_{A_k}(\omega) = 1$$

Ma ciò è vero in quanto:

$$\begin{aligned}
 \mathbb{1}_{\{\bigcap_{k=1}^n A_k\}}(\omega) = 1 &\iff \omega \in \bigcap_{k=1}^n A_k \\
 &\iff \omega \in A_k \quad \forall k = 1, \dots, n \\
 &\iff \mathbb{1}_{A_k}(\omega) = 1 \quad \forall k = 1, \dots, n \\
 &\iff \prod_{k=1}^n \mathbb{1}_{A_k}(\omega) = 1 \quad \square
 \end{aligned}$$

##### Proposizione A.12

Siano  $A, B \subseteq \Omega$ . Allora  $\mathbb{1}_{\{A \cup B\}} = \mathbb{1}_A + \mathbb{1}_B - \mathbb{1}_A \cdot \mathbb{1}_B$ .

## A.2 Calcolo combinatorio

### A.2.1 Principi del calcolo combinatorio

Siano  $A$  e  $B$  due insiemi. Se  $\#A = \#B$  allora esiste una funzione biettiva  $f$  tale che  $f : A \rightarrow B$ .

Si ricordi inoltre che:

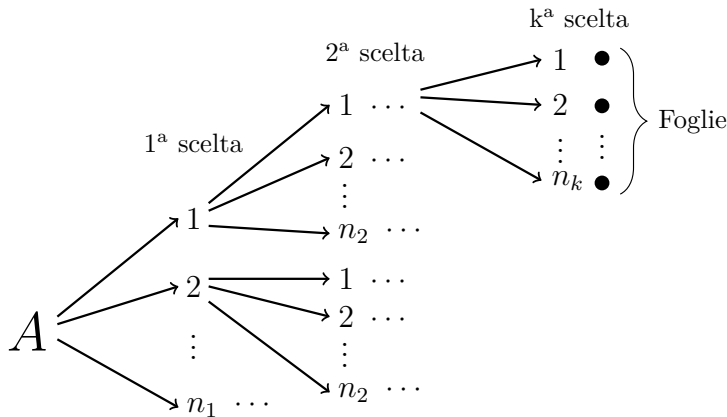
- se  $A \cap B = \emptyset$ , allora  $A$  e  $B$  sono detti disgiunti e  $\#(A \cup B) = \#A + \#B$ ;
- se  $A \cap B \neq \emptyset$ , allora  $\#(A \cup B) = \#A + \#B - \#(A \cap B)$ .

#### Definizione A.13: Principio fondamentale del calcolo combinatorio

Sia  $A$  un insieme finito. Si supponga che ogni elemento di  $A$  possa essere determinato mediante una sequenza di  $k$  scelte successive, in modo che la prima scelta è tra  $n_1$  possibilità, la seconda tra  $n_2$ , e in generale la  $i$ -esima tra  $n_i$ , dove  $n_1, \dots, n_k$  sono numeri naturali. Se a ogni sequenza di scelte corrisponde uno e un solo elemento di  $A$ , allora vale anche il viceversa, ovvero a ogni elemento di  $A$  corrisponde una e una sola sequenza di scelte. In tal caso:  $\#A = n_1 \cdot n_2 \cdot \dots \cdot n_k$ .

Questo principio viene anche detto *principio di enumerazione* o *schema delle scelte successive*.

Tale schema di ragionamento equivale graficamente ad un diagramma ad albero:



Si osservi che a prescindere da quale freccia è stata percorsa alla 1ª scelta, la 2ª scelta avrà sempre  $n_2$  possibilità, e così via per tutte le altre scelte.

Si riportano alcuni risultati notevoli deducibili dal principio appena riportato.

## A.2.2 Permutazioni

### Definizione A.14

Sia dato un insieme  $A$  di  $n$  oggetti distinti. Ogni possibile ordinamento di questo insieme prende il nome di **permutazione**.

Le *permutazioni semplici* di  $n$  elementi distinti, che differiscono per l'ordine, sono:

$$P_n = n! = n \cdot (n - 1) \cdot (n - 2) \cdots 2 \cdot 1$$

Le *permutazioni con ripetizione* di  $n$  elementi, dove un elemento è ripetuto  $h_1$  volte, un altro  $h_2$  volte, un terzo  $h_3$  volte, e così via, sono:

$$P_n^{h_1, h_2, h_3, \dots} = \frac{n!}{h_1 \cdot h_2 \cdot h_3 \dots}$$

## A.2.3 Combinazioni

### Definizione A.15

Sia dato un insieme  $A$  di  $n$  oggetti distinti. Ogni sottoinsieme di  $A$  costituito da  $k$  elementi, con  $0 \leq k \leq n$ , è detto **combinazione semplice** di  $n$  oggetti di classe  $k$ .

Le combinazioni semplici si esprimono attraverso il coefficiente binomiale:

$$C_{n,k} = \binom{n}{k} = \frac{n!}{k!(n-k)!}$$

### Definizione A.16

Sia dato un insieme  $A$  di  $n$  oggetti distinti. Ogni insieme di  $k$  elementi di  $A$  costituito anche da oggetti ripetuti è detto **combinazione con ripetizione** di  $n$  oggetti di classe  $k$ .

Le combinazioni con ripetizione sono:

$$C_{n,k}^* = \binom{n+k-1}{k}$$

Si osservi che le combinazioni non si distinguono per l'ordinamento.

È utile la conoscenza della **formula di Stiefel**:

$$\binom{n}{k} = \binom{n-1}{k} + \binom{n-1}{k-1}, \quad \text{oppure} \quad \binom{n+1}{k} = \binom{n}{k-1} + \binom{n}{k}$$

La sua dimostrazione viene lasciata al lettore come esercizio.



### A.2.4 Disposizioni

#### Definizione A.17

Sia dato un insieme  $A$  di  $n$  oggetti distinti. Ogni possibile sottoinsieme ordinato dell'insieme  $A$  di  $k$  elementi prende il nome di **disposizioni semplici**.

Dunque due disposizioni possono differire o per gli oggetti che contengono, o per l'ordinamento degli stessi. Le disposizioni semplici sono:

$$D_{n,k} = n \cdot (n-1) \cdot (n-2) \dots (n-k+1) = C_{n,k} \cdot P_k$$

#### Definizione A.18

Sia dato un insieme  $A$  di  $n$  oggetti distinti. Ogni possibile insieme ordinato formato da  $m$  oggetti anche ripetuti dell'insieme  $A$  si dice **disposizione con ripetizione** di  $n$  oggetti di classe  $k$ .

Dunque  $m$  può essere sia maggiore che minore che uguale a  $n$ . Le disposizioni con ripetizione sono:

$$D_{n,k}^* = n^m$$

## A.3 Integrazione complessa

La teoria dell'integrazione per  $h : (\mathbb{R}^n, \mathcal{B}^n, \mathbb{P}) \rightarrow (\mathbb{R}, \mathcal{B})$  si estende tale e quale per  $h : (\mathbb{R}^n, \mathcal{B}^n, \mathbb{P}) \rightarrow (\mathbb{C}, \mathcal{B})$ . Si può infatti considerare l'unità immaginaria  $i$  come una costante e scrivere la funzione complessa in forma algebrica:

$$h(x) = h_1(x) + i \cdot h_2(x), \quad \text{con } h_1, h_2 : \mathbb{R}^n \rightarrow \mathbb{R}$$

Ora è possibile integrare  $h$  nel seguente modo:

$$\int_{\mathbb{R}^n} h(x) \, d\mathbb{P} = \int_{\mathbb{R}^n} h_1(x) \, d\mathbb{P} + i \int_{\mathbb{R}^n} h_2(x) \, d\mathbb{P}$$

In corsi successivi tale teoria verrà propriamente dimostrata nel dettaglio.<sup>70</sup>

Queste proprietà sono utili quando si utilizza la funzione caratteristica, che è, di fatto, un valore atteso (ovvero un integrale) ad argomento complesso.

