

# Probabilità\*

Questi appunti sono stati presi durante le lezioni dell'insegnamento "Probabilità" tenuto dalla prof. Guatteri durante l'A. A. 2022/23. Non sono stati revisionati da alcun docente (potrebbero contenere errori, in forma e in sostanza, di qualsiasi tipo) e non sono in alcun modo sostitutivi della frequentazione delle lezioni.

---

\*[mariachiara.menicucci@mail.polimi.it](mailto:mariachiara.menicucci@mail.polimi.it) per segnalare errori, richiedere il codice LaTeX ecc.

# Indice

<b>1</b>	<b>Fondamenti</b>	<b>4</b>
<b>2</b>	<b>Definizione assiomatica di probabilità</b>	<b>6</b>
2.1	Spazio campionario . . . . .	7
2.2	$\sigma$ -algebre . . . . .	8
2.3	Misure di probabilità . . . . .	12
2.3.1	Assegnazione di una probabilità . . . . .	16
<b>3</b>	<b>Probabilità condizionata</b>	<b>19</b>
<b>4</b>	<b>Indipendenza stocastica</b>	<b>24</b>
<b>5</b>	<b>Assegnazione della probabilità su <math>\mathcal{B}(\mathbb{R})</math></b>	<b>27</b>
<b>6</b>	<b>Variabili aleatorie</b>	<b>33</b>
6.1	Relazione tra variabili aleatorie . . . . .	39
6.2	Integrazione rispetto a una misura di probabilità . . . . .	42
6.2.1	Spazi $L^p$ e varianza . . . . .	52
6.2.2	Calcolo del valore atteso di una variabile aleatoria . . . . .	55
<b>7</b>	<b>Variabili aleatorie discrete</b>	<b>57</b>
<b>8</b>	<b>Variabili aleatorie assolutamente continue</b>	<b>61</b>
8.1	Integrazione rispetto a una misura . . . . .	63
<b>9</b>	<b>Variabili aleatorie indipendenti e spazi prodotto</b>	<b>75</b>
<b>10</b>	<b>Vettori aleatori</b>	<b>85</b>
10.1	Vettori aleatori discreti . . . . .	86
10.2	Vettori assolutamente continui . . . . .	88
10.3	Covarianza . . . . .	92
<b>11</b>	<b>Funzione caratteristica</b>	<b>97</b>

<b>12 Vettori gaussiani</b>	<b>103</b>
12.1 Funzioni non lineari di vettori gaussiani e statistiche campionarie . . . . .	109
<b>13 Leggi condizionali</b>	<b>111</b>
13.1 Attese condizionate . . . . .	117
13.2 Varianza condizionata . . . . .	124
<b>14 Convergenza di successioni di variabili aleatorie</b>	<b>126</b>
14.1 Convergenza debole . . . . .	133
<b>15 Teoremi limite</b>	<b>140</b>
15.1 Applicazioni . . . . .	143
15.2 Funzione di ripartizione empirica . . . . .	143
<b>16 Catene di Markov</b>	<b>145</b>
16.1 Classificazione degli stati . . . . .	150
16.2 Comportamento asintotico di catene . . . . .	161

# 1 Fondamenti

In genere indicheremo una funzione usando la notazione  $X : \Omega \rightarrow E$ , dove  $\Omega$  è il dominio e  $E$  è il codominio; si indicano con  $A$  i sottinsiemi di  $\Omega$ , con  $B$  quelli di  $E$ , con  $\omega$  il generico elemento di  $\Omega$ , con  $x$  l'immagine di  $\omega$  attraverso  $X$  ( $X(\omega) = x \in E$ ).  $\{\omega\} \subseteq \Omega$ . Di solito per noi  $E$  sarà  $\mathbb{R}$  o  $\mathbb{R}^n$ , mentre per  $\Omega$  abbiamo una vasta scelta.

**Def** Data  $X : \Omega \rightarrow E$ ,  $A \subseteq \Omega$ , si dice immagine di  $A$  attraverso  $X$ , e si indica con  $X(A)$ ,  $\{x \in E : \exists \omega \in A : X(\omega) = x\}$ .

L'insieme degli  $X(A)$  al variare di  $A$ , cioè  $\{X(A) : A \subseteq \Omega\}$  è una famiglia di sottinsiemi di  $E$ .  $X(A) \subseteq E$ .

**Def** Data  $X : \Omega \rightarrow E$ ,  $B \subseteq E$ , si dice controimmagine di  $B$  attraverso  $X$ , e si indica con  $X^{-1}(B)$ ,  $\{\omega \in \Omega : X(\omega) \in B\}$ .

Vale  $X^{-1}(B) \subseteq \Omega$ .  $X^{-1}(B)$  è un oggetto ben definito per ogni  $B \subseteq E$ . Inoltre è possibile che  $X(\Omega) \subset E$ , mentre  $X^{-1}(E) = \Omega$ .

L'insieme degli  $X^{-1}(B)$  al variare di  $B$ , cioè  $\{X^{-1}(B) : B \subseteq E\}$  è una famiglia di sottinsiemi di  $\Omega$ .

1. Sia  $X : \mathbb{R} \rightarrow \mathbb{R}$ ,  $X(\omega) = \omega^2$ .  $X(\Omega) = [0, +\infty)$ , che è un sottinsieme proprio di  $E$ . Siano  $A = (0, 1) \subseteq \Omega$ ,  $B = (0, 1) \subseteq E$ .  $X(A) = (0, 1)$ ,  $X^{-1}(B) = (-1, 0) \cup (0, 1) \subseteq \Omega$ .  $A^c = (-\infty, 0] \cup [1, +\infty)$ :  $X(A^c) = [0, +\infty)$ , mentre  $(X(A))^c = (-\infty, 0] \cup [1, +\infty)$ , quindi  $X(A^c) \neq (X(A))^c$ : le due operazioni non commutano.

Sia  $\tilde{A} = (-1, 0)$ :  $\tilde{A} \cap A = \emptyset$ , cioè  $\tilde{A}$  e  $A$  sono disgiunti.  $X(A) = (0, 1)$ ,  $X(\tilde{A}) = (0, 1)$ , quindi  $X(A) \cap X(\tilde{A}) = (0, 1)$ , che è diverso da  $X(\tilde{A} \cap A) = \emptyset$ : le operazioni di intersezione e immagine non commutano.

$X^{-1}(E) = \Omega$ .  $X^{-1}(B) = \emptyset$  per ogni  $B \subseteq (-\infty, 0) = (X(\Omega))^c$ , invece  $X^{-1}((-\infty, 0]) = \{0\}$ .

$B^c = (-\infty, 0] \cup [1, +\infty)$ :  $X^{-1}(B^c) = (-\infty, -1] \cup \{0\} \cup [1, +\infty)$ , mentre  $(X^{-1}(B))^c = ((-1, 0) \cup (0, 1))^c = (-\infty, -1] \cup \{0\} \cup [1, +\infty)$ : in questo caso le operazioni di controimmagine e complementazione commutano, cioè  $X^{-1}(B^c) = (X^{-1}(B))^c$ .

Sia  $\tilde{B} = (-1, 0) \subseteq E$ .  $\tilde{B} \cap B = \emptyset$  e  $X^{-1}(\tilde{B} \cap B) = \emptyset$ , ma anche  $X^{-1}(B) \cap X^{-1}(\tilde{B}) = ((-1, 0) \cup (0, 1)) \cap \emptyset = \emptyset$ : in questo caso le operazioni di intersezione e controimmagine commutano.

In generale valgono le seguenti proprietà.

### Proprietà (controimmagine di una funzione)

Hp:  $X : \Omega \rightarrow E, A \subseteq \Omega, B \subseteq E, I$  qualsiasi

$$\text{Ts: (i) } X(X^{-1}(B)) \subseteq B$$

$$\text{(ii) } X^{-1}(X(A)) \supseteq A$$

$$\text{(iii) } X^{-1}(B^c) = (X^{-1}(B))^c$$

$$\text{(iv) } \forall (B_\alpha)_{\alpha \in I} : B_\alpha \subseteq E \forall \alpha \in I, X^{-1}(\cap_\alpha B_\alpha) = \cap_\alpha X^{-1}(B_\alpha)$$

$$\text{e } X^{-1}(\cup_\alpha B_\alpha) = \cup_\alpha X^{-1}(B_\alpha)$$

Un esempio di validità di (i), (ii) si ha quando  $B = E, A = \Omega$ : (ii) vale con l'uguaglianza se  $A = \Omega$ . In generale, con l'operazione di immagine è possibile "perdere" qualcosa, mentre con la controimmagine è possibile "guadagnare" qualcosa. (iv) è significativa in particolare quando  $I$  è non numerabile: afferma che la controimmagine dell'intersezione è l'intersezione delle controimmagini.

In probabilità  $X^{-1}(B)$  si indica con  $(X \in B)$ . Quindi e. g.  $X^{-1}(B^c) = (X \in B^c) = (X \notin B) = (X^{-1}(B))^c = (X \in B)^c$ : per le proprietà della controimmagine, il simbolo di complementare passa fuori.

Date  $X : \Omega \rightarrow E, h : E \rightarrow E'$ , si ha la funzione composta  $(h \circ X)(\omega) = Y(\omega) = h(X(\omega))$ , funzione da  $\Omega$  in  $E'$ . Per ogni  $C' \subseteq E' (h \circ X)^{-1}(C') = X^{-1}(h^{-1}(C'))$ .

**Funzioni a valori in  $\mathbb{R}$**   $\mathbb{R}$  è chiuso rispetto alle operazioni di addizione, sottrazione, divisione, moltiplicazione, massimo tra due elementi (indicato con  $\vee$ ) e minimo tra due elementi (indicato con  $\wedge$ ): forma un reticolo. Quindi, date  $X_1 : \Omega \rightarrow \mathbb{R}, X_2 : \Omega \rightarrow \mathbb{R}$ , è ben definita e. g.  $X_1 + X_2 : \Omega \rightarrow \mathbb{R}$ , cioè  $(X_1 + X_2)(\omega) = X_1(\omega) + X_2(\omega)$ .

1. Sia  $\Omega = \{t, c\} \times \{t, c\} = \{(t, t), (t, c), (c, t), (c, c)\} = \{\omega = (\omega_1, \omega_2) : \omega_k \in \{t, c\} \forall k = 1, 2\}$ , che modella i

possibili esiti di due lanci consecutivi di una moneta.  $X_1 : \Omega \rightarrow \mathbb{R}, X_1(\omega) = \begin{cases} 1 & \text{se } \omega_1 = t \\ 0 & \text{altrimenti} \end{cases}, X_2 : \Omega \rightarrow \mathbb{R},$

$X_2(\omega) = \begin{cases} 1 & \text{se } \omega_2 = t \\ 0 & \text{altrimenti} \end{cases}$ .  $X_i$  è la funzione indicatrice dell'evento "al lancio  $i$  è uscita testa".  $X_1 \neq X_2$

perché ad esempio se  $\omega = (t, c)$   $X_1, X_2$  hanno valori diversi.  $X_1$  può essere rappresentata come  $\begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}$ , dove

in ogni quadratino è presente il valore di  $X_1$  associato a un certo  $\omega \in \Omega$ : infatti  $|\Omega| = 4$ .  $a_{11}$  è il valore di  $X_1$  per  $\omega = (c, t)$ ,  $a_{12}$  per  $\omega = (c, c)$ ,  $a_{21}$  per  $\omega = (t, c)$ ,  $a_{22}$  per  $\omega = (t, t)$ . Analogamente si rappresenta  $X_2$  con

$$\begin{matrix} 0 & 0 \\ 1 & 1 \end{matrix} . \text{ Allora } X_1 + X_2 \text{ è } \begin{matrix} 1 & 0 \\ 2 & 1 \end{matrix} : (X_1 + X_2)(\omega) = \begin{cases} 2 \text{ se } \omega = (t, t) \\ 0 \text{ se } \omega = (c, c) : X_1 + X_2 \text{ conta il numero di teste} \\ 1 \text{ altrimenti} \end{cases}$$

uscite nei due lanci.

$$2X_1 - X_2 \text{ è rappresentabile come } \begin{matrix} 2 & 0 \\ 1 & -1 \end{matrix}, X_1 \vee X_2 \text{ con } \begin{matrix} 1 & 0 \\ 1 & 1 \end{matrix}, X_1 X_2 \text{ con } \begin{matrix} 0 & 0 \\ 1 & 0 \end{matrix}. \text{ L'insieme } (X_1 = 0) \text{ è}$$

l'insieme delle coppie per cui il primo elemento non è  $t$ , dunque  $\{(c, t), (c, c)\}$ :  $(X_1 = 0) = X_1^{-1}(\{0\}) = \{\omega \in \Omega : X_1(\omega) = 0\} = \{\omega \in \Omega : \omega_1 \neq t\}$ .