## A.4 Logaritmo complesso

Illustriamo ora alcune proprietà del logaritmo complesso, utile strumento per la dimostrazione del teorema centrale del limite (17.3).

<sup>70</sup>E invece era solo una spudorata menzogna

L'equazione  $e^w = z$ , con  $z \neq 0$ , ammette infinite soluzioni se l'incognita  $w$  appartiene a  $\mathbb{C}$ . Infatti, scrivendo  $z$  in forma polare, si ottiene:

$$\begin{aligned} e^w = z &= |z|e^{i\vartheta} && \text{con } \vartheta = \arg z \in [0, 2\pi] \\ &= |z| e^{i(\vartheta+2k\pi)} && \text{(angoli uguali a meno di giri)} \\ \implies w = \log z &= \ln|z| + i \cdot \arg z + i \cdot 2k\pi && \text{con } k \in \mathbb{Z} \end{aligned}$$

Nell'ultimo passaggio abbiamo sfruttato la proprietà del logaritmo di un prodotto. Si noti che con  $\log$  indichiamo il logaritmo "complesso", mentre con  $\ln$  indichiamo il normale logaritmo reale.

Per convenzione si sceglie la soluzione con  $k = 0$ ; si ottiene così la funzione chiamata **logaritmo principale**:

$$\log : \mathbb{C} \setminus \{0\} \rightarrow \mathbb{C}, \quad z \mapsto \log z = \ln|z| + i \arg z, \quad \text{con } -\pi \leq \arg z \leq \pi$$

Dunque, definendo l'insieme  $D = \{z \in \mathbb{C} : |z| < 1\}$ , la funzione  $f : D \rightarrow \mathbb{C}$ ,  $f(z) = \log(1+z)$  è sempre ben definita. La particolarità di questa funzione, che non sarà dimostrata in questo testo, è che ha lo stesso sviluppo di Taylor della sua controparte reale; quindi per  $x_0 = 0$  si ha che:

$$\log(1+z) = z - \frac{1}{2}z^2 + \frac{1}{3}z^3 + \dots$$

**NB:** Anche nel caso complesso l'asintotico del logaritmo rimane invariato; infatti, sfruttando lo sviluppo di Taylor appena visto:

$$\frac{\log(1+z)}{z} = \frac{z + o(z)}{z} = 1 + o(1) \xrightarrow{z \rightarrow 0} 1$$

## A.5 Radice quadrata di una matrice

Per **radice quadrata di una matrice**  $C$  si intende una matrice  $\sqrt{C}$  tale per cui:

$$\sqrt{C} \cdot \sqrt{C} = C$$

La trattazione sarà sviluppata solo per uno specifico tipo di matrice quadrata, in quanto il resto esula dalle necessità di questa trattazione; ma è bene tener presente che, in generale, una matrice ammette più radici quadrate.

Sia  $C \in \mathbb{R}^{n \times n}$  una matrice simmetrica e semi-definita positiva. Essa è diagonalizzabile e ha tutti gli autovalori reali<sup>71</sup> e non negativi:

$$C = C^T \text{ e } C \geq 0 \iff \lambda_k \geq 0$$

---

<sup>71</sup>«Teorema della diagonalizzazione, uno dei più importanti di tutta l'umanità»

Questo implica che  $\exists U$  ortogonale (ovvero tale che  $U^T = U^{-1}$ ) che verifica la seguente relazione:

$$UCU^T = \Lambda = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix} \quad \text{con } \lambda_k \in \mathbb{R} \quad \forall k$$

La radice della matrice  $C$  in questo caso risulta essere nella seguente forma:

$$\sqrt{C} = U^T \begin{bmatrix} \sqrt{\lambda_1} & & \\ & \ddots & \\ & & \sqrt{\lambda_n} \end{bmatrix} U$$

La verifica della definizione è immediata:

$$\begin{aligned} \sqrt{C} \cdot \sqrt{C} &= U\sqrt{\Lambda}U^T \cdot U\sqrt{\Lambda}U^T \\ &= U\sqrt{\Lambda}\sqrt{\Lambda}U^T && (U^T U = I) \\ &= U\Lambda U^T && (\text{prop. matrici diagonali}) \\ &= C \end{aligned}$$

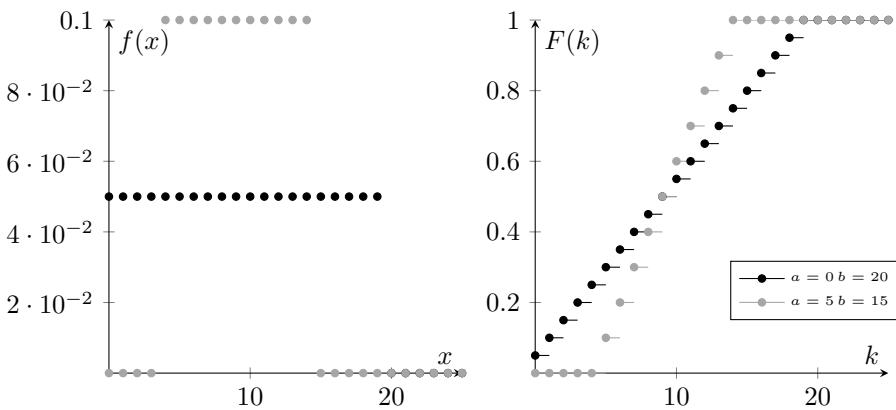
## Appendice B. Distribuzioni discrete

Riportiamo qui una lista di alcune distribuzioni discrete notevoli corredate da grafici e formule.

### B.1 Uniforme discreta

$$U_{disc}(a, b)$$

Chiamata così<sup>72</sup> in quanto distribuisce la probabilità in ugual quantità su  $n$  elementi, ovvero ciascuno di essi ha la stessa probabilità di realizzarsi.



**Supporto:**  $\{a, \dots, a + \frac{k-1}{n-1}(b-a), \dots, b\}$

**Funzione di densità:**  $\mathbb{P}(X = x) = \frac{1}{n} \quad \forall x$

**Funzione di ripartizione:**  $F(k) = \frac{k}{n}$  per  $a + \frac{k-1}{n-1}(b-a)$

**Valore atteso:**  $\mathbb{E}[X] = \frac{a+b}{2}$

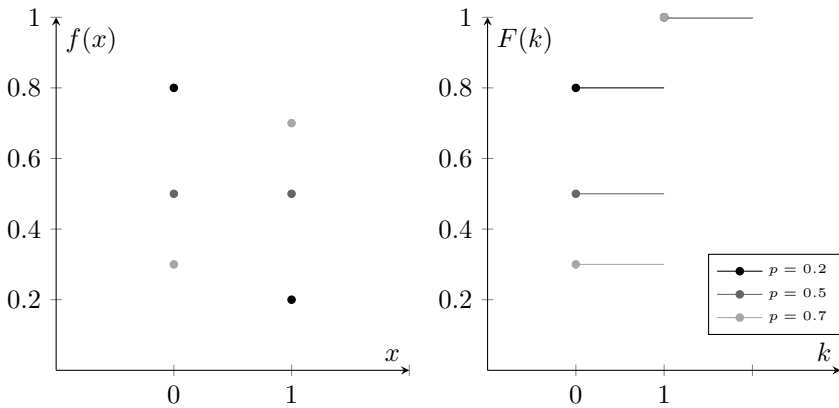
**Varianza:**  $Var(X) = \frac{n^2-1}{12}$

<sup>72</sup>Non è solo una divisa che tende a non dare nell'occhio; è anche una distribuzione a tutti gli effetti.