**Spazio di Bernoulli** Considero gli esperimenti che possono avere solo due esiti: successo o insuccesso.

Si definisce  $\Omega = \{0, 1\}^{\mathbb{N}}$  come l'insieme delle successioni per cui ogni elemento è 0 o 1, cioè  $\Omega = \{0, 1\}^{\mathbb{N}} = \{(\omega_n)_{n \in \mathbb{N}} : \omega_n \in \{0, 1\} \forall n\}$ .  $\Omega$  è quindi uno spazio di successioni (detto spazio di Bernoulli), cioè il generico elemento di  $\Omega$  è e. g.  $\bar{\omega} = (0, 1, \dots, 0, 1, \dots)$ .  $\Omega$  è non numerabile, per cui si vedrà che è ragionevole scegliere  $\mathcal{A} = \sigma\{(E_k)_{k \geq 1}\}$ , dove  $E_k$  è l'evento che rappresenta il successo al tentativo  $k$ .

Data una successione  $\omega \in \Omega$ , si definisce  $X_n : \Omega \rightarrow \mathbb{R}, X_n(\omega) = \omega_n$ : è una funzione che a ogni successione associa il suo elemento  $n$ -esimo, quindi una funzione proiezione che restituisce l'esito dell' $n$ -esimo esperimento.

Se l' $n$ -esimo elemento di  $\bar{\omega}$  è 1,  $X_n(\bar{\omega}) = \bar{\omega}_n = 1$ ;  $X_0(\bar{\omega}) = 0, X_1(\bar{\omega}) = 1, X_{n-1}(\bar{\omega}) = 0$ .  $(X_n)_{n \geq 0}$  è una successione di funzioni, ciascuna delle quali, al variare di  $\omega \in \Omega$ , ha immagine  $X_n(\Omega) = \{0, 1\}$ .

1.  $(X_2 = 0)$  è  $X_2^{-1}(\{0\})$ , cioè l'insieme delle successioni di  $\Omega$  che hanno il secondo elemento nullo:  $\{\omega \in \Omega : \omega_2 = 0\}$ .
2.  $\bigcap_{n=0}^{+\infty} (X_n = 0)$  è l'insieme delle successioni di  $\Omega$  tali che  $X_i = 0 \forall i = 1, \dots, n$ , quindi  $\{\omega_n : \omega_n = 0 \forall n\}$ , cioè la successione nulla.

**Funzione indicatrice** Dato  $A \subseteq \Omega$ , si definisce funzione indicatrice dell'insieme  $A$  la funzione  $I_A : \Omega \rightarrow$

$$\mathbb{R}, I_A(\omega) = \begin{cases} 1 \text{ se } \omega \in A \\ 0 \text{ altrimenti} \end{cases}. \text{ Quindi } (I_A = 1) = I_A^{-1}(\{1\}) = A, \text{ mentre } (I_A = 0) = A^c. \text{ Se } A_1, \dots, A_n \subseteq \Omega, \text{ la}$$

funzione indicatrice della loro intersezione è  $I_{\bigcap_{k=1}^n A_k} = \prod_{k=1}^n I_{A_k} = I_{A_1} I_{A_2} \dots I_{A_n}$ . Inoltre  $I_{A_1 \cup A_2} = I_{A_1} + I_{A_2} - I_{A_1} I_{A_2}$ .

## 2 Definizione assiomatica di probabilità

Si dice esperimento aleatorio un esperimento, non necessariamente ripetibile nel tempo, di cui non è noto a priori l'esito. E. g., si può supporre di avere un'urna che contiene palline rosse e bianche e di voler contare quante palline

rosse ci sono: il numero non è noto a priori. La probabilità studia modelli per questi esperimenti aleatori. Ci interessa in particolare - come modello di un esperimento aleatorio - la terna di Kolmogorov  $(\Omega, \mathcal{A}, \mathbf{P})$ , che è alla base dell'approccio assiomatico alla probabilità. L'approccio assiomatico di Kolmogorov consiste nel fatto che la sua teoria non dice come calcolare probabilità, ma quali proprietà deve avere un certo oggetto matematico per poter essere chiamato modello di un esperimento aleatorio.

## 2.1 Spazio campionario

**Def** Si dice spazio campionario o spazio degli esiti, e si indica con  $\Omega$ , un insieme che contiene tutti i possibili esiti dell'esperimento aleatorio di cui si vuole costruire il modello. Un generico elemento di  $\Omega$ ,  $\omega \in \Omega$ , cioè un possibile risultato dell'esperimento, si dice realizzazione dell'esperimento.

E' conveniente che  $\Omega$  contenga tutti e soli i risultati dell'esperimento; in generale però la scelta di  $\Omega$  dipende dallo sperimentatore e dalla grandezza di interesse, non è univocamente determinata dall'esperimento aleatorio; quando è possibile, è conveniente scegliere  $\Omega \subseteq \mathbb{R}$ .

$\Omega$  può essere finito, infinito numerabile o infinito più che numerabile; quando  $\Omega$  è al più numerabile (cioè è finito o infinito numerabile) si dice che  $\Omega$  è discreto.

1. Se l'esperimento aleatorio è il lancio di una moneta e il risultato di interesse è la faccia che esce, ci sono solo due possibili esiti: esce testa o esce croce. Posso quindi scegliere uno spazio campionario categorico  $\Omega = \{t, c\}$ , oppure - se 0 indica che esce croce -  $\Omega = \{0, 1\}$ .
2. Se lancio un dado e il risultato di interesse è il numero che esce sulla faccia superiore,  $\Omega = \{1, 2, 3, 4, 5, 6\}$ .
3. Se lancio  $n$  volte una moneta e il risultato di interesse è il numero di volte che esce testa,  $\Omega = \{0, 1, \dots, n\}$ .
4. Se ho un'urna con  $M$  palline rosse e bianche, tra cui almeno una rossa e almeno una bianca, la grandezza di interesse è il numero di palline rosse e conto il numero di palline rosse,  $\Omega = \{1, \dots, M-1\}$ .
5. Se si considera il tempo di vita di una lampadina,  $\Omega = [0, T]$ , o  $\Omega = [0, +\infty]$ .

Si dice evento una proposizione logica di cui, dopo l'esperimento aleatorio, è noto il valore di verità o falsità. Ogni realizzazione di  $\Omega$ ,  $\omega \in \Omega$ , corrisponde a un evento  $\{\omega\} \subseteq \Omega$ , detto evento elementare. Ogni evento è definito da un insieme  $A \subseteq \Omega$  (quindi un sottinsieme di  $\Omega$ ) tale che la proposizione logica che descrive l'evento è vera se e solo se il risultato dell'esperimento  $\omega$  appartiene a  $A$ , e allora si dice che l'evento si è verificato. Se invece  $\omega \notin A$ , si è verificato  $A^c$ .

1. Se lancio un dado,  $\Omega = \{1, 2, 3, 4, 5, 6\}$ ; la proposizione "esce il numero 1" è un evento, che posso indicare con  $A$  e a cui corrisponde un insieme a cui il risultato dell'esperimento appartiene se e solo se la proposizione è vera.  $A = \{1\} \subseteq \Omega$ .  $C = \{2, 4, 6\}$  è l'evento "esce un numero pari"; pur non essendo  $C$  un evento elementare, può essere visto come unione di eventi elementari:  $C = \{2\} \cup \{4\} \cup \{6\}$ .

## 2.2 $\sigma$ -algebre

In generale non è possibile descrivere per elencazione l'insieme di tutti gli eventi relativi a un esperimento; se ne descrivono le proprietà.

**Def 1.1** Dato un qualsiasi insieme  $\Omega$  e l'insieme delle parti di  $\Omega$   $2^\Omega$ , dato  $\mathcal{A} \subseteq 2^\Omega$ , si dice che  $\mathcal{A}$  è una  $\sigma$ -algebra su  $\Omega$  se valgono le seguenti proprietà:

**A1**  $\Omega \in \mathcal{A}, \emptyset \in \mathcal{A}$

**A2**  $A \in \mathcal{A} \implies A^c \in \mathcal{A}$

**A3** Se  $(A_n)_{n \geq 1}$  è una successione di eventi con  $A_n \in \mathcal{A} \forall n$ , allora  $\bigcup_{n=1}^{+\infty} A_n \in \mathcal{A}$  e  $\bigcap_{n=1}^{+\infty} A_n \in \mathcal{A}$

$\mathcal{A}$  è "l'insieme degli eventi relativi a un esperimento che ci possono interessare" (il generico elemento di  $\mathcal{A}$  ha il significato modellistico di evento) e le tre proprietà richieste sono ragionevoli per poter avere un insieme abbastanza ricco di eventi: se ci interessano gli eventi  $A_1, A_2$ , ci interesseranno anche i loro complementari, le loro unioni e le loro intersezioni. Una generica  $\sigma$ -algebra è una famiglia di sottinsiemi di  $\Omega$  e non necessariamente coincide con l'insieme delle parti di  $\Omega$ : può esserne anche un sottinsieme proprio.

A2 afferma che  $\mathcal{A}$  è chiusa rispetto all'operazione di complementare, A3 che è chiusa rispetto all'intersezione e unione numerabile.  $(A_n)_{n \geq 1}$  con  $A_n \in \mathcal{A} \forall n$  è una successione di eventi. La coppia  $(\Omega, \mathcal{A})$  si dice spazio probabilizzabile o spazio misurabile.

Se  $\Omega$  è al più numerabile, classicamente si considera come  $\sigma$ -algebra su  $\Omega$  l'insieme delle parti di  $\Omega$ ; se invece  $\Omega = \mathbb{R}$  noi, come  $\sigma$ -algebra su  $\mathbb{R}$ , non considereremo mai  $\mathcal{P}(\mathbb{R})$ , perché così facendo non si riesce a definire una misura di probabilità che sia invariante per traslazioni. Questo si farà in genere se  $\Omega$  non è discreto.

La definizione può essere raffinata. Infatti, dalle richieste  $\Omega \in \mathcal{A}$  e A2 segue che  $\emptyset \in \mathcal{A}$ . Inoltre, per le leggi di De Morgan,  $(\bigcup_{\alpha} A_{\alpha})^c = \bigcap_{\alpha} A_{\alpha}^c$ : quindi, se vale A2 e se  $(A_n)_{n \geq 1}$  successione con  $A_n \in \mathcal{A} \forall n \implies \bigcup_{n=1}^{+\infty} A_n \in \mathcal{A}$ , si ha che  $A_n^c \in \mathcal{A} \forall n$  e inoltre  $\bigcup_{n=1}^{+\infty} A_n^c \in \mathcal{A}$ . Allora, per A2,  $(\bigcup_{n=1}^{+\infty} A_n^c)^c = \bigcap_{n=1}^{+\infty} A_n \in \mathcal{A}$ . La definizione ripulita delle ridondanze diventa quindi



**Def 1.2** Dato un qualsiasi insieme  $\Omega$  e l'insieme delle parti di  $\Omega$   $2^\Omega$ , dato  $\mathcal{A} \subseteq 2^\Omega$ , si dice che  $\mathcal{A}$  è una  $\sigma$ -algebra su  $\Omega$  se

**A1'**  $\Omega \in \mathcal{A}$

**A2**  $A \in \mathcal{A} \implies A^c \in \mathcal{A}$

**A3'**  $(A_n)_{n \geq 1}$  è una successione di eventi con  $A_n \in \mathcal{A} \forall n \implies \bigcup_{n=1}^{+\infty} A_n \in \mathcal{A}$

Nella definizione, la chiusura è rispetto a un'unione numerabile, quindi infinita: è però naturale chiedere che la chiusura sia anche rispetto a un'unione finita (si sta affermando che ogni  $\sigma$ -algebra è un'algebra).

**Prop 1.3 (chiusura per unione e intersezione finita)**

Hp:  $\mathcal{A}$  è una  $\sigma$ -algebra su  $\Omega$ ,  $n$  è finito,  $A_1, A_2, \dots, A_n$

sono tali che  $A_k \in \mathcal{A} \forall k = 1, \dots, n$

$$\text{Ts: } \bigcup_{k=1}^n A_k \in \mathcal{A}, \bigcap_{k=1}^n A_k \in \mathcal{A}$$

La tesi significa che  $\mathcal{A}$  è chiusa anche per unione e intersezione finita. Questo implica che sia anche chiusa per differenze, perché  $A_1 \setminus A_2 = A_1 \cap A_2^c$ .

**Dim** Sia  $n \in \mathbb{N}$ ,  $A_1, A_2, \dots, A_n : A_k \in \mathcal{A} \forall k = 1, \dots, n$ . Devo dimostrare che  $\bigcup_{k=1}^n A_k \in \mathcal{A}$ : mi devo quindi inventare una successione di eventi in modo da ottenere  $\bigcup_{k=1}^n A_k$  come unione numerabile di eventi, per usare A3. Definisco allora  $(B_k)_{k \geq 1}$ , famiglia numerabile di eventi, in modo che  $B_k \in \mathcal{A} \forall k$  e  $\bigcup_{k=1}^n A_k = \bigcup_{k=1}^{+\infty} B_k$ . Prendo

$$B_k = \begin{cases} A_k & \text{se } k \leq n \\ \emptyset & \text{se } k \geq n+1 \end{cases} : B_k \in \mathcal{A} \forall k \text{ (} A_k \text{ per ipotesi, } \emptyset \text{ per A1) e } \bigcup_{k=1}^{+\infty} B_k = \bigcup_{k=1}^n A_k \cup \emptyset \cup \dots = \bigcup_{k=1}^n A_k, \text{ ma}$$

$\bigcup_{k=1}^{+\infty} B_k \in \mathcal{A}$  per A3, dunque anche  $\bigcup_{k=1}^n A_k \in \mathcal{A}$ .

Allora  $\mathcal{A}$  è chiusa anche per intersezione finita, perché  $\bigcap_{k=1}^n A_k = (\bigcup_{k=1}^n A_k^c)^c$ , e  $(\bigcup_{k=1}^n A_k^c)^c \in \mathcal{A}$  per quanto appena mostrato e per A2. ■

Si è quindi visto come dalla sola richiesta (oltre ad A1 e A2) della chiusura rispetto all'unione numerabile si siano dedotte anche la chiusura rispetto all'intersezione numerabile, l'intersezione finita e l'unione finita.

Data  $(\mathcal{A}_\alpha)$  una famiglia di  $\sigma$ -algre su  $\Omega$ ,  $\bigcap_\alpha \mathcal{A}_\alpha$  è ancora una  $\sigma$ -algebra su  $\Omega$ , perché  $\emptyset, \Omega \in \mathcal{A}_\alpha \forall \alpha$ ; dato un generico elemento in  $\bigcap_\alpha \mathcal{A}_\alpha$ , questo si trova in ciascuna  $\mathcal{A}_\alpha$ , dunque il suo complementare si troverà in ciascuna  $\mathcal{A}_\alpha$  e dunque anche in  $\bigcap_\alpha \mathcal{A}_\alpha$ ; dati  $n$  elementi di  $\bigcap_\alpha \mathcal{A}_\alpha$ , questi si trovano in ciascuna  $\mathcal{A}_\alpha$ , dunque la loro unione si troverà in ciascuna  $\mathcal{A}_\alpha$  e dunque anche in  $\bigcap_\alpha \mathcal{A}_\alpha$ . L'unione di  $\sigma$ -algre non è invece una  $\sigma$ -algebra.

1. Dato  $\Omega$ , la  $\sigma$ -algebra  $\{\emptyset, \Omega\}$ , per cui sono banalmente veri A1, A2 e A3, è detta  $\sigma$ -algebra banale.

2. Dato  $\Omega$ , l'insieme delle parti di  $\Omega$   $2^\Omega$  è la più grande  $\sigma$ -algebra su  $\Omega$ , nel senso che qualsiasi  $\sigma$ -algebra su  $\Omega$  è sottinsieme di  $2^\Omega$ .
3. Dato  $\Omega$ , se  $A$  è un qualsiasi elemento di  $2^\Omega$  (cioè un qualsiasi sottinsieme di  $\Omega$ ),  $\{\emptyset, \Omega, A, A^c\}$  è una  $\sigma$ -algebra.
4. Considero  $\Omega = \{1, 2, 3, 4, 5, 6\}$  e l'evento  $A = \{2, 4, 6\}$  ("esce un numero pari"). La più piccola  $\sigma$ -algebra su  $\Omega$  che contenga  $A$  è  $\{\emptyset, \Omega, \{2, 4, 6\}, \{1, 3, 5\}\}$ .

**Def** Dato un qualsiasi insieme  $\Omega$  e l'insieme delle parti di  $\Omega$   $2^\Omega$ , dato  $\mathcal{A} \subseteq 2^\Omega$ , si dice che  $\mathcal{A}$  è un'algebra su  $\Omega$  se

$$\mathbf{A1} \quad \Omega \in \mathcal{A}, \emptyset \in \mathcal{A}$$

$$\mathbf{A2} \quad A \in \mathcal{A} \implies A^c \in \mathcal{A}$$

$$\mathbf{A3} \quad n \in \mathbb{N}, A_1, \dots, A_n \in \mathcal{A} \implies \bigcup_{k=1}^n A_k \in \mathcal{A} \text{ e } \bigcap_{k=1}^n A_k \in \mathcal{A}$$

Ogni  $\sigma$ -algebra su  $\Omega$  è quindi un'algebra su  $\Omega$ .

**Def 1.4** Dato  $\Omega$  e  $\mathcal{C} \subseteq 2^\Omega$  una classe di sottinsiemi di  $\Omega$ , si dice  $\sigma$ -algebra generata dalla classe  $\mathcal{C}$ , e si indica con  $\sigma(\mathcal{C})$ , la più piccola  $\sigma$ -algebra su  $\Omega$  contenente  $\mathcal{C}$ .

Con  $\mathcal{C}$  in generale si indica un insieme di eventi, quindi un sottinsieme di  $2^\Omega$ . Con "la più piccola  $\sigma$ -algebra" si intende l'intersezione tra tutte le  $\sigma$ -algre che contengono  $\mathcal{C}$ . Questa definizione serve perché è possibile che si sia interessati in particolare ad alcuni eventi (cioè alcuni elementi di  $2^\Omega$ ), e si desidera quindi trovare una  $\sigma$ -algebra adeguata per costruire un modello probabilistico dell'esperimento; in genere è conveniente lavorare con una  $\sigma$ -algebra piccola.

**Prop 1.5 (esistenza della  $\sigma$ -algebra generata da un insieme di eventi)**

Hp:  $\Omega$  è un insieme qualsiasi,  $\mathcal{C} \subseteq 2^\Omega$  una classe di sottinsiemi di  $\Omega$

Ts:  $\exists \sigma(\mathcal{C})$

Questo significa che la definizione 1.4 è ben posta.

**Dim\*** Essendo  $\mathcal{C} \subseteq 2^\Omega$ , poiché  $2^\Omega$  è una  $\sigma$ -algebra, esiste una  $\sigma$ -algebra che contiene  $\mathcal{C}$ . L'intersezione di  $\sigma$ -algre è ancora una  $\sigma$ -algebra: quindi se considero la famiglia di tutte le  $\sigma$ -algre su  $\Omega$  che contengono  $\mathcal{C}$  e la chiamo  $\mathcal{F}$ , vale  $\bigcap_{\alpha \in \mathcal{F}} \mathcal{A}_\alpha = \sigma(\mathcal{C})$ . ■

1. Se  $\Omega$  è un insieme e  $A \subseteq \Omega$ ,  $\sigma(A) = \{\Omega, \emptyset, A, A^c\}$ .

**Def 1.6** Dato  $\Omega$ , si dice partizione discreta di  $\Omega$ , e si indica con  $\mathcal{C}$ , una famiglia di sottinsiemi di  $\Omega$   $\{C_k\}_{k \in I}$  tale che  $I$  è al più numerabile,  $C_k \cap C_h = \emptyset \forall h \neq k$ ,  $\bigcup_{k \in I} C_k = \Omega$ .

La seconda richiesta significa che i  $C_k$  sono a due a due disgiunti, la terza che almeno uno degli eventi  $C_k$  si verifica sempre (essendo gli eventi disgiunti, in realtà se ne verifica sempre esattamente uno).

**Prop 1.6 ( $\sigma$ -algebra generata da una partizione discreta)**

Hp:  $\Omega$  è un insieme qualsiasi,  $\mathcal{C}$  è una partizione discreta di  $\Omega$

$$\text{Ts: } \sigma(\mathcal{C}) = \left\{ A \subseteq \Omega : A = \bigcup_{k \in J} C_k \text{ al variare di } J \subseteq I \right\}$$

Infatti  $\{A \subseteq \Omega : A = \bigcup_{k \in J} C_k, J \subseteq I\}$  è un insieme costruito in modo da essere chiuso rispetto all'unione numerabile. Contiene  $\emptyset$  (si prende  $J = \emptyset$ ) e  $\Omega$  (si prende  $J = I$ ). Se  $\Omega$  è finito,  $|\sigma(\mathcal{C})| = 2^{|I|}$ .

1. Sia  $\Omega = \{1, \dots, 6\}$ .  $C_1 = \{1, 6\}, C_2 = \{2, 5\}, C_3 = \{3, 4\}$ :  $\{C_k\}$  è una partizione discreta di  $\Omega$ , indicizzata su  $I = \{1, 2, 3\}$ . I  $J$  da considerare sono quindi  $\emptyset, \{1\}, \{2\}, \{3\}, \{1, 2\}, \{1, 3\}, \{2, 3\}, \{1, 2, 3\}$  e si ottiene  $\sigma(\mathcal{C}) = \{\emptyset, \{1, 6\}, \{2, 5\}, \{3, 4\}, \{1, 2, 5, 6\}, \{1, 3, 4, 6\}, \{2, 3, 4, 5\}, \{1, 2, 3, 4, 5, 6\}\}$ , che è effettivamente una  $\sigma$ -algebra.
2. Se  $\Omega$  è al più numerabile, la famiglia  $\mathcal{C} = \{\{\omega_i\}\}_{i \in I}$  di tutti i singoletti di  $\Omega$  è una partizione di  $\Omega$  e  $\sigma(\mathcal{C}) = 2^\Omega$ .

**Def 1.7** Si definisce insieme dei boreliani di  $\mathbb{R}$ , e si indica con  $\mathcal{B}(\mathbb{R})$ , la  $\sigma$ -algebra generata dalla classe  $\mathcal{C} = \{(a, b) : -\infty \leq a < b \leq +\infty\}$ .

Si potrebbe equivalentemente usare  $[a, b]$  o  $(a, b]$  per la definizione.  $\mathcal{C}$  è l'insieme di tutti gli intervalli aperti di  $\mathbb{R}$ , e evidentemente non è una  $\sigma$ -algebra perché non è chiusa per complementazione:  $(0, 1)^c$  non è un intervallo aperto.  $\mathcal{B}(\mathbb{R})$  contiene tutti gli intervalli di  $\mathbb{R}$  (ad esempio,  $(a, b] = \bigcap_{n=1}^{+\infty} (a, b + \frac{1}{n}) \in \mathcal{B}(\mathbb{R})$ ), le loro unioni e le loro intersezioni,  $\mathbb{Q}$  e quindi  $I$ : in generale contiene tutti i sottinsiemi non "patologici" di  $\mathbb{R}$  (non l'insieme di Vitali, ad esempio), quindi  $\mathcal{B}(\mathbb{R}) \subsetneq 2^\mathbb{R}$ . In generale, si può mostrare che se 1.7 è data usando gli intervalli di un certo tipo, allora la  $\sigma$ -algebra generata contiene tutti gli intervalli di qualsiasi altro tipo: essenzialmente, quindi,  $\mathcal{B}(\mathbb{R})$  è la più piccola  $\sigma$ -algebra che contenga tutti gli intervalli di  $\mathbb{R}$ .

La definizione può essere data equivalentemente dicendo che  $\mathcal{B}(\mathbb{R})$  è la  $\sigma$ -algebra generata dalla topologia di  $\mathbb{R}$ , cioè gli aperti: l'equivalenza vale per ogni insieme aperto può essere scritto come unione numerabile di intervalli aperti disgiunti.

**Prop 1.8 (genesi dei boreliani)**

$$\text{Hp:} \quad \mathcal{C} = \{(-\infty, q] : q \in \mathbb{Q}\}$$

$$\text{Ts:} \quad \mathcal{B}(\mathbb{R}) = \sigma(\mathcal{C})$$

Quindi la  $\sigma$ -algebra generata dall'insieme degli intervalli inferiormente illimitati avente estremo destro razionale è la stessa generata dall'insieme degli aperti di  $\mathbb{R}$ . L'utilità della proposizione sta nel fatto che l'insieme delle semirette è più semplice rispetto a quello degli aperti.

## 2.3 Misure di probabilità

Storicamente ci sono stati tre principali approcci alla probabilità:

1. approccio soggettivista: la probabilità di un evento è il grado di fiducia dello sperimentatore nel verificarsi di tale evento. Dipende dal grado di avversione al rischio dello sperimentatore.
2. approccio frequentista: si suppone di poter ripetere più volte l'esperimento. Se si fanno  $n$  prove e l'evento di interesse si è verificato  $n_A$  volte, la frequenza relativa dell'evento  $A$  di interesse è  $f_n(A) = \frac{n_A}{n}$ . Allora si definisce la probabilità di  $A$  come  $\mathbf{P}(A) = \lim_{n \rightarrow +\infty} f_n(A)$ .
3. approccio classico: per ogni esperimento che abbia un numero finito di esiti possibili (occorre che la cardinalità di  $\Omega$  sia finita) si definisce  $\mathbf{P}(A)$  come numero di casi favorevoli fratto numero di casi possibili.

Se ad esempio lancio un dado non truccato, posso usare tutti e tre gli approcci e ottengo lo stesso risultato; se invece lancio un dado truccato (e. g. con una faccia di materiale diverso, che renda più probabile l'uscita della faccia opposta), l'unico approccio sensato è quello frequentista, perché non si dispone di informazioni sulla situazione. Per valutare la probabilità che ci siano elezioni politiche quest'anno è invece sensato usare l'approccio soggettivista (l'esperimento è non ripetibile).