## B.2 Bernoulli

$$Be(p)$$

Misura l'esito di un singolo esperimento il cui esito può essere vero (1) o falso (0), con proporzione  $p \in [0, 1]$ .



**Supporto:**  $\{0, 1\}$

**Funzione di densità:**  $\mathbb{P}(X = 1) = p$   
 $\mathbb{P}(X = 0) = 1 - p$

**Funzione di ripartizione:**  $F(k) = \begin{cases} 0 & \text{per } k < 0 \\ 1 - p & \text{per } 0 \leq k < 1 \\ 1 & \text{per } k \geq 1 \end{cases}$

**Funzione caratteristica:**  $\varphi(u) = pe^{iu} + 1 - p$

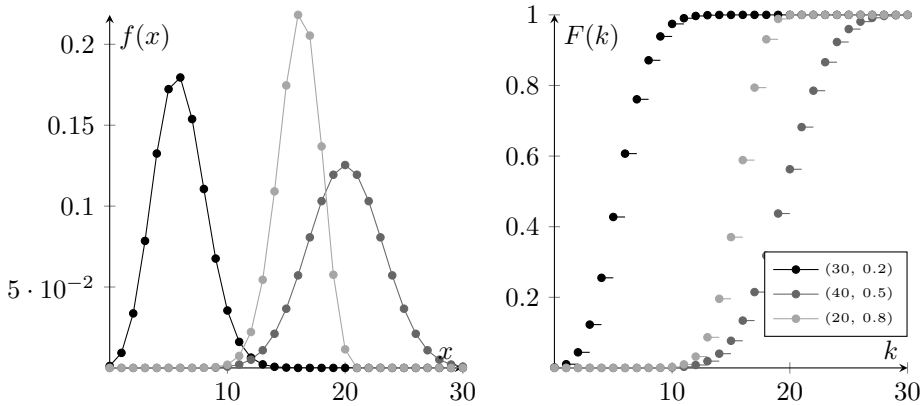
**Valore atteso:**  $\mathbb{E}[X] = p$

**Varianza:**  $Var(X) = p(1 - p)$

### B.3 Binomiale

$$Bi(n, p) = \sum_{i=1}^n Be(p)$$

Misura la somma di successi in  $n \in \mathbb{N}$  esperimenti di Bernoulli di identica proporzione  $p \in [0, 1]$ . Coincide quindi con la somma di  $n$  bernoulliane.



**Supporto:**  $\{0, 1, 2, \dots, n\}$

---

**Funzione di densità:**  $P(X = x) = \binom{n}{x} p^x (1 - p)^{n-x}$

---

**Funzione di ripartizione:** Tu, in realtà, non la vuoi sapere.

---

**Funzione caratteristica:**  $\varphi(u) = (pe^{iu} + 1 - p)^n$

---

**Valore atteso:**  $\mathbb{E}[X] = np$

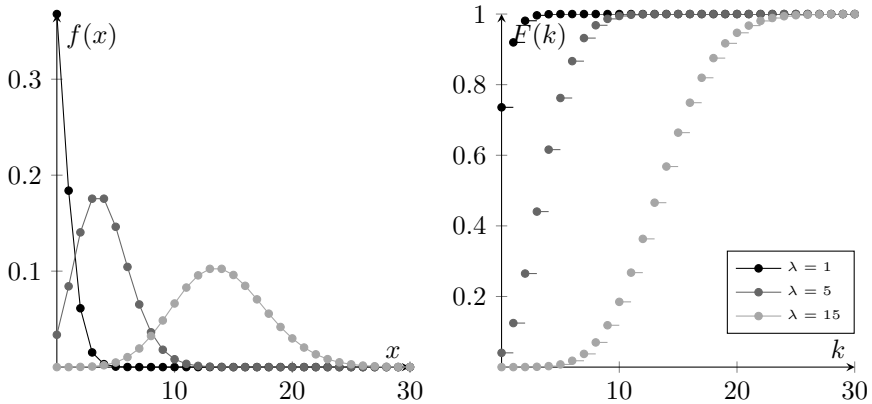
---

**Varianza:**  $Var(X) = np(1 - p)$

## B.4 Poisson

$$Po(\lambda)$$

Chiamata anche legge degli eventi rari, esprime le probabilità per il numero di eventi che si verificano successivamente ed indipendentemente in un dato intervallo di tempo, sapendo che mediamente se ne verifica un numero  $\lambda > 0$ . È il limite delle serie di distribuzioni binomiali con  $\lambda = np$  (si ha convergenza in legge).



**Supporto:**

$\mathbb{N}$

**Funzione di densità:**

$$\mathbb{P}(X = x) = e^{-\lambda} \frac{\lambda^x}{x!}$$

**Funzione di ripartizione:**

Tù ne quaèsieris.

**Funzione caratteristica:**

$$\varphi(u) = \exp\{\lambda(e^{iu} - 1)\}$$

**Valore atteso:**

$$\mathbb{E}[X] = \lambda$$

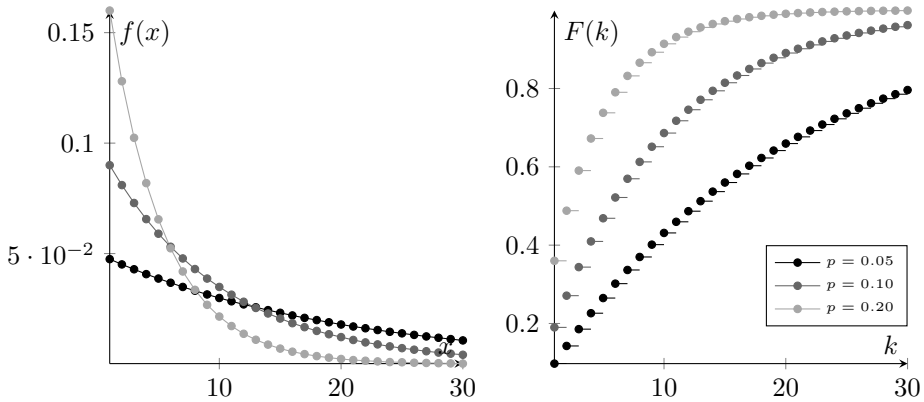
**Varianza:**

$$\text{Var}(X) = \lambda$$

## B.5 Geometrica

$$\mathcal{G}(p)$$

Misura il numero di fallimenti prima di un successo in un processo di Bernoulli di proporzione  $p \in [0, 1]$ . È priva di memoria.