**Def 2.1** Dato lo spazio misurabile  $(\Omega, \mathcal{A})$ , si dice misura di probabilità su  $(\Omega, \mathcal{A})$  una funzione  $\mathbf{P} : \mathcal{A} \rightarrow [0, 1]$  tale che

$$\mathbf{A1} \quad \mathbf{P}(\Omega) = 1$$

$$\mathbf{A2} \quad \text{se } (A_n)_{n \geq 1} \text{ è una famiglia numerabile tale che } A_n \in \mathcal{A} \forall n \geq 1 \text{ e } A_h \cap A_k = \emptyset \forall h \neq k, \text{ allora } \mathbf{P}\left(\bigcup_{n=1}^{+\infty} A_n\right) = \sum_{n=1}^{+\infty} \mathbf{P}(A_n)$$

A1 è ben posta perché  $\Omega \in \mathcal{A}$ .  $(A_n)_{n \geq 1}$  è quindi una famiglia numerabile di eventi. A2 è ben posta perché  $\bigcup_{n=1}^{+\infty} A_n$  è un'unione numerabile di elementi di  $\mathcal{A}$ , quindi appartiene a  $\mathcal{A}$  e  $\mathbf{P}$  è ivi definita.  $\Omega$  è detto evento certo, per il quale si ha il massimo grado di fiducia. A2 permette di calcolare la probabilità che almeno uno degli  $A_n$  (di fatto, esattamente uno, perché sono eventi disgiunti) si verifichi, se si conosce la probabilità di ciascun evento. A2 è detta  $\sigma$ -additività, ed è fondamentale il fatto che si richiede che  $\mathbf{P}$  sia additiva rispetto a famiglie numerabili di eventi. Se si richiede l'additività rispetto a una qualsiasi famiglia di eventi, anche non numerabile - oltre a non sapere se l'unione non numerabile di eventi appartiene ad  $\mathcal{A}$  - si incorre immediatamente in una contraddizione, perché se si considera  $(\Omega, \mathcal{A})$  con  $\Omega = [a, b]$ ,  $\mathcal{A}$  una  $\sigma$ -algebra qualsiasi su  $\Omega$  e una  $\mathbf{P} : \mathbf{P}(\{x\}) = 0 \forall x \in [a, b]$ , si ha che  $\mathbf{P}\left(\bigcup_{n=1}^{+\infty} A_n\right) = \mathbf{P}\left(\bigcup_{x \in [a, b]} \{x\}\right) = \mathbf{P}(\Omega) = \sum_{x \in [a, b]} 0$ , che contraddice A1.

La probabilità assegna a ogni evento, cioè a degli insiemi, un numero reale, che fornisce una misura degli eventi, proprio come, misurando le aree e i volumi, si assegna a un insieme un numero reale.

### Teo 2.2 (proprietà della probabilità)

Hp:  $\mathbf{P}$  è una misura di probabilità su  $(\Omega, \mathcal{A})$

Ts: (P1)  $\mathbf{P}(\emptyset) = 0$

(P2) Se  $\{A_1, \dots, A_n\}$  è un insieme tale che  $A_n \in \mathcal{A} \forall n \geq 1$

$$\text{e } A_h \cap A_k = \emptyset \forall h \neq k, \mathbf{P}\left(\bigcup_{k=1}^n A_k\right) = \sum_{k=1}^n \mathbf{P}(A_k)$$

P2 è detta finita additività: se  $\mathbf{P}$  è  $\sigma$ -additiva, allora è anche additiva rispetto a una famiglia finita di eventi disgiunti. Non vale invece il viceversa.

**Dim** (P1) E' logico pensare di usare A2. Considero la famiglia numerabile  $(A_n)_{n \geq 1}$  con  $A_n = \emptyset \forall n \geq 1$ : vale  $A_n \in \mathcal{A} \forall n$  e  $A_h \cap A_k = \emptyset \forall h \neq k$ , allora, poiché  $\mathbf{P}$  è una misura di probabilità, per A1  $\mathbf{P}\left(\bigcup_{n=1}^{+\infty} A_n\right) = \mathbf{P}(\emptyset) = \sum_{n=1}^{+\infty} \mathbf{P}(\emptyset)$ . Quindi, poiché  $\mathbf{P} \in [0, 1]$ ,  $\mathbf{P}(\emptyset) = 0$ . (Oppure:  $\mathbf{P}(\Omega \cup \emptyset) = 1$ )

(P2) Considero la famiglia numerabile  $(B_k)_{k \geq 1}$  definita come segue:  $B_k = \begin{cases} A_k & \text{se } k \leq n \\ \emptyset & \text{se } k > n \end{cases}$ . Vale  $B_k \in \mathcal{A} \forall k$

e  $B_h \cap B_k = \emptyset \forall h \neq k$  per le ipotesi e poi si hanno solo insiemi vuoti; inoltre  $\bigcup_{k=1}^{+\infty} B_k = \bigcup_{k=1}^n A_k$ , ma quindi per A2  $\mathbf{P}\left(\bigcup_{k=1}^{+\infty} B_k\right) = \sum_{k=1}^{+\infty} \mathbf{P}(B_k) = \sum_{k=1}^n \mathbf{P}(A_k) + \mathbf{P}(\emptyset) + \dots + \mathbf{P}(\emptyset) = \sum_{k=1}^n \mathbf{P}(A_k)$  e  $\bigcup_{k=1}^n A_k = \sum_{k=1}^n \mathbf{P}(A_k)$ .

■

### Teo 2.3 (proprietà della probabilità)

Hp:  $\mathbf{P}$  è una misura di probabilità su  $(\Omega, \mathcal{A})$

Ts: (P3) (monotonia) se  $A, A' \in \mathcal{A}$  e  $A \subseteq A'$ ,  $\mathbf{P}(A) \leq \mathbf{P}(A')$

(P4) (complementare) se  $A \in \mathcal{A}$ ,  $\mathbf{P}(A^c) = 1 - \mathbf{P}(A)$

(P5) (differenza) se  $A, A' \in \mathcal{A}$ ,  $\mathbf{P}(A \setminus A') = \mathbf{P}(A) - \mathbf{P}(A' \cap A)$

(P6) (unione) se  $A, A' \in \mathcal{A}$ ,  $\mathbf{P}(A \cup A') = \mathbf{P}(A) + \mathbf{P}(A') - \mathbf{P}(A \cap A')$

(P7) se  $(A_n)_{n \geq 1}$  è una famiglia numerabile tale che  $A_n \in \mathcal{A} \forall n$ ,  $\mathbf{P}\left(\bigcup_{n=1}^{+\infty} A_n\right) \leq \sum_{n=1}^{+\infty} \mathbf{P}(A_n)$

(P8) disuguaglianza di Bonferroni: se  $\{A_1, \dots, A_n\}$  è un insieme tale che  $A_n \in \mathcal{A}$

$$\forall n \geq 1, \mathbf{P}(A_1 \cap \dots \cap A_n) \geq \mathbf{P}(A_1) + \dots + \mathbf{P}(A_n) - (n - 1) \quad \forall n \geq 2$$

P3 mantiene la relazione d'ordine da  $\mathcal{A}$  a  $\mathbb{R}$ . p5 generalizza P4. P6 implica  $\mathbf{P}(A \cup A') \leq \mathbf{P}(A) + \mathbf{P}(A')$ , proprietà che è detta subadditività. P7 diventa inutile se la serie al lato destro diverge.

**Dim** (P3) Sarà naturale usare ancora P2. Voglio scrivere  $A'$  come unione di insiemi disgiunti: disintegrando grazie ad  $A$ ,  $A' = (A' \cap A) \cup (A' \cap A^c) = A \cup (A' \cap A^c) = A \cup (A' \setminus A)$ .  $A' \cap A, A' \cap A^c$  sono disgiunti e ciascuno dei due appartiene a  $\mathcal{A}$  per ipotesi e A2 e A3. Quindi per P2  $\mathbf{P}(A') = \mathbf{P}(A) + \mathbf{P}(A' \setminus A) \geq \mathbf{P}(A)$  perché  $\mathbf{P}(A' \setminus A) \geq 0$ .

(P4) Vale  $\Omega = A \cup A^c$ , quindi  $\mathbf{P}(A \cup A^c) = 1 = \mathbf{P}(A) + \mathbf{P}(A^c)$ , perché  $A, A^c$  sono disgiunti e  $A, A^c \in \mathcal{A}$ , quindi si usa P2.

(P5) Vale  $A = (A \setminus A') \cup (A' \cap A)$ , quindi, essendo  $A \setminus A', A' \cap A$  eventi disgiunti e appartenenti ad  $\mathcal{A}$ , per P2  $\mathbf{P}(A \setminus A') + \mathbf{P}(A' \cap A) = \mathbf{P}(A)$ .

(P6) Di nuovo, scrivo  $A \cup A'$  come unione di insiemi disgiunti:  $A \cup A' = (A \setminus A') \cup (A \cap A') \cup (A' \setminus A)$ . Per P5  $\mathbf{P}(A \setminus A') = \mathbf{P}(A) - \mathbf{P}(A' \cap A)$ , mentre  $\mathbf{P}(A' \setminus A) = \mathbf{P}(A') - \mathbf{P}(A' \cap A)$ , quindi per P2  $\mathbf{P}(A \cup A') = \mathbf{P}(A \setminus A') + \mathbf{P}(A \cap A') + \mathbf{P}(A' \setminus A) = \mathbf{P}(A) - \mathbf{P}(A' \cap A) + \mathbf{P}(A \cap A') + \mathbf{P}(A') - \mathbf{P}(A' \cap A) = \mathbf{P}(A) + \mathbf{P}(A') - \mathbf{P}(A' \cap A)$ .

(P8) Si procede per induzione. Se  $n = 2$ , per P6  $\mathbf{P}(A_1 \cap A_2) = \mathbf{P}(A_1) + \mathbf{P}(A_2) - \mathbf{P}(A_1 \cup A_2) \geq \mathbf{P}(A_1) + \mathbf{P}(A_2) - 1$ . Ora suppongo  $\mathbf{P}(A_1 \cap \dots \cap A_n) \geq \mathbf{P}(A_1) + \dots + \mathbf{P}(A_n) - (n - 1)$  vera per  $n$  generico e mostro che è vera anche per  $n + 1$ . Sia  $A' = A_1 \cap \dots \cap A_n$ , che appartiene ad  $\mathcal{A}$ :  $\mathbf{P}(A_1 \cap \dots \cap A_n \cap A_{n+1}) = \mathbf{P}(A' \cap A_{n+1}) \geq \mathbf{P}(A') + \mathbf{P}(A_{n+1}) - 1$ , che per ipotesi induttiva è maggiore o uguale di  $\mathbf{P}(A_1) + \dots + \mathbf{P}(A_n) - (n - 1) + \mathbf{P}(A_{n+1}) - 1 = \mathbf{P}(A_1) + \dots + \mathbf{P}(A_n) + \mathbf{P}(A_{n+1}) - (n + 1 - 1)$ . ■

1. Dimostro che, dato  $(\Omega, \mathcal{A}, \mathbf{P})$  e  $A \in \mathcal{A} : \mathbf{P}(A) = 1$ ,  $\mathbf{P}(A \cap B) = \mathbf{P}(B) \forall B \in \mathcal{A}$ . Infatti  $\mathbf{P}(A \cap B) = \mathbf{P}(A) + \mathbf{P}(B) - \mathbf{P}(A \cup B)$ . Ma  $\mathbf{P}(A \cup B) \geq \mathbf{P}(A) = 1$  per P3, essendo  $A \subseteq A \cup B$ : quindi  $\mathbf{P}(A \cup B) = 1$  e

$$\mathbf{P}(A \cap B) = 1 + \mathbf{P}(B) - 1 = \mathbf{P}(B).$$

**Successioni di insiemi inscatolati** Data una  $\sigma$ -algebra  $\mathcal{A}$ , data una famiglia numerabile  $(A_n)_{n \geq 1} : A_n \in \mathcal{A}$

$\forall n$ , si dice che  $A_n$  cresce ad  $A$ , e si scrive  $A_n \uparrow A$ , se vale  $\begin{cases} A_n \subseteq A_{n+1} \\ \bigcup_{n=1}^{+\infty} A_n = A \end{cases}$ ; si dice che  $A_n$  decresce ad  $A$ , e si

scrive  $A_n \downarrow A$ , se vale  $\begin{cases} A_{n+1} \subseteq A_n \\ \bigcap_{n=1}^{+\infty} A_n = A \end{cases}$ . Il fatto che tra insiemi non sia definita una relazione d'ordine totale, a

differenza di quanto accade in  $\mathbb{R}$ , non rende possibile estendere la definizione di limite per successioni alle successioni di insiemi.

$\bigcup_{n=1}^{+\infty} A_n$  è ben definita, senza bisogno di inventarsi una definizione di limite:  $\bigcup_{n=1}^{+\infty} A_n = \{a : \exists i : a \in A_i\}$ .

**Def** Dato un insieme  $\Omega$  e una successione di insiemi  $A_1, A_2, \dots : A_i \subseteq \Omega \forall i$ , si definisce limite superiore degli  $A_n$ , e si indica con  $\limsup_n A_n$ ,  $\bigcap_{n=1}^{+\infty} \bigcup_{k=n}^{+\infty} A_n$ .

**Def** Dato un insieme  $\Omega$  e una successione di insiemi  $A_1, A_2, \dots : A_i \subseteq \Omega \forall i$ , si definisce limite inferiore degli  $A_n$ , e si indica con  $\liminf_n A_n$ ,  $\bigcup_{n=1}^{+\infty} \bigcap_{k=n}^{+\infty} A_n$ .