**Supporto:**

$$\mathbb{N}^+$$

**Funzione di densità:**

$$\mathbb{P}(X = x) = p(1 - p)^{x-1}$$

**Funzione di ripartizione:**

$$F(k) = 1 - (1 - p)^k$$

**Funzione caratteristica:**

$$\varphi(u) = \frac{pe^{iu}}{1 - e^{iu}(1 - p)}$$

**Valore atteso:**

$$\mathbb{E}[X] = \frac{1}{p}$$

**Varianza:**

$$\text{Var}(X) = \frac{1 - p}{p^2}$$

Si può definire anche la distribuzione **geometrica traslata**:

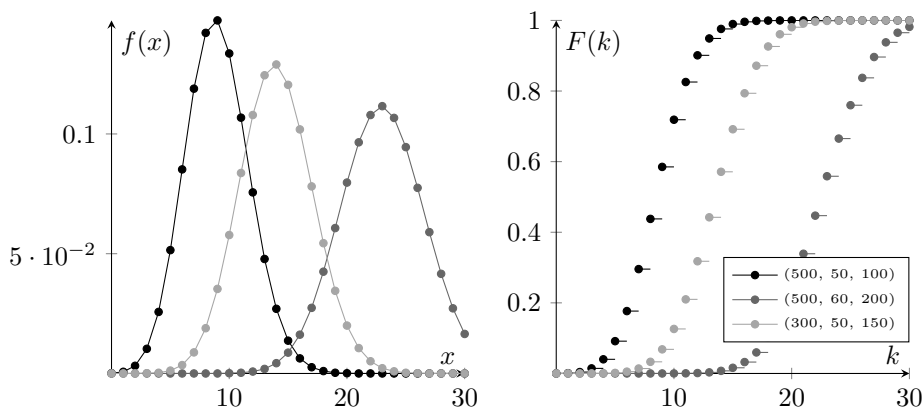
$$\mathbb{P}(X = x) = p(1 - p)^x, \quad F(k) = 1 - (1 - p)^{(k+1)}$$



## B.6 Ipergeometrica

$$\mathcal{H}(n, h, r)$$

Descrive l'estrazione senza reimmissione di palline da un'urna con  $n$  palline di cui  $h$  del tipo A e  $n-h$  del tipo B (per esempio, le prime verdi e le seconde gialle). Misura la probabilità di ottenere  $k$  palline del tipo A estraendone  $r$  dall'urna.



**Supporto:**  $k \in \{ \max\{0, r + h - n\}, \dots, \min\{r, h\} \}$

**Funzione di densità:**  $\mathbb{P}(X = x) = \frac{\binom{h}{x} \binom{n-h}{r-x}}{\binom{n}{r}}$

**Funzione di ripartizione:** Scire nefàs.

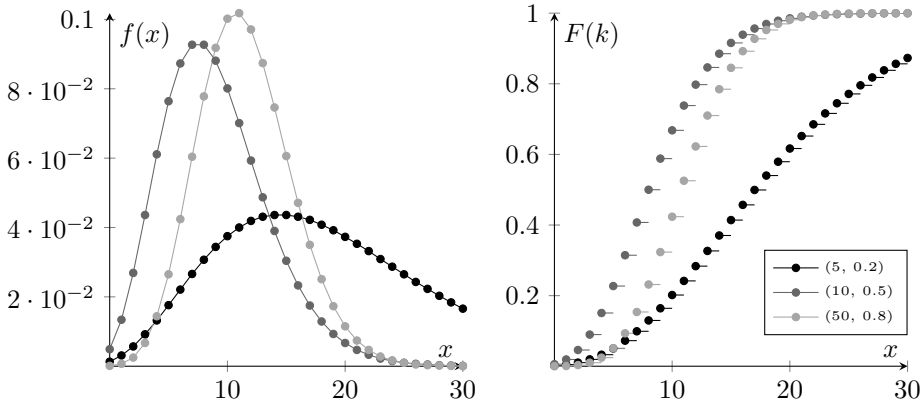
**Valore atteso:**  $\mathbb{E}[X] = \frac{rh}{n}$

**Varianza:**  $Var(X) = \frac{h(n-h)r(n-r)}{n^2(n-1)}$

## B.7 Pascal

$$NB(p, n)$$

Nota anche come **distribuzione binomiale negativa** o **distribuzione di Polya**, misura il numero di fallimenti precedenti il successo  $n$ -esimo in un processo di Bernoulli di proporzione  $p$ .



**Supporto:**  $\mathbb{N}$

**Funzione di densità:**  $\mathbb{P}(X = x) = \binom{x+n-1}{x} p^n (1-p)^x = \binom{-n}{x} p^n (p-1)^x$

**Funzione di ripartizione:** Hahahah, rassegnati caro.

**Funzione caratteristica:**  $\varphi(u) = \left( \frac{p}{1 - e^{iu}(1-p)} \right)^n$

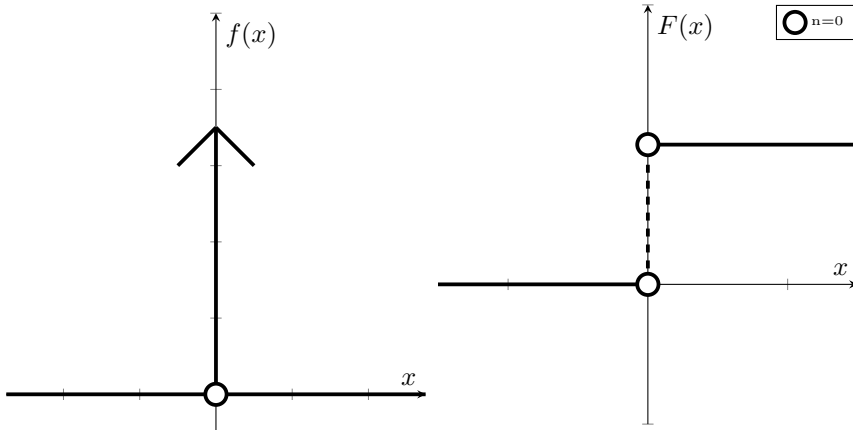
**Valore atteso:**  $\mathbb{E}[X] = \frac{n}{p}$

**Varianza:**  $Va[X] = n \frac{1-p}{p^2}$

## B.8 Delta di Dirac

$$\delta(n)$$

Chiamata anche **massa** o **distribuzione degenere**, concentra tutta la probabilità nel punto  $n$ . Ciò significa che vale zero in ogni punto fuorché  $n$ , mentre l'integrale sull'intero dominio vale 1. A causa della discontinuità della  $F$ , questa distribuzione non ammette densità.



**Supporto:**  $\{n\}$

**Funzione di ripartizione:**  $F(x) = \begin{cases} 1 & \text{per } x \geq n \\ 0 & \text{per } x < n \end{cases}$

**Funzione caratteristica:**  $\varphi(u) = e^{inu}$

**Valore atteso:**  $\mathbb{E}[X] = n$

**Varianza:**  $Var(X) = 0$

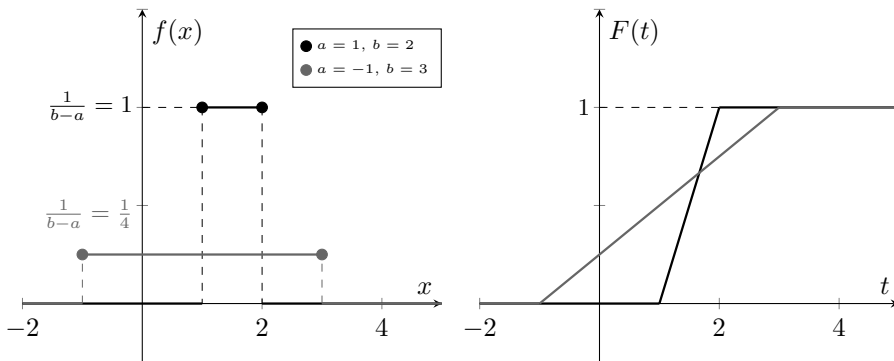
## Appendice C. Distribuzioni continue

Riportiamo qui una lista di alcune distribuzioni continue notevoli corredate da grafici e formule.

### C.1 Uniforme continua

$$U_{cont}(a, b)$$

Probabilità uniforme su un intervallo continuo  $[a, b]$ , con  $a < b$ .