Se  $(\Omega, \mathcal{A})$  è uno spazio misurabile e  $A_i \in \mathcal{A} \forall i$ , l'evento  $\limsup_n A_n$  può essere espresso a parole come "ω appartiene ad  $A_n$  per infiniti  $n$ "; l'evento  $\liminf_n A_n$  può essere espresso a parole come "ω appartiene definitivamente ad  $A_n$ ", cioè  $\exists k : \omega \in A_n \forall n \geq k$ . Se e. g.  $A_n$  rappresenta la testa in un lancio di moneta,  $\limsup_n A_n$  si verifica se e solo se esce testa infinite volte,  $\liminf_n A_n$  si verifica se e solo se esce sempre testa da un certo punto in poi.

#### Teo 2.4 (continuità della probabilità)

Hp:  $(\Omega, \mathcal{A}, \mathbf{P})$  è uno spazio di probabilità,  $(A_n)_{n \geq 1} : A_n \in \mathcal{A} \forall n, A_n \uparrow A$

$$\text{Ts: (P9)} \quad \mathbf{P}\left(\bigcup_{n=1}^{+\infty} A_n\right) = \lim_{n \rightarrow +\infty} \mathbf{P}(A_n)$$

Hp:  $(\Omega, \mathcal{A}, \mathbf{P})$  è uno spazio di probabilità,  $(A_n)_{n \geq 1} : A_n \in \mathcal{A} \forall n, A_n \downarrow A$

$$\text{Ts: (P10)} \quad \mathbf{P}\left(\bigcap_{n=1}^{+\infty} A_n\right) = \lim_{n \rightarrow +\infty} \mathbf{P}(A_n)$$

La tesi è ben posta perché  $\bigcup_{n=1}^{+\infty} A_n \in \mathcal{A}$ : ogni  $\sigma$ -algebra è chiusa per limiti crescenti.

**Dim (P9)** Devo scrivere  $A$  come unione di insiemi disgiunti. Costruisco  $(B_n)_{n \geq 1} : B_1 = A_1, B_2 = A_2 \setminus A_1, \dots, B_n = A_n \setminus A_{n-1}$  con  $n \geq 2$ : per  $(B_n)_{n \geq 1}$  vale che  $\bigcup_{n=1}^{+\infty} B_n = A$ ,  $\bigcup_{k=1}^n B_k = A_k$ ,  $B_h \cap B_k = \emptyset \forall h \neq k$ . Allora per la  $\sigma$ -additività sui  $B_n$ , cioè A2, vale  $\mathbf{P}(A) = \mathbf{P}\left(\bigcup_{n=1}^{+\infty} B_n\right) = \sum_{n=1}^{+\infty} \mathbf{P}(B_n) = \lim_{k \rightarrow +\infty} \sum_{n=1}^k \mathbf{P}(B_n) = \lim_{k \rightarrow +\infty} \mathbf{P}(A_k)$  usando la successione delle somme parziali, e per P2 su  $A_k$ .

(P10) Uso le leggi di De Morgan. Vale  $\bigcap_{n=1}^{+\infty} A_n = \left(\bigcup_{n=1}^{+\infty} A_n^c\right)^c$ , per cui  $\mathbf{P}\left(\bigcap_{n=1}^{+\infty} A_n\right) = \mathbf{P}\left(\bigcup_{n=1}^{+\infty} A_n^c\right)^c = 1 - \mathbf{P}\left(\bigcup_{n=1}^{+\infty} A_n^c\right) = 1 - \lim_{n \rightarrow +\infty} \mathbf{P}(A_n^c) = \lim_{n \rightarrow +\infty} \mathbf{P}(A_n)$ , usando P4 e P9. ■

**Teo (lemma di Borel-Cantelli)\***

Hp:  $(\Omega, \mathcal{A}, \mathbf{P})$  è uno spazio di probabilità,  $(A_n)_{n \geq 1} : A_n \in \mathcal{A} \forall n, \sum_{n=1}^{+\infty} \mathbf{P}(A_n) < +\infty$

$$\text{Ts: (P9) } \mathbf{P}\left(\limsup_n A_n\right) = 0$$

**Dim** Per definizione  $\limsup_n A_n = \bigcap_{n=1}^{+\infty} \bigcup_{k=n}^{+\infty} A_k$ . Ponendo  $B_n = \bigcup_{k=n}^{+\infty} A_k$ , i  $B_n$  sono una successione decrescente, quindi per continuità e subadditività della probabilità  $\mathbf{P}\left(\bigcap_{n=1}^{+\infty} B_n\right) = \lim_{n \rightarrow +\infty} \mathbf{P}\left(\bigcup_{k=n}^{+\infty} A_k\right) \leq \lim_{n \rightarrow +\infty} \sum_{k=n}^{+\infty} \mathbf{P}(A_k) = \lim_{n \rightarrow +\infty} \left(\sum_{k=1}^{+\infty} \mathbf{P}(A_k) - \sum_{k=1}^{n-1} \mathbf{P}(A_k)\right) = \lim_{n \rightarrow +\infty} 0 = 0$ . ■

### 2.3.1 Assegnazione di una probabilità

E' ora naturale chiedersi come si fa, dato uno spazio misurabile  $(\Omega, \mathcal{A})$ , ad assegnarvi una probabilità  $\mathbf{P}$ . Partiamo dal caso semplice di assegnazione di probabilità, con  $\Omega$  discreto.

**Teo 3.1 (assegnazione di probabilità su  $\Omega$  discreto)**

Hp:  $\Omega$  è al più numerabile

Ts: (i)  $\mathbf{P} : 2^\Omega \rightarrow [0, 1]$  è caratterizzata da  $p : \Omega \rightarrow [0, 1]$  tale che  $p(\omega) = \mathbf{P}(\{\omega\})$

(ii) data  $p : \Omega \rightarrow \mathbb{R}, \exists ! \mathbf{P} : 2^\Omega \rightarrow [0, 1] : \mathbf{P}(\{\omega\}) = p(\omega) \forall \omega \in \Omega$

$$\iff (\text{A}) p(\omega) \geq 0 \forall \omega \in \Omega \text{ e } (\text{B}) \sum_{\omega \in \Omega} p(\omega) = 1$$

$\mathbf{P}$  è considerata definita su  $2^\Omega$  a priori perché se  $\Omega$  è numerabile classicamente si sceglie  $2^\Omega$  come  $\sigma$ -algebra su  $\Omega$ .

(i), cioè che  $\mathbf{P}$  è caratterizzata dal suo valore sugli eventi elementari, significa che - data una certa  $p$  - assumendo che  $\mathbf{P}$  in corrispondenza degli eventi elementari (atomi)  $\omega$  valga  $p(\omega)$ , si può calcolare  $\mathbf{P}(A) \forall A \in 2^\Omega$ , cioè  $p$  determina univocamente  $\mathbf{P}$ ; è vero, ma poco significativo, anche il viceversa. L'utilità di (i) sta nel fatto che si può costruire  $\mathbf{P}$  assegnando una probabilità a ogni evento elementare, lavorando quindi con un dominio più semplice.

La funzione  $p : \Omega \rightarrow [0, 1]$ , che soddisfa (A) e (B), è detta distribuzione di probabilità.

**Dim** (i) Ogni  $A \in 2^\Omega$  può essere scritto in modo unico come unione di eventi elementari, cioè  $A = \bigcup_{\omega \in A} \{\omega\}$ , perché  $A$  è un sottinsieme di  $\Omega$  e quindi unione di elementi di  $\Omega$ . Essendo  $\Omega$  al più numerabile, anche  $A$  è al più numerabile e  $\bigcup_{\omega \in A} \{\omega\}$  è un'unione finita o infinita numerabile di eventi disgiunti: allora per A2 e per la finita additività  $\mathbf{P}(A) = \mathbf{P}\left(\bigcup_{\omega \in A} \{\omega\}\right) = \sum_{\omega \in A} \mathbf{P}(\{\omega\})$ , che è una somma se  $A$  ha cardinalità finita, una serie se  $A$  ha cardinalità infinita. Poiché si sa come calcolare  $\mathbf{P}(A) \forall A \in 2^\Omega$ ,  $\mathbf{P}$  è univocamente determinata.



(ii) Mostro l'implicazione da sinistra a destra. Se  $\mathbf{P}(\{\omega\}) = p(\omega) \forall \omega \in \Omega$ , poiché  $\mathbf{P}$  è una funzione a valori in  $[0, 1]$ , dovrà essere  $p \in [0, 1]$  e quindi  $p(\omega) \geq 0 \forall \omega$ . Inoltre  $\sum_{\omega \in \Omega} p(\omega) = \sum_{\omega \in \Omega} \mathbf{P}(\{\omega\}) = \mathbf{P}(\bigcup_{\omega \in \Omega} \{\omega\}) = \mathbf{P}(\Omega) = 1$  perché  $\Omega$  è al più numerabile, quindi si applicano A2, la finita o  $\sigma$  additività e A1.

Mostro l'implicazione da destra a sinistra. Sia  $p(\omega) \geq 0 \forall \omega \in \Omega$  e  $\sum_{\omega \in \Omega} p(\omega) = 1$ . Dato un qualsiasi sottinsieme di  $\Omega$   $A \in 2^\Omega$ , poiché può essere scritto come  $\bigcup_{\omega \in A} \{\omega\}$ , definisco  $\mathbf{P}(A) = \mathbf{P}(\bigcup_{\omega \in A} \{\omega\}) := \sum_{\omega \in A} p(\omega)$ . Verifico che soddisfa A1 e A2:  $\mathbf{P}(\Omega) = \sum_{\omega \in \Omega} p(\omega) = 1$ , quindi va bene A1; A2 si può dimostrare grazie al fatto che si possono riordinare i termini di una serie senza modificarne la somma. ■

1. Considero un'urna che contiene  $N$  palline numerate da 1 a  $N$ , un infante bendato che ne pesca una pallina e osserva il numero. Definisco il modello probabilistico dell'esperimento:  $\Omega = \{1, \dots, N\}$ ,  $\mathcal{A} = 2^\Omega$  dato che  $\Omega$  è finito. Ora uso il teorema (ii): scelgo  $p$  tale che  $p(\omega) \geq 0 \forall \omega \in \Omega$  e  $\sum_{n=1}^N p(\omega_n) = 1$ , così da individuare  $\mathbf{P}$ . Avrei infinite possibili scelte: ma nelle realtà, essendo l'infante bendato, è sensato assumere che tutte le palline siano equiprobabili, per cui si avrà la stessa  $p(\omega) = c$  per ogni  $\omega \in \Omega$ . Per trovarne il valore si risolve il sistema 
$$\begin{cases} p(\omega) = c \geq 0 \\ \sum_{\omega \in \Omega} p(\omega) = 1 \end{cases},$$
 da cui si ricava  $Nc = 1$ , cioè  $c = \frac{1}{N}$ : l'ipotesi che ogni evento elementare sia equiprobabile ha portato a  $\mathbf{P}(\{\omega\}) = \frac{1}{N} \forall \omega \in \Omega$ . Di conseguenza, se  $A$  è un qualsiasi sottinsieme di  $\Omega$  con cardinalità  $|A|$ ,  $A = \bigcup_{i=1}^{|A|} \{\omega_i\}$ ,  $\mathbf{P}(A) = \sum_{i=1}^{|A|} \mathbf{P}(\{\omega_i\}) = \frac{|A|}{N} = \frac{|A|}{|\Omega|}$ : l'approccio assiomatico alla probabilità ci ha riportato alla definizione classica di probabilità come casi favorevoli fratto casi possibili. Il modello di probabilità  $(\Omega, 2^\Omega, \mathbf{P})$  visto, con  $\Omega$  finito e  $\mathbf{P}$  tale che  $\mathbf{P}(\{\omega_i\}) = \frac{1}{N} \forall i$ , è detto modello di spazio equiprobabile, e la  $p$  scelta è detta distribuzione uniforme.

**Def** Dato  $(\Omega, \mathcal{A}, \mathbf{P})$  spazio di probabilità, se  $\Omega$  è finito,  $\mathcal{A} = 2^\Omega$  e  $\mathbf{P}$  è tale che  $\mathbf{P}(\{\omega\}) = \frac{1}{|\Omega|} \forall \omega \in \Omega$ ,  $(\Omega, \mathcal{A}, \mathbf{P})$  si dice spazio probabilistico discreto e uniforme.

Per 3.1  $\mathbf{P}$  è ben definita.