**Supporto:**

$$[a, b] \in \mathbb{R}$$

**Funzione di densità:**

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{per } x \in [a, b] \\ 0 & \text{altrimenti} \end{cases}$$

**Funzione di ripartizione:**

$$F(t) = \begin{cases} 0 & \text{per } t < a \\ \frac{t-a}{b-a} & \text{per } t \in [a, b] \\ 1 & \text{per } t > b \end{cases}$$

**Funzione caratteristica:**

$$\varphi(u) = \exp\left\{i\frac{a+b}{2}u\right\} \frac{\sin\left(\frac{b-a}{2}u\right)}{\left(\frac{b-a}{2}u\right)}$$

**Valore atteso:**

$$\mathbb{E}[X] = \frac{1}{2}(a+b)$$

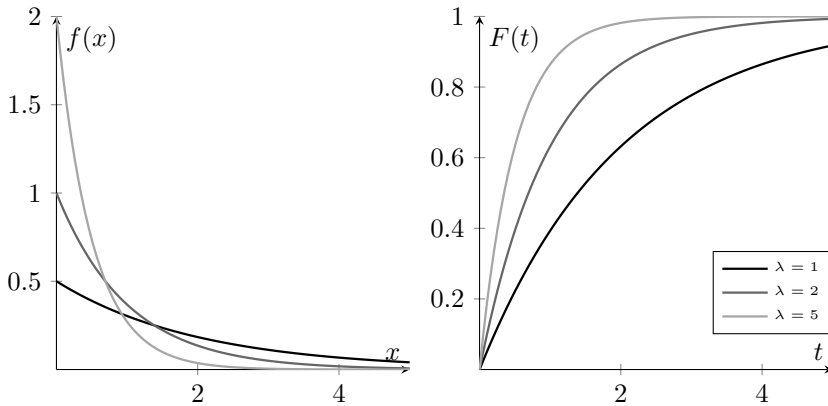
**Varianza:**

$$Var(X) = \frac{1}{12}(b-a)^2$$

## C.2 Esponenziale

$$\mathcal{E}(\lambda)$$

Descrive il tempo di attesa al verificarsi di un successo che avviene periodicamente: misura dunque la durata tra eventi in un processo di Poisson di parametro  $\lambda$ . In quanto priva di memoria, può modellizzare il tempo di vita di oggetti che non invecchiano.



**Supporto:**  $[0, +\infty)$

**Funzione di densità:**  $f(x) = \lambda e^{-\lambda x} \mathbf{1}_{(0, +\infty)}(x)$

**Funzione di ripartizione:**  $F(t) = 1 - e^{-\lambda t} \mathbf{1}_{(0, +\infty)}(t)$

**Funzione caratteristica:**  $\varphi(u) = \frac{\lambda}{\lambda - iu}$

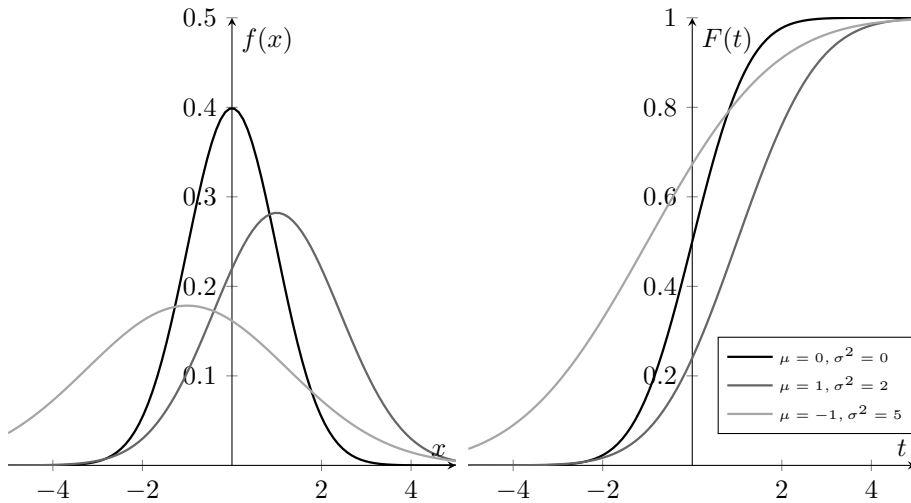
**Valore atteso:**  $\mathbb{E}[X] = \frac{1}{\lambda}$

**Varianza:**  $Var(X) = \frac{1}{\lambda^2}$

### C.3 Normale

$$\mathcal{N}(\mu, \sigma^2)$$

Detta anche **distribuzione gaussiana**, riveste un ruolo cruciale nella teoria della probabilità e della statistica. I suoi parametri  $\mu$  e  $\sigma^2$  coincidono rispettivamente con la sua media e varianza. Nel caso di media nulla e varianza unitaria si ha una **normale standard**:  $Z \sim \mathcal{N}(0, 1)$ .

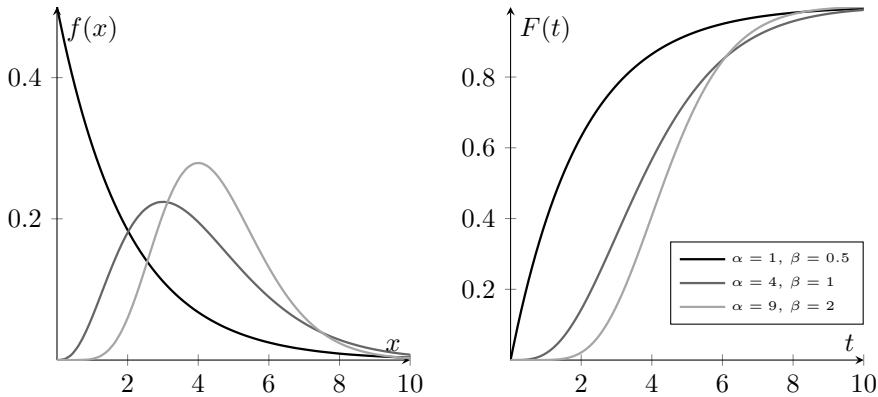


<b>Supporto:</b>	$\mathbb{R}$
<b>Funzione di densità:</b>	$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}$
<b>Funzione di ripartizione:</b>	$F(t) = \Phi\left(\frac{t-\mu}{\sigma}\right) = \frac{1}{2} \left[1 + \operatorname{erf}\left(\frac{t-\mu}{\sqrt{2}\sigma}\right)\right]$
<b>Funzione caratteristica:</b>	$\varphi(u) = \exp\left\{iu\mu - \frac{\sigma^2}{2}u^2\right\}$
<b>Valore atteso:</b>	$\mathbb{E}[X] = \mu$
<b>Varianza:</b>	$\operatorname{Var}(X) = \sigma^2$
<b>Momento quarto:</b>	$\mathbb{E}[(X-\mu)^4] = 3\sigma^4$
<b>Trasformazioni affini:</b>	<p>Date <math>X \sim \mathcal{N}(\mu_X, \sigma_X^2)</math>, <math>Y \sim \mathcal{N}(\mu_Y, \sigma_Y^2)</math></p> <p><math>AX + b \sim \mathcal{N}(A\mu_X + b, A\sigma_X^2 A^T)</math></p> <p><math>aX + bY \sim \mathcal{N}(a\mu_X + b\mu_Y, a^2\sigma_X^2 + b^2\sigma_Y^2 + 2ab\sigma_{XY})</math></p> <p><math>a^2\sigma_X^2 + b^2\sigma_Y^2 + 2ab\sigma_{XY} \geq 0</math> (Condizione da verificare)</p>

## C.4 Gamma

$$\Gamma(\alpha, \beta)$$

Comprende come casi particolari le distribuzioni esponenziali e chi-quadro. È definita come la somma di VA indipendenti con distribuzione esponenziale ( $\alpha$  VA esponenziali di parametro  $\beta$ ), con  $\alpha > 0$  e  $\beta > 0$ .



**Supporto:**  $[0, +\infty)$

**Funzione di densità:**  $f(x) = \frac{\beta^\alpha x^{\alpha-1} e^{-\beta x}}{\Gamma(\alpha)} \mathbf{1}_{(0, +\infty)}(x)$

**Funzione di ripartizione:**  $F(t) = \frac{\gamma(\alpha, \beta t)}{\Gamma(\alpha)}$

**Funzione caratteristica:**  $\varphi(u) = \left( \frac{\beta}{\beta - iu} \right)^\alpha$

**Valore atteso e momenti:**  $\mathbb{E}[X^k] = \frac{\alpha(\alpha+1)\dots(\alpha+k-1)}{\beta^k}$

**Varianza:**  $Var(X) = \frac{\alpha}{\beta^2}$

**Prodotto per scalare:**  $X \sim \Gamma(\alpha, \beta), Y = cX, c > 0 \implies Y \sim \Gamma(\alpha, \frac{\beta}{c})$

Dove la *funzione* Gamma di Eulero è definita come:

$$\Gamma(x) = \int_0^{+\infty} t^{x-1} e^{-t} dt$$

Un caso particolare della distribuzione gamma è la distribuzione erlanghiana (da A. K. Erlang) dove i parametri  $\alpha$  e  $\beta$  sono interi positivi. Solo in questo caso la funzione di ripartizione si può scrivere come:

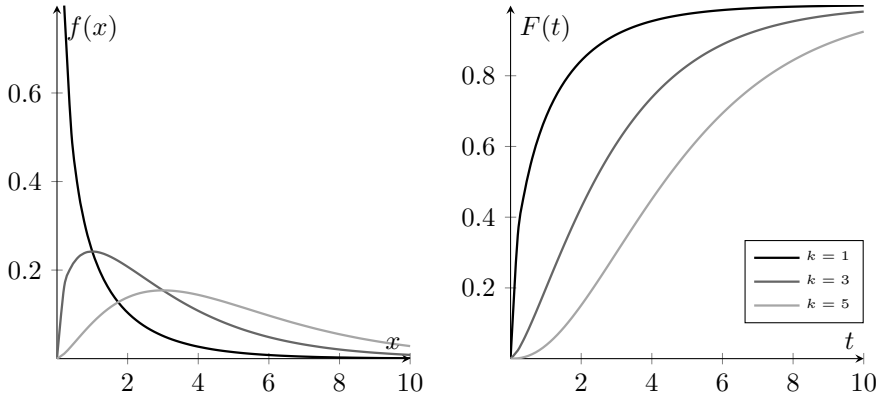
$$F(t) = \left( 1 - \sum_{i=0}^{\alpha-1} e^{-\beta t} \frac{(\beta t)^i}{i!} \right)$$



## C.5 Chi-quadro

$$\chi^2(k) = \Gamma\left(\frac{k}{2}, \frac{1}{2}\right) = \sum_{i=1}^k Z_i^2$$

Somma di  $k$  normali standard, ovvero  $Z \sim \mathcal{N}(0, 1)$ , ciascuna elevata al quadrato.



**Supporto:**  $[0, +\infty)$

**Funzione di densità:**  $f(x) = \frac{1}{2^{\frac{k}{2}} \Gamma\left(\frac{k}{2}\right)} x^{\frac{k}{2}-1} e^{-\frac{x}{2}} \mathbf{1}_{(0, +\infty)}$

**Funzione di ripartizione:**  $F(t) = \frac{\gamma\left(\frac{k}{2}, \frac{t}{2}\right)}{\Gamma\left(\frac{k}{2}\right)}$

**Funzione caratteristica:**  $\varphi(u) = (1 - 2iu)^{-\frac{k}{2}}$

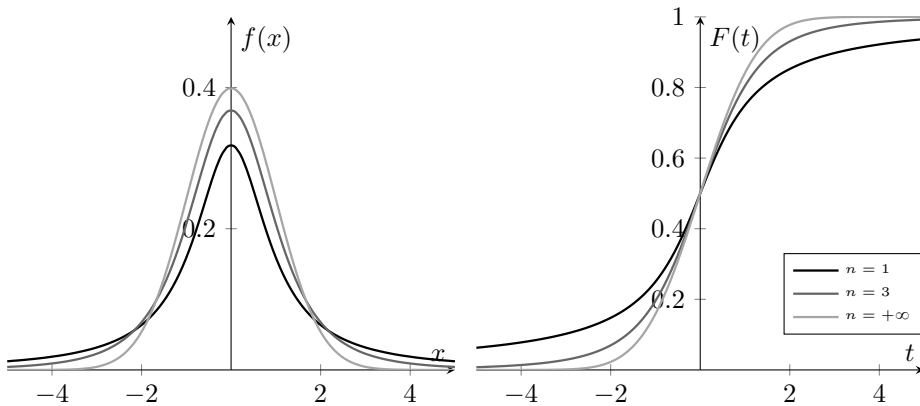
**Valore atteso:**  $\mathbb{E}[X] = k$

**Varianza:**  $Var(X) = 2k$

### C.6 $t$ di Student

$$t(n) = \frac{\mathcal{N}(0, 1)}{\sqrt{\frac{\chi^2(n)}{n}}}$$

Definita come rapporto fra una normale standard e un'espressione legata alla chi-quadro, nell'ipotesi che queste due variabili aleatorie siano indipendenti. Riveste un ruolo importante nella statistica, in particolare nella stima della media di una popolazione. Il parametro  $n$  viene detto *numero di gradi di libertà*.



**Supporto:**  $\mathbb{R}$

**Funzione di densità:**  $f(x) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\sqrt{\pi n}} \cdot \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}}$

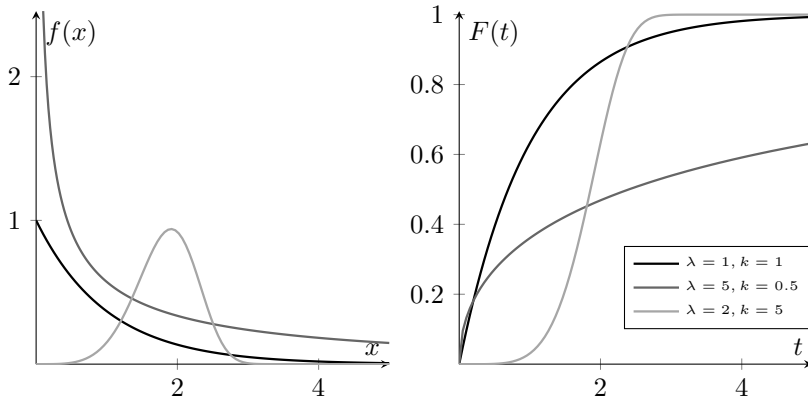
**Valore atteso:**  $\mathbb{E}[X] = 0$  se  $n > 1$  oppure indefinito

**Varianza:**  $Var(X) = \frac{n}{n-2}$  se  $n > 2$  oppure indefinita

## C.7 Weibull

$$W(\lambda, k)$$

Qui trovate rappresentata la più comune parametrizzazione:  $\lambda > 0$ ,  $k > 0$ .