1. Sia  $\Omega = \{0, \dots, n\}$ ,  $\mathcal{A} = 2^\Omega$ .  $|\mathcal{A}| = 2^{n+1}$ . Scelgo  $p : \Omega \rightarrow [0, 1]$  tale che  $p(\omega) \geq 0 \forall \omega \in \Omega$  e  $\sum_{\omega \in \Omega} p(\omega) = 1$ , così da individuare  $\mathbf{P}$ . Dato  $p \in (0, 1)$  assegnato, definisco  $p(k) = \binom{n}{k} p^k (1-p)^{n-k}$ : è vero che  $p(k) \geq 0 \forall k \in \Omega$ , e inoltre  $\sum_{k=0}^n \binom{n}{k} p^k (1-p)^{n-k} = (p + 1 - p)^n = 1$ , perché questa somma è lo sviluppo del binomio di Newton. La  $p$  scelta è detta distribuzione binomiale  $(Bi(n, p))$ .
2. Sia  $\Omega = \mathbb{N}$ ,  $\mathcal{A} = 2^\mathbb{N}$ ,  $\lambda > 0$ . Scelgo  $p : \Omega \rightarrow [0, 1]$  tale che  $p(\omega) \geq 0 \forall \omega \in \Omega$  e  $\sum_{\omega \in \Omega} p(\omega) = 1$ , così da individuare  $\mathbf{P}$ . Definisco  $p(k) = e^{-\lambda} \frac{\lambda^k}{k!} \forall k \in \mathbb{N}$ : è vero che  $p(k) \geq 0 \forall k \in \Omega$ , e inoltre  $\sum_{k=0}^{+\infty} e^{-\lambda} \frac{\lambda^k}{k!} = e^{-\lambda} e^\lambda = 1$  perché  $\sum_{k=0}^{+\infty} \frac{\lambda^k}{k!}$  è la serie di Taylor di  $e^\lambda$ . Il modello di probabilità  $(\mathbb{N}, 2^\mathbb{N}, \mathbf{P})$  visto è detto modello di Poisson e la  $p$  scelta è detta distribuzione di Poisson  $(poiss(\lambda))$ .

Ora si passa all'assegnazione di una probabilità con  $\Omega$  qualsiasi.

**Teo 3.2 (assegnazione di probabilità su  $\Omega$  qualsiasi)**

Hp:  $\Omega$  è un insieme qualsiasi,  $(C_k)_{k \in I}$  è una partizione discreta di  $\Omega$ ,  $\mathcal{A} = \sigma((C_k)_k : k \in I)$

Ts: (i)  $\mathbf{P} : \mathcal{A} \rightarrow [0, 1]$  è univocamente determinata da  $p : I \rightarrow [0, 1]$  tale che  $p(n) = \mathbf{P}(C_n)$

(ii) data  $p : I \rightarrow \mathbb{R}$ ,  $\exists ! \mathbf{P} : \mathcal{A} \rightarrow [0, 1] : \mathbf{P}(C_n) = p(n) \quad \forall n \in I$

$$\iff (A) \ p(n) \geq 0 \quad \forall n \in I \text{ e } (B) \sum_{n \in I} p(n) = 1$$

$(C_k)_{k \in I}$  partizione discreta di  $\Omega$  significa che è  $(C_k)_{k \geq 1}$  è una famiglia al più numerabile di insiemi disgiunti la cui unione è  $\Omega$ .  $\mathcal{A}$  in questo caso non può essere  $2^\Omega$ , che potrebbe essere "troppo grande", perché  $\Omega$  è qualsiasi. Si è visto che  $\sigma((C_k)_k : k \in I) = \{A \subseteq \Omega : A = \bigcup_{k \in J} C_k, J \subseteq I\}$ . Gli eventi a cui è assegnata una probabilità sono dunque solo quelli del tipo  $A = \bigcup_{k \in J} C_k, J \subseteq I$ . 3.2 diventa 3.1 nel caso particolare in cui  $\Omega$  è discreto e, se  $\Omega$  è infinito numerabile,  $I = \mathbb{N}$  e  $C_k = \{\omega_k\} \quad \forall k$ .

C'è un'evidente analogia con il teorema precedente: essendo  $\Omega$  qualsiasi e non necessariamente discreto, si definisce la probabilità non sugli atomi ma su ogni elemento di una partizione discreta. Se  $\Omega$  non è discreto si compie dunque un compromesso: si rinuncia a "vedere" alcuni eventi, limitandosi ad assegnare una probabilità agli elementi della partizione discreta, ma in questo modo ci si riconduce al caso discreto, riuscendo a scrivere ogni evento come unione al più numerabile di insiemi e calcolando ogni probabilità mediante una funzione definita su un insieme molto semplice quale  $I$ .

**Dim** (i) E' noto che per 1.6 che se  $\mathcal{C}$  è una partizione discreta di  $\Omega$  allora  $\sigma(\mathcal{C}) = \{A \subseteq \Omega : A = \bigcup_{k \in J} C_k \text{ al variare di } J \subseteq I\}$ , cioè ogni elemento della  $\sigma$ -algebra scelta può essere scritto come unione finita o infinita numerabile dei  $C_k$ . Quindi  $\forall A \in \mathcal{A}$  vale, per  $\sigma$ -additività,  $\mathbf{P}(A) = \mathbf{P}(\bigcup_{k \in J} C_k) = \sum_{k \in J} \mathbf{P}(C_k) = \sum_{k \in J} p(k)$ , e  $\mathbf{P}$  è univocamente determinata.

(ii) Mostro l'implicazione da sinistra a destra. Se  $\mathbf{P}(C_n) = p(n) \quad \forall n \in I$ , poiché  $\mathbf{P}$  è una funzione a valori in  $[0, 1]$ , dovrà essere  $p \in [0, 1]$  e quindi  $p(n) \geq 0 \quad \forall n \in I$ . Inoltre  $\sum_{n \in I} p(n) = \sum_{n \in I} \mathbf{P}(C_n) = \mathbf{P}(\bigcup_{n \in I} C_n) = \mathbf{P}(\Omega) = 1$  per definizione di partizione.

Mostro l'implicazione da destra a sinistra. Sia  $p(n) \geq 0 \quad \forall n \in I$  e  $\sum_{n \in I} p(n) = 1$ . Dato un qualsiasi sottinsieme di  $\Omega$   $A \in \mathcal{A}$ , poiché può essere scritto come  $\bigcup_{k \in J} C_k$ , definisco  $\mathbf{P}(A) = \mathbf{P}(\bigcup_{k \in J} C_k) := \sum_{k \in J} p(k)$ . Verifico che soddisfa A1 e A2:  $\mathbf{P}(\Omega) = \sum_{n \in I} p(n) = 1$ , quindi va bene A1; A2 si può dimostrare grazie al fatto che si possono riordinare i termini di una serie senza modificarne la somma. ■

### 3 Probabilità condizionata

1. Lancio uno dopo l'altro due dadi non truccati e voglio calcolare la probabilità dell'evento  $A$  = "la somma dei due risultati è 12". La terna di Kolmogorov dell'esperimento è  $\Omega = \{1, \dots, 6\} \times \{1, \dots, 6\}$ ,  $\mathcal{A} = 2^\Omega$ , e - poiché è sensato che ogni coppia sia equiprobabile - si può usare la distribuzione uniforme, per cui  $\mathbf{P}(A) = \frac{|A|}{|\Omega|}$ :  $|\Omega| = 36$ ,  $|A| = |\{(h, k) : (h, k) \in \Omega \wedge h + k = 12\}| = |\{(6, 6)\}| = 1$ , per cui  $\mathbf{P}(A) = \frac{1}{36}$ .

Ora considero l'evento  $E$  = "il primo lancio dà 6":  $\mathbf{P}(A)$ , con questa informazione aggiuntiva, coincide con la probabilità che il secondo lancio dia 6, quindi  $\frac{1}{6}$ : è aumentata, perché ora i casi possibili sono solo le coppie sulla prima riga.

(6, 1)	(6, 2)	(6, 3)	(6, 4)	(6, 5)	(6, 6)
(5, 1)	(5, 2)	(5, 3)	(5, 4)	(5, 5)	(5, 6)
(4, 1)	(4, 2)	(4, 3)	(4, 4)	(4, 5)	(4, 6)
(3, 1)	(3, 2)	(3, 3)	(3, 4)	(3, 5)	(3, 6)
(2, 1)	(2, 2)	(2, 3)	(2, 4)	(2, 5)	(2, 6)
(1, 1)	(1, 2)	(1, 3)	(1, 4)	(1, 5)	(1, 6)

Dunque  $\frac{1}{6} = \frac{|A|}{|E|}$ ; dato che  $A \subseteq E$ , vale  $\mathbf{P}(A) = \mathbf{P}(A \cap E)$ , dunque  $\frac{1}{6} = \frac{|A \cap E|}{|E|} = \frac{|A \cap E|}{|\Omega|} \frac{|\Omega|}{|E|} = \frac{P(A \cap E)}{P(E)}$ .

Questo esempio rende ragionevole la seguente definizione.

**Def 3.3** Dato  $(\Omega, \mathcal{A}, \mathbf{P})$  spazio di probabilità,  $E \in \mathcal{A}$  evento tale che  $\mathbf{P}(E) > 0$  e  $A \in \mathcal{A}$  evento qualsiasi, si dice probabilità condizionata di  $A$  dato  $E$ , e si indica con  $\mathbf{P}(A|E)$ , il numero  $\frac{\mathbf{P}(A \cap E)}{\mathbf{P}(E)}$ .

Il significato della definizione è il calcolo della probabilità di un evento quando si è in possesso di un'informazione aggiuntiva.

La definizione è ben posta perché si richiede  $P(E) > 0$  (cioè che  $E$  sia un evento non trascurabile).

Evidentemente il ruolo di  $A$  e  $E$  non può essere scambiato.

Se  $E \subseteq A$ , l'intuizione dice che  $\mathbf{P}(A|E) = 1$ , perché  $E$  implica  $A$ : questo è in accordo con la definizione, perché  $\mathbf{P}(A|E) = \frac{\mathbf{P}(A \cap E)}{\mathbf{P}(E)} = \frac{\mathbf{P}(E)}{\mathbf{P}(E)} = 1$ . Se invece  $A \cap E = \emptyset$ , l'intuizione dice che  $\mathbf{P}(A|E) = 0$ , perché  $A$  e  $E$  sono incompatibili: questo è in accordo con la definizione, perché  $\mathbf{P}(A|E) = \frac{\mathbf{P}(A \cap E)}{\mathbf{P}(E)} = \frac{0}{\mathbf{P}(E)} = 0$ .

La probabilità condizionata è utile in generale quando un esperimento è composto da più fasi:

1. Se lancio due dadi e  $A$  = "ad almeno uno dei due lanci esce un numero pari" e  $E$  = "al primo lancio esce un numero pari", evidentemente  $E \subseteq A$  e  $\mathbf{P}(A|E) = 1$ . NB: un evento è un elemento di  $\mathcal{A}$ , quindi necessariamente una coppia, non può essere relativo a solo una fase dell'esperimento.

2. Considerando lo stesso esperimento di prima, sia  $F$  = "uno dei due lanci ha dato 6" (si intende almeno uno).

In tal caso, essendo sempre in regime equiprobabile,  $\mathbf{P}(A|F) = \frac{\mathbf{P}(A \cap F)}{\mathbf{P}(F)} = \frac{|A \cap F|}{|\Omega|} \cdot \frac{|\Omega|}{|F|} = \frac{1}{36} \cdot \frac{36}{11} = \frac{1}{11}$ . Noto che  $\frac{1}{36} < \frac{1}{11} < \frac{1}{6}$ : al crescere della quantità di informazioni possedute la probabilità aumenta, infatti nel primo caso non si sa niente, nel secondo si sa cosa è successo in uno dei due lanci, nel terzo si sa cosa è successo nel primo lancio.

**Prop 3.4 (la probabilità condizionata è una misura di probabilità nel suo primo argomento)**

Hp:  $(\Omega, \mathcal{A}, \mathbf{P})$  spazio di probabilità,  $E \in \mathcal{A}$ ,  $\mathbf{P}(E) > 0$ ,  $A \in \mathcal{A}$  fissato

Ts:  $\mathbf{P}_E : \mathcal{A} \rightarrow [0, 1]$ ,  $\mathbf{P}_E(A) = \mathbf{P}(A|E)$  è una misura di probabilità su  $\mathcal{A}$

$\mathbf{P}_E$  è inoltre assolutamente continua rispetto a  $\mathbf{P}$ .

**Dim** Affinché sia vera la tesi  $\mathbf{P}_E$  deve soddisfare A1, A2.  $\mathbf{P}_E(\Omega) = \mathbf{P}(\Omega|E) = \frac{\mathbf{P}(\Omega \cap E)}{\mathbf{P}(E)} = \frac{\mathbf{P}(E)}{\mathbf{P}(E)} = 1$ , quindi vale A1. Se  $(A_n)_{n \geq 1}$  è una famiglia numerabile di eventi disgiunti,  $\mathbf{P}_E\left(\bigcup_{n=1}^{+\infty} A_n\right) = \mathbf{P}\left(\bigcup_{n=1}^{+\infty} A_n | E\right) = \frac{\mathbf{P}\left(\left(\bigcup_{n=1}^{+\infty} A_n\right) \cap E\right)}{\mathbf{P}(E)} = \frac{\mathbf{P}\left(\bigcup_{n=1}^{+\infty} (A_n \cap E)\right)}{\mathbf{P}(E)}$  per la proprietà distributiva dell'intersezione rispetto all'unione; poiché  $\mathbf{P}$  è una misura di probabilità (e, essendo gli  $A_n$  disgiunti, anche gli  $A_n \cap E$  sono disgiunti) si ottiene  $\frac{\sum_{n=1}^{+\infty} \mathbf{P}(A_n \cap E)}{\mathbf{P}(E)} = \sum_{n=1}^{+\infty} \frac{\mathbf{P}(A_n \cap E)}{\mathbf{P}(E)} = \sum_{n=1}^{+\infty} \mathbf{P}_E(A_n)$ . ■

Quindi valgono tutte le solite proprietà della probabilità, e. g.  $\mathbf{P}_E(A^c) = \mathbf{P}(A^c|E) = 1 - \mathbf{P}(A|E)$ ,  $\mathbf{P}_E(A \cup A') = \mathbf{P}(A \cup A'|E) = \mathbf{P}_E(A) + \mathbf{P}_E(A') - \mathbf{P}_E(A \cap A')$ . Non è invece vero che  $\mathbf{P}_A(E) = \mathbf{P}(A|E)$  è una misura di probabilità: infatti, dati  $A, B$  disgiunti,  $\mathbf{P}(A|A \cup B) = \frac{\mathbf{P}(A)}{\mathbf{P}(A \cup B)} \neq \mathbf{P}(A|A) + \mathbf{P}(A|B) = 1 + \mathbf{P}(A|B)$ .