**Supporto:**  $[0, +\infty)$

**Funzione di densità:**  $f(x) = \frac{k}{\lambda^k} x^{k-1} e^{-(x/\lambda)^k} \mathbb{1}_{(0, +\infty)}(x)$

**Funzione di ripartizione:**  $F(t) = 1 - e^{-(\frac{t}{\lambda})^k}$

**Funzione caratteristica:**  $\varphi(u) = \sum_{n=0}^{+\infty} \frac{(iu)^n \lambda^n}{n!} \Gamma\left(1 + \frac{n}{k}\right)$

**Valore atteso:**  $\mathbb{E}[X] = \frac{\lambda}{k} \Gamma\left(\frac{1}{k}\right)$

**Varianza:**  $Var(X) = \frac{\lambda^2}{k^2} [2k\Gamma\left(\frac{2}{k}\right) - \Gamma^2\left(\frac{1}{k}\right)]$

È spesso impiegata per la stimare il tempo medio di guasto (o MTTF, mean time to failure) di un sistema. L'asse delle ascisse rappresenta il tempo, o il numero di cicli a cui è sottoposto il campione (per esempio cadute o oscillazioni caldo-freddo); l'asse delle ordinate nella densità rappresenta la probabilità che un oggetto si guasti nel tempo, e nella funzione di ripartizione la percentuale di oggetti guasti.

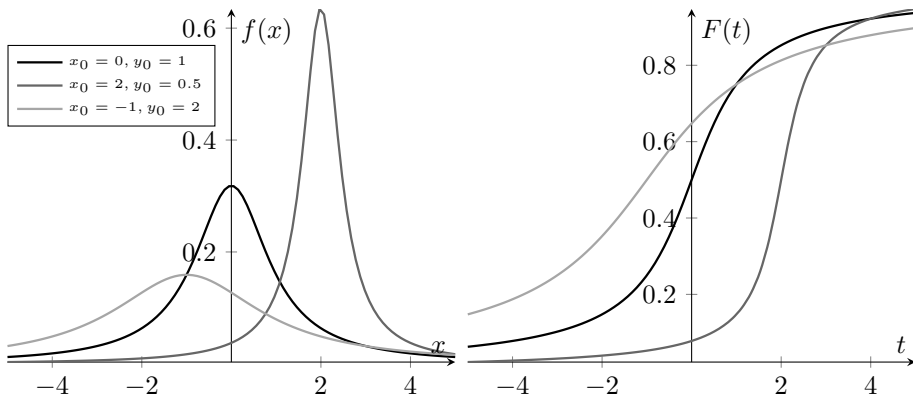
Si noti che il parametro  $k$ , denominato *tasso di guasto* o *modulo di Weibull*, fornisce un'importante informazione:

- $k < 1$ : la probabilità di guastarsi si riduce nel tempo, per esempio nel caso del rodaggio di un macchinario, producendo quindi un'alta "mortalità infantile";
- $k = 1$ : la probabilità di guastarsi rimane invariata, e questo è il caso particolare della distribuzione esponenziale  $\mathcal{E}(\lambda)$ ;
- $k > 1$ : la probabilità di guastarsi aumenta nel tempo, per esempio in un sistema dove le sollecitazioni precedenti danneggiano senza guastare.

## C.8 Cauchy

$$\mathcal{C}(x_0, y_0)$$

Caso insolito tra le distribuzioni canoniche, in quanto una variabile con questa legge non è in  $L^1$  e pertanto non ha né valore atteso né varianza. Ha i parametri  $x_0 \in \mathbb{R}$ ,  $y_0 > 0$ .



**Supporto:**  $\mathbb{R}$

**Funzione di densità:**  $f(x) = \frac{1}{\pi} \frac{y_0}{(x - x_0)^2 + y_0^2}$

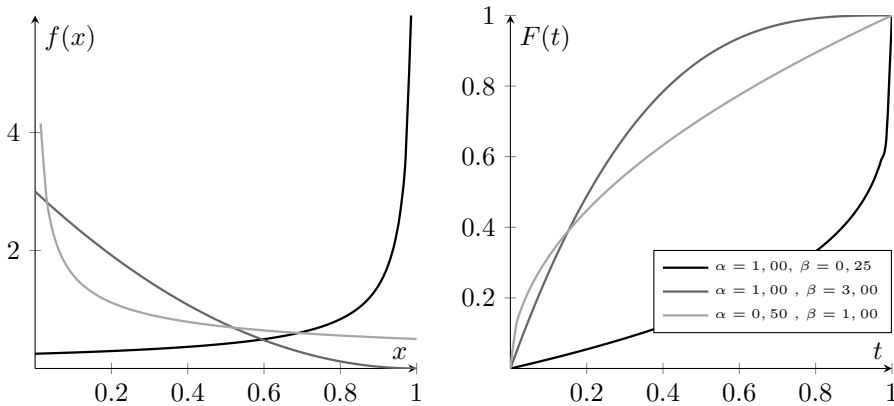
**Funzione di ripartizione:**  $F(t) = \frac{1}{\pi} \arctan\left(\frac{t - x_0}{y_0}\right) + \frac{1}{2}$

**Funzione caratteristica:**  $\varphi(u) = \exp\{ix_0u - y_0|u|\}$

### C.9 Beta

$$B(\alpha, \beta)$$

Governa la probabilità  $p$ , a priori distribuita uniformemente, di un processo di Bernoulli dopo aver osservato  $\alpha - 1$  successi e  $\beta - 1$  fallimenti; con  $\alpha > 0$  e  $\beta > 0$ .



**Supporto:**  $[0, 1]$

**Funzione di densità:**  $f(x) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha, \beta)}$

**Valore atteso:**  $\mathbb{E}[X] = \frac{\alpha}{\alpha + \beta}$

**Varianza:**  $Var(X) = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}$

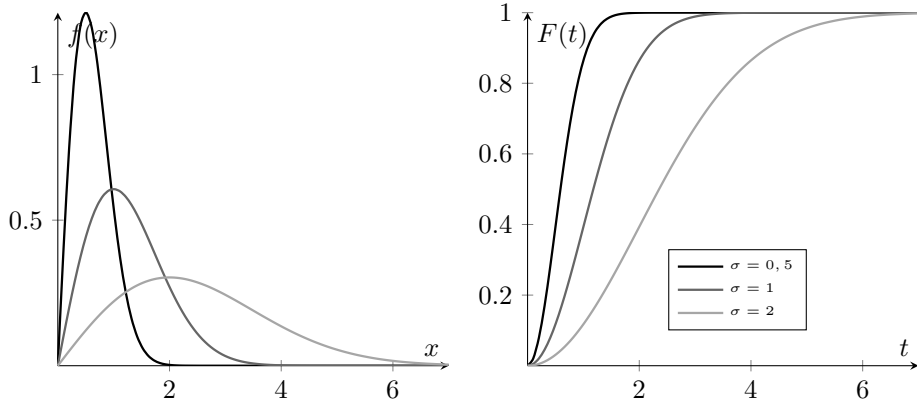
Dove a denominatore della densità troviamo la *funzione* Beta di Eulero, così definita:

$$B(x, y) = \int_0^1 t^{x-1}(1-t)^{y-1} dt$$

## C.10 Rayleigh

$$Z(\sigma)$$

Descrive la variabile aleatoria  $Z = \sqrt{X^2 + Y^2}$ , dove  $X$  e  $Y$  hanno legge  $\mathcal{N}(0, \sigma^2)$ ; con  $\sigma > 0$ .