1. Dato  $(\Omega, \mathcal{A}, \mathbf{P})$  spazio di probabilità,  $A, B, C \in \mathcal{A}$ ,  $\mathbf{P}(A \cap B) > 0$ , vale  $\mathbf{P}_A(C|B) = \mathbf{P}_{A \cap B}(C)$ . Infatti

$$\mathbf{P}_A(C|B) = \frac{\mathbf{P}_A(C \cap B)}{\mathbf{P}_A(B)} = \frac{\mathbf{P}(C \cap B|A)}{\mathbf{P}(B|A)} = \frac{\mathbf{P}(C \cap B \cap A)}{\mathbf{P}(B \cap A) \mathbf{P}(A)}, \text{ mentre } \mathbf{P}_{A \cap B}(C) = \mathbf{P}(C|A \cap B) = \frac{\mathbf{P}(C \cap B \cap A)}{\mathbf{P}(A \cap B)}.$$

Nell'ipotesi  $\mathbf{P}(E) > 0$ , l'uguaglianza  $\mathbf{P}(A \cap E) = \mathbf{P}(A|E) \mathbf{P}(E)$  fa intuire che il concetto di probabilità condizionata può essere usato anche per assegnare un modello, ad esempio per esperimenti che si svolgono in più fasi, come è spiegato nel contesto seguente.

Se ho due partizioni discrete  $(A_n)_{n \geq 1}, (B_n)_{n \geq 1}$  di  $\Omega$ , ciascuna formata da eventi non trascurabili, allora anche  $\{A_h \cap B_k \text{ al variare di } h, k \geq 1\}$  è una partizione discreta di  $\Omega$ : tutte le possibili intersezioni tra gli insiemi delle due partizioni determinano una partizione più fine di  $\Omega$ . Supponendo che le partizioni siano finite, vale  $A_i = \bigcup_{h=1}^k (A_i \cap B_h) = A_i \cap \left(\bigcup_{h=1}^k B_h\right) = A_i \cap \Omega$ .

**Teo 4.1 (assegnazione di probabilità su  $\Omega$  qualsiasi II)**

Hp:  $\Omega$  è un insieme,  $(A_n)_{n \geq 1}, (B_n)_{n \geq 1}$  sono partizioni discrete di  $\Omega$ ,  $(A_n)_{n \geq 1}$  eventi

non trascurabili,  $\mathcal{A} = \sigma((A_h \cap B_k) : h, k \geq 1)$ ,  $p : I \rightarrow [0, 1] : p(n) > 0 \forall n \geq 1$  e

$$\sum_{n=1}^{+\infty} p(n) = 1, \forall n \text{ fissato } q(k|n) : J \rightarrow [0, 1] : q(k|n) \geq 0 \forall k, \sum_{k=1}^{+\infty} q(k|n) = 1$$

Ts:  $\exists ! \mathbf{P} : \mathcal{A} \rightarrow [0, 1] : \mathbf{P}(A_n) = p(n)$  e  $\mathbf{P}(B_k|A_n) = q(k|n) \forall k, \forall n$

cosa significa non trascurabili? non vuoti? o di probabilità non nulla? ma rispetto a quale prob?

Il teorema spiega come modellizzare un esperimento a partire dalla probabilità condizionata. Dalla tesi segue che avendo  $p$  e  $q$  si può costruire una probabilità sulla  $\sigma$ -algebra generata dalla partizione più fine, perché  $\mathbf{P}(B_k|A_n) = \frac{\mathbf{P}(A_n \cap B_k)}{\mathbf{P}(A_n)}$ , per cui  $\mathbf{P}(A_n \cap B_k) = \mathbf{P}(B_k|A_n) \mathbf{P}(A_n)$ : si vedono solo tutte le possibili unioni dei "quadrati", cioè degli elementi della partizione fine. (NB: lei non aveva messo hp  $q(k|n) \geq 0 \forall k$ ). La sommatoria è su tutti i naturali e non su  $I$  generico perché si stanno considerando partizioni infinite. L'implicazione opposta è ovvia.

**Dim** Voglio usare il teorema 3.2. Se quindi definisco  $p(n, k) : p(n, k) \geq 0 \forall n, k$  e  $\sum_{n, k \geq 1} p(n, k) = 1$ , allora  $\exists ! \mathbf{P} : \sigma((A_h \cap B_k) : h, k \geq 1) \rightarrow [0, 1] : \mathbf{P}(A_h \cap B_k) = p(n, k) \forall n \in I$ . Fisso allora  $p(n, k) = p(n) q(k|n)$ :  $p(n, k) \geq 0 \forall n, \forall k$  è vero per ipotesi, mentre  $\sum_{n, k \geq 1} p(n, k) = \sum_{n, k \geq 1} p(n) q(k|n) = \sum_n p(n) \sum_k q(k|n) = \sum_n p(n) = 1$  perché le serie sono a termini positivi e per le ipotesi. Allora  $\exists ! \mathbf{P} : \mathcal{A} \rightarrow [0, 1] : \mathbf{P}(A_n \cap B_k) = p(n, k)$ . Ne segue che  $\mathbf{P}(A_n) = \mathbf{P}\left(\bigcup_{k=1}^{+\infty} (A_n \cap B_k)\right) = \sum_{k=1}^{+\infty} \mathbf{P}(A_n \cap B_k) = \sum_{k=1}^{+\infty} p(n) q(k|n) = p(n) \sum_{k=1}^{+\infty} q(k|n) = p(n)$  e  $\mathbf{P}(B_k|A_n) = \frac{\mathbf{P}(A_n \cap B_k)}{\mathbf{P}(A_n)} = \frac{p(n) q(k|n)}{p(n)} = q(k|n)$  per definizione di probabilità condizionata e per quanto mostrato sopra. ■

Questo teorema permette quindi di assegnare una probabilità che coglie bene l'aspetto modellistico che ci interessa.

1. Considero un'urna con  $n \geq 2$  palline nere e  $b \geq 2$  palline bianche: faccio due estrazioni senza reimmissione e osservo il colore delle palline estratte. Lo spazio campionario è  $\Omega = \{(b, b), (b, n), (n, b), (n, n)\}$ . Sia  $N_1$  l'evento "al lancio 1 è uscita una pallina nera" ( $\{(n, n), (n, b)\}$ ), e analogamente  $B_1 = \{(b, b), (b, n)\}$ ,  $N_2, B_2$ .  $\{B_1, N_1\}$  è una partizione di  $\Omega$  (cioè gli  $(A_n)$ ), e  $\{B_2, N_2\}$  è un'altra partizione di  $\Omega$  (cioè i  $(B_n)$ ). Qual è  $\mathbf{P}(B_1 \cap B_2)$ ? Posso usare il teorema sopra per risalire a  $\mathbf{P}$  definita sulla  $\sigma$ -algebra generata da  $\{B_h \cap N_k : h, k \in \{1, 2\}\}$ : è sufficiente definire  $p(n, k)$  in corrispondenza degli elementi della partizione. Trovandosi in un caso in cui gli esiti sono tutti equiprobabili, si usa la distribuzione uniforme, per cui  $p_1 = \mathbf{P}(B_1) = \frac{b}{n+b}$ ,  $p_2 = \mathbf{P}(N_1) = \frac{n}{n+b}$ ,

$$q_{1|1} = \mathbf{P}(B_2|B_1) = \frac{b-1}{n+b-1}, q_{2|1} = \mathbf{P}(N_2|B_1) = \frac{n}{n+b-1}, q_{1|2} = \mathbf{P}(B_2|N_1) = \frac{b}{n+b-1}, q_{2|2} = \mathbf{P}(N_2|N_1) = \frac{n-1}{n+b-1}. \text{ Allora } \mathbf{P}(B_1 \cap B_2) = p_1 q_{1|1} = \frac{b}{n+b} \frac{b-1}{n+b-1}.$$

Come fare per calcolare  $\mathbf{P}(N_2)$  o  $\mathbf{P}(N_1|N_2)$  nell'esempio precedente? Il seguente teorema fornisce una risposta.

**Teo 4.2 (disintegrazione e probabilità totali)**

Hp:  $(\Omega, \mathcal{A}, \mathbf{P})$  è uno spazio di probabilità,  $(E_n)_{n \geq 1}$  è una partizione discreta di  $\Omega$  con  $E_n \in \mathcal{A} \forall n$

$$\text{Ts: (i) (formula di disintegrazione)} \forall A \in \mathcal{A} \mathbf{P}(A) = \sum_{n=1}^{+\infty} \mathbf{P}(A \cap E_n)$$

$$\text{(ii) (formula delle probabilità totali) se inoltre } \mathbf{P}(E_n) > 0 \forall n, \forall A \in \mathcal{A} \mathbf{P}(A) = \sum_{n=1}^{+\infty} \mathbf{P}(A|E_n) \mathbf{P}(E_n)$$

La (i) è detta formula di disintegrazione perché si sta disintegrando  $A$  con gli  $E_n$ . La (ii) serve quando le condizioni dell'esperimento sono incerte.

**Dim** (i) Poiché  $(E_n)_{n \geq 1}$  è una partizione discreta di  $\Omega$ ,  $A = \bigcup_{n=1}^{+\infty} (A \cap E_n)$ . Gli  $A \cap E_n$  sono eventi disgiunti perché gli  $E_n$  lo sono, quindi per la  $\sigma$ -additività (A2)  $\mathbf{P}(A) = \mathbf{P}\left(\bigcup_{n=1}^{+\infty} (A \cap E_n)\right) = \sum_{n=1}^{+\infty} \mathbf{P}(A \cap E_n)$ .

(ii) Se  $\mathbf{P}(E_n) > 0 \forall n$ , è ben definita  $\forall n \mathbf{P}(A|E_n) = \frac{\mathbf{P}(A \cap E_n)}{\mathbf{P}(E_n)}$ , e usando (i) si ha  $\mathbf{P}(A) = \sum_{n=1}^{+\infty} \mathbf{P}(A \cap E_n) = \sum_{n=1}^{+\infty} \mathbf{P}(A|E_n) \mathbf{P}(E_n)$ . ■

1. Nell'esempio precedente,  $\mathbf{P}(N_2) = \sum_{n=1}^2 \mathbf{P}(N_2|E_n) \mathbf{P}(E_n) = \mathbf{P}(N_2|N_1) \mathbf{P}(N_1) + \mathbf{P}(N_2|B_1) \mathbf{P}(B_1) = \frac{n-1}{n+b-1} \frac{n}{n+b} + \frac{n}{n+b-1} \frac{b}{n+b} = \frac{n}{n+b-1} \frac{n-1+b}{n+b} = \frac{n}{n+b}$ : è la stessa probabilità di estrarre una pallina nera alla prima estrazione.

Infatti, se non so cosa è uscito alla prima estrazione, è come se non avessi tolto alcuna pallina (risultato controintuitivo).

In realtà, nel teorema le ipotesi possono essere indebolite: non è necessario avere una partizione.

Hp:  $(\Omega, \mathcal{A}, \mathbf{P})$  è uno spazio di probabilità,  $(E_n)_{n \geq 1}$  è una famiglia di sottinsiemi di  $\Omega$

al più numerabile tale che  $E_n \in \mathcal{A} \forall n$ ,  $\mathbf{P}(E_h \cap E_k) = 0 \forall h \neq k$ ,  $\mathbf{P}\left(\bigcup_{n=1}^{+\infty} E_n\right) = 1$

$$\text{Ts: (i) (formula di disintegrazione)} \forall A \in \mathcal{A} \mathbf{P}(A) = \sum_{n=1}^{+\infty} \mathbf{P}(A \cap E_n)$$

$$\text{(ii) (formula delle probabilità totali) se inoltre } \mathbf{P}(E_n) > 0 \forall n, \mathbf{P}(A) = \sum_{n=1}^{+\infty} \mathbf{P}(A|E_n) \mathbf{P}(E_n)$$

Infatti, dato  $A$ , affinché sia  $\mathbf{P}(A) = 1$  è sufficiente, ma non necessario, che  $A = \Omega$ , e affinché sia  $\mathbf{P}(A) = 0$  è sufficiente, ma non necessario, che  $A = \emptyset$ .