**Supporto:**  $[0, +\infty)$

**Funzione di densità:**  $f(x) = \frac{x}{\sigma^2} e^{-\frac{x^2}{2\sigma^2}}$

**Funzione di ripartizione:**  $F(t) = 1 - e^{-\frac{t^2}{2\sigma^2}}$

**Valore atteso:**  $\mathbb{E}[X] = \sigma\sqrt{\frac{\pi}{2}}$

**Varianza:**  $Var(X) = (2 - \frac{\pi}{2})\sigma^2$





# Indice analitico

- $\sigma$ -additività, 16
- $\sigma$ -algebra, 11
  - generata da una collezione, 12
  - generata da una VA, 178
  - prodotto, 110
- additività, 16
- algebra, 11
- approssimazione di una VAR, 74
- asintotica normalità, 211
- attesa condizionata, 177
  - da una  $\sigma$ -algebra, 181
  - da una VA, 177
- Bayes, formula di, 29
- bilinearità della covarianza, 138
- Bonferroni, disuguaglianza di, 20
- Borel
  - $\sigma$ -algebra di, 14
  - paradosso di, 168
- campione casuale, 162
- catena, 221
- catena di Markov, 224
  - aperiodica, 233
  - irriducibile, 235
  - periodo di una, 233
- Cauchy-Schwarz, disuguaglianza di, 82
- Chapman-Kolmogorov, equazione di, 231
- Chebyshev, disuguaglianza di, 85
- chi-quadro ( $\chi^2$ ), legge, 161
- chiusura
  - per complementazione, 11
  - per differenze, 32
  - per intersezioni finite, 31
  - per limiti crescenti, 31
  - per unioni e intersezioni, 11
- classe (collezione), 11
- chiusa, 235
- irriducibile, 235
- limite, 251
- classi monotone, teorema delle, 32
- coefficiente di correlazione lineare, 137
- composizione di funzioni misurabili, 54
- controimmagine, 252
- convergenza, 187
  - certa, 187
  - debole, 194
  - dominata, teorema della, 73
  - in  $L^p$ , 189
  - in legge ( $\mathcal{L}$ ), 195
  - in probabilità ( $\mathbb{P}$ ), 190
  - monotona, teorema della, 73
  - quasi certa (qc), 187
  - velocità di, 210
- convoluzione, 151
- covarianza, 135
  - condizionata, 168
- De Morgan, leggi di, 252
- delta di Dirac, 42
- densità di probabilità
  - continua, 44
  - discreta, 122
  - invariante, 243
  - reversibile, 246
  - stazionaria, 243
- disgiunti a coppie, eventi, 16
- distribuzione, 58
  - Cauchy, 220
  - esponenziale, 45
  - gaussiana, 46
  - normale, 46
  - uniforme, 45
- Doob, lemma di misurabilità di, 179
- equazione del bilancio dettagliato, 246

errore quadratico medio (MSE), 184, 216  
 esperimento aleatorio, 9  
 evento, 10  
     improbabile, 22  
     quasi certo, 22  
 famiglia  
     di eventi indipendenti, 26  
     di VA indipendenti, 118  
 Fatou, lemma di, 73  
 Fourier, trasformata di, 143  
 Fubini-Tonelli, teorema di, 112  
     multidimensionale, 119  
 funzione  
     boreliana, 56  
     caratteristica, 143  
     di ripartizione, 37  
     di ripartizione empirica, 218  
     di ripartizione in  $\mathbb{R}^n$ , 122  
     generatrice dei momenti, 147  
     indicatrice, 254  
     misurabile, 52  
 Glivenko-Cantelli, teorema di, 219  
 immagine, 252  
 indipendenza  
     di eventi, 24  
     di VA, 103  
     tra una VA e una  $\sigma$ -algebra, 182  
 insieme delle parti, 10  
 integrale (valore atteso), 63  
 interpretazioni modellistiche della probabilità, 15  
 iper-quadranti di sud-ovest, 55, 111  
 istante di primo arrivo, 236  
     in una classe, 240  
 $\mathcal{L}^1$ , spazio, 72  
 $L^1$ , spazio, 82  
 $L^p$ , spazi, 82  
 Lebesgue  
     integrale di, 97  
     integrale in  $\mathbb{R}^n$ , 128  
     misura di, 95  
     misura su  $\mathbb{R}^n$ , 127  
 legge, 58  
     condizionata, 167  
     congiunta, 116, 124  
     dei grandi numeri (LGN), 203  
     iniziale (catene di Markov), 224  
     marginale, 116, 124  
 Levy, teorema di continuità di, 199  
 limite inferiore e superiore  
     di una successione insiemistica, 253  
     di una successione numerica, 251  
 macchina da scrivere, 189  
 Markov  
     catena di, 224  
     disuguaglianza di, 85  
     proprietà di, 224  
 Markov-Kakutani, teorema di, 244  
 matrice di transizione, 222  
     bistocastica, 243  
     regolare, 249  
 matrice varianza, 139  
 media  
     (valore atteso), 63  
     campionaria, 162  
     orizzontale, 247  
     verticale, 247  
 metodo delta 1, 213  
 misura di probabilità, 16  
 misurabilità, 52  
 momento, 145  
 omogeneità (catene di Markov), 224  
 parte  
     negativa di una VA, 71  
     positiva di una VA, 71  
 partizione, 28

periodo di uno stato, 232  
 principio fondamentale del calcolo  
     combinatorio, 255  
 probabilità, 16  
     condizionata, 23  
     di arrivo (catene di Markov),  
         236  
     di assorbimento, 240  
     di transizione, 222  
     su un'algebra, 31  
 probabilità totali, formula delle,  
     28  
 processo stocastico, 221  
 prodotto  
     cerchiato, 110  
     di convoluzione, 151  
 proporzione campionaria, 218  
 proprietà di Markov, 225  
  
 quasi certo (qc)  
     evento, 22  
     proprietà, 62  
     uguaglianza, 60  
 quasi ovunque (qo), 98  
  
 regola del valore atteso, 88, 93,  
     101  
 rovina del giocatore, 226  
  
 scomposizione della varianza,  
     185  
 scorrelazione, 135  
 semi-definita positività, 139  
 seno cardinale, 149  
 Slutsky, teorema di, 202  
 sotto- $\sigma$ -algebra, 181  
 spazio  
     campionario, 9  
     degli stati, 49  
     di probabilità, 21  
     misurabile, 12  
 stati, spazio degli, 221  
 stato  
     aperiodico, 233  
     assorbente, 235  
     che comunica con un altro,  
         235  
     che conduce a un altro, 233  
     periodico, 233  
     ricorrente, 236  
         transitorio, 236  
 Stiefel, formula di, 256  
 stimatore corretto o non distorto,  
     215  
 subadditività, 20  
 successione  
     di eventi, 17  
     di VA semplici, 74  
  
 tempo atteso di assorbimento,  
     240  
 teorema  
     centrale del limite (TCL),  
         207  
     ergodico, 247  
 topologia euclidea di  $\mathbb{R}$ , 14  
 traiettoria, 222  
 trasformazione affine, 148  
     per vettori normali, 157  
  
 urna di Ehrenfest, 226  
  
 valore atteso, 63  
     condizionato, 168  
     generalizzato, 71  
     in  $\mathbb{R}^n$ , 139  
     per VAR continue, 101  
     per VAR discrete, 93  
     per VAR positive, 69  
     per VAR semplici, 65  
     per vettori continui, 130  
     per vettori discreti, 125  
     su un insieme, 91  
 variabile aleatoria, 49  
     congiuntamente normale,  
         154  
     reale, 63  
     reale continua, 98  
     reale discreta, 93  
     semplice, 63  
 varianza, 83  
     campionaria, 162  
     condizionata, 168  
     condizionata (VA), 185  
 vettore aleatorio, 121  
     continuo, 129  
     discreto, 122  
     gaussiano, 154  
 vettore media (valore atteso),  
     139