**Dim** (i) Disintegro  $A$  usando gli  $E_n$ :  $A = \left(A \cap \bigcup_{n=1}^{+\infty} E_n\right) \cup \left(A \cap \left(\bigcup_{n=1}^{+\infty} E_n\right)^c\right)$ . Avendo un'unione finita di eventi disgiunti, vale  $\mathbf{P}(A) = \mathbf{P}\left(A \cap \bigcup_{n=1}^{+\infty} E_n\right) + \mathbf{P}\left(A \cap \left(\bigcup_{n=1}^{+\infty} E_n\right)^c\right)$ . Poiché  $\mathbf{P}\left(\bigcup_{n=1}^{+\infty} E_n\right) = 1$ ,

$\mathbf{P}\left(A \cap \left(\bigcup_{n=1}^{+\infty} E_n\right)^c\right) \leq \mathbf{P}\left(\left(\bigcup_{n=1}^{+\infty} E_n\right)^c\right) = 0$ , quindi  $\mathbf{P}\left(A \cap \left(\bigcup_{n=1}^{+\infty} E_n\right)^c\right) = 0$ . Inoltre  $\forall N \mathbf{P}\left(A \cap \bigcup_{n=1}^N E_n\right) = \mathbf{P}\left(\bigcup_{n=1}^N (A \cap E_n)\right) = \sum_{n=1}^N \mathbf{P}(A \cap E_n) - \sum_{1 \leq i < j \leq N} \mathbf{P}((A \cap E_i) \cap (A \cap E_j)) + \sum_{1 \leq i < j < k \leq N} \mathbf{P}((A \cap E_i) \cap (A \cap E_j) \cap (A \cap E_k)) - \dots + (-1)^{N-1} \mathbf{P}((A \cap E_1) \cap \dots \cap (A \cap E_N))$ , che è uguale a  $\sum_{n=1}^{+\infty} \mathbf{P}(A \cap E_n)$  perché  $\mathbf{P}(E_h \cap E_k) = 0 \forall h \neq k$ .  
Dunque  $\mathbf{P}(A) = \sum_{n=1}^{+\infty} \mathbf{P}(A \cap E_n)$ . ■

Si è visto che, dato  $E : \mathbf{P}(E) > 0$ ,  $\mathbf{P}(A|E) = \frac{\mathbf{P}(A \cap E)}{\mathbf{P}(E)}$ , fissato  $E$ , è una misura di probabilità su  $\mathcal{A}$  al variare di  $A \in \mathcal{A}$ . Quindi, se si conosce  $\mathbf{P}_E(A)$ , si ha che  $\mathbf{P}_E(A^c) = 1 - \mathbf{P}_E(A) = 1 - \mathbf{P}(A|E)$ . In generale invece  $\mathbf{P}(A|E)$ , fissato  $A$ , non è una misura di probabilità al variare di  $E$  (infatti in generale  $\mathbf{P}(A|E) \neq \mathbf{P}(E|A)$ ). Tuttavia esiste una relazione tra le due, espressa dal seguente teorema.

#### Teo 4.3 (Bayes)

Hp:  $(\Omega, \mathcal{A}, \mathbf{P})$  è uno spazio di probabilità,  $(E_n)_{n \geq 1}$  è una partizione discreta

di  $\Omega$  con  $E_n \in \mathcal{A} \forall n$  e  $\mathbf{P}(E_n) > 0 \forall n$ ,  $A \in \mathcal{A}$  è non trascurabile

$$\text{Ts: } \mathbf{P}(E_n|A) = \frac{\mathbf{P}(A|E_n) \mathbf{P}(E_n)}{\mathbf{P}(A)} = \frac{\mathbf{P}(A|E_n) \mathbf{P}(E_n)}{\sum_{n=1}^{+\infty} \mathbf{P}(A|E_n) \mathbf{P}(E_n)} \quad \forall h \geq 1$$

Il teorema evidenzia che non c'è una relazione lineare tra  $\mathbf{P}(E_n|A)$  e  $\mathbf{P}(A|E_n) \mathbf{P}(E_n)$

**Dim** Per definizione  $\mathbf{P}(E_n|A) = \frac{\mathbf{P}(E_n \cap A)}{\mathbf{P}(A)}$ , ben definita perché  $\mathbf{P}(A) > 0$ . Essendo anche  $\mathbf{P}(E_n) > 0 \forall h$ , è ben definita  $\mathbf{P}(A|E_n) = \frac{\mathbf{P}(E_n \cap A)}{\mathbf{P}(E_n)}$ : sostituendo nell'uguaglianza precedente si ha  $\mathbf{P}(E_h|A) = \frac{\mathbf{P}(A|E_n) \mathbf{P}(E_n)}{\mathbf{P}(A)}$ . La seconda uguaglianza della tesi vale perché  $\mathbf{P}(A) = \sum_{n=1}^{+\infty} \mathbf{P}(A|E_n) \mathbf{P}(E_n)$  per la formula delle probabilità totali. ■

1. Nell'esempio precedente,  $\mathbf{P}(N_1|N_2) = \frac{\mathbf{P}(N_2|N_1) \mathbf{P}(N_1)}{\mathbf{P}(N_2)} = \frac{n-1}{n+b-1} \frac{n}{n+b} \frac{n+b}{n} = \frac{n-1}{n+b-1}$ , che coincide con  $\mathbf{P}(N_2|N_1)$ .

#### Teo 4.4 (formula di moltiplicazione)

Hp:  $(\Omega, \mathcal{A}, \mathbf{P})$  è uno spazio di probabilità,  $A_1, \dots, A_n \in \mathcal{A}$ ,  $\mathbf{P}(A_1 \cap \dots \cap A_{n-1}) > 0$

Ts:  $\mathbf{P}(A_1 \cap \dots \cap A_n) = \mathbf{P}(A_1) \mathbf{P}(A_2|A_1) \mathbf{P}(A_3|A_2 \cap A_1) \dots \mathbf{P}(A_n|A_{n-1} \cap \dots \cap A_1)$

Poiché  $\mathbf{P}(A_1 \cap \dots \cap A_{n-1}) > 0$  e  $A_1 \supseteq A_1 \cap A_2 \supseteq \dots \supseteq A_1 \cap \dots \cap A_{n-1}$ , per  $\mathbf{P}3 \forall i = 1, \dots, n-1 \mathbf{P}(A_1 \cap \dots \cap A_i) \geq \mathbf{P}(A_1 \cap \dots \cap A_{i+1})$ , e questo implica  $\mathbf{P}(A_1 \cap \dots \cap A_i) \geq \dots \geq \mathbf{P}(A_1 \cap \dots \cap A_{n-1}) > 0 \forall i = 1, \dots, n-1$ , dunque le probabilità condizionate nella tesi sono tutte ben definite.

La formula è utile quando è facile calcolare le probabilità condizionate.

**Dim** Scrivo il lato destro usando la definizione di probabilità condizionata:  $\mathbf{P}(A_1) \mathbf{P}(A_2|A_1) \mathbf{P}(A_3|A_2 \cap A_1) \dots \mathbf{P}(A_n|A_{n-1} \cap \dots \cap A_1) = \mathbf{P}(A_1) \frac{\mathbf{P}(A_2 \cap A_1)}{\mathbf{P}(A_1)} \frac{\mathbf{P}(A_3 \cap A_2 \cap A_1)}{\mathbf{P}(A_2 \cap A_1)} \dots \frac{\mathbf{P}(A_n \cap A_{n-1} \cap \dots \cap A_1)}{\mathbf{P}(A_{n-1} \cap \dots \cap A_1)} = \mathbf{P}(A_n \cap A_{n-1} \cap \dots \cap A_1)$ . La dimostrazione si potrebbe fare anche per induzione. ■

## 4 Indipendenza stocastica

1. Nell'ultimo esempio visto, con estrazione senza reimmissione,  $\mathbf{P}(N_2|N_1) = \frac{n-1}{n+b-1} \neq \mathbf{P}(N_2) = \frac{n}{n+b}$ . Se invece si estrae con reimmissione, si sta effettuando un esperimento diverso, descritto da un'altra terna di Kolmogorov, con misura di probabilità  $\tilde{\mathbf{P}} : \tilde{\mathbf{P}}(N_2|N_1) = \frac{n}{n+b} = \tilde{\mathbf{P}}(N_2)$ : in assenza di reimmissione  $\mathbf{P}(N_2|N_1) \neq \mathbf{P}(N_2)$ , cioè la probabilità è influenzata dall'essersi verificato  $N_1$ , mentre ciò non accade con reimmissione, perché  $\tilde{\mathbf{P}}(N_2|N_1) = \frac{\tilde{\mathbf{P}}(N_2 \cap N_1)}{\tilde{\mathbf{P}}(N_1)} = \tilde{\mathbf{P}}(N_2)$ . Questa intuizione modellistica è formalizzata dalla definizione di indipendenza.

**Def 4.5** Dato  $(\Omega, \mathcal{A}, \mathbf{P})$  spazio di probabilità,  $A, E \in \mathcal{A}$  si dicono indipendenti, e si scrive  $A \perp E$ , se  $\mathbf{P}(A \cap E) = \mathbf{P}(A) \mathbf{P}(E)$ .

La situazione precedentemente descritta,  $\frac{\tilde{\mathbf{P}}(N_2 \cap N_1)}{\tilde{\mathbf{P}}(N_1)} = \tilde{\mathbf{P}}(N_2)$ , permette quindi di affermare che con reimmissione  $N_1$  e  $N_2$  sono eventi indipendenti, senza reimmissione non lo sono. Si noti che, per astrarre la definizione di eventi indipendenti, nell'ultima uguaglianza si sono moltiplicati entrambi i lati per  $\tilde{\mathbf{P}}(N_1)$ , per rimuovere la necessità di  $\tilde{\mathbf{P}}(N_1) > 0$ . La relazione di indipendenza tra eventi gode della proprietà di simmetria, ma non della riflessività (in generale  $\mathbf{P}(A) \neq \mathbf{P}^2(A)$ , quindi non è una relazione di equivalenza).

Se inoltre  $\mathbf{P}(A), \mathbf{P}(E) > 0$ ,  $A \perp E \iff \mathbf{P}(A|E) = \mathbf{P}(A) \iff \mathbf{P}(E|A) = \mathbf{P}(E)$ , come si ricava dalla definizione, dividendo per  $\mathbf{P}(E), \mathbf{P}(A)$  rispettivamente. In questa uguaglianza sta il significato modellistico dell'indipendenza tra eventi: il verificarsi di uno non influenza la probabilità del verificarsi dell'altro.

La definizione di indipendenza aiuta inoltre a costruire i modelli probabilistici.

1. Lancio un dado rosso e un dado bianco. Chiamo  $R_k$  l'evento "esce la faccia  $k$  sul dado rosso", analogamente  $B_k$ ;  $\Omega = \{(h, k) : h, k \in \{1, \dots, 6\}\}$ ;  $(R_k)_{k \in \{1, \dots, 6\}}, (B_k)_{k \in \{1, \dots, 6\}}$  sono due partizioni di  $\Omega$  (gli  $A_n$  e  $B_n$  del teorema).  $\mathcal{A} = \sigma(B_k \cap R_h, h, k \in \{1, \dots, 6\})$ ; l'evento  $B_k \cap R_h$  è identificato dalla coppia  $(k, h)$ . Per il teorema 4.1, basta imporre  $p_k = \frac{1}{6} \forall k = 1, \dots, 6$  e  $q_{h|n} = \frac{1}{6} \forall h, \forall k = 1, \dots, 6$ :  $p_k$  rappresenta  $\mathbf{P}(R_k)$ ,  $q_{h|k}$  è effettivamente  $\mathbf{P}(B_k|R_k) = \mathbf{P}(B_k)$ , perché i risultati sui due dadi sono indipendenti. Allora  $\mathbf{P}(R_h \cap B_k) = \mathbf{P}(R_h) \mathbf{P}(B_k) = \frac{1}{6} \frac{1}{6} \forall h, k = 1, \dots, 6$ , cioè  $\mathbf{P}((h, k)) = \frac{1}{36} \forall h, k = 1, \dots, 6$ : l'indipendenza dei due dadi induce il modello uniforme nello spazio misurabile  $(\Omega, \mathcal{A})$ .



**Prop (indipendenza tra eventi)**

(i) Hp:  $(\Omega, \mathcal{A}, \mathbf{P})$  è uno spazio di probabilità,  $A \in \mathcal{A}$ ,  $\mathbf{P}(A) = 0$  o  $\mathbf{P}(A) = 1$

Ts:  $\forall E \in \mathcal{A}, A \perp E$

(ii) Hp:  $(\Omega, \mathcal{A}, \mathbf{P})$  è uno spazio di probabilità,  $A, B \in \mathcal{A}$ ,  $A \subseteq B$ ,  $\mathbf{P}(A) > 0$ ,  $\mathbf{P}(B) < 1$

Ts:  $A \not\perp B$

(iii) Hp:  $(\Omega, \mathcal{A}, \mathbf{P})$  è uno spazio di probabilità,  $A, B \in \mathcal{A}$ ,  $A \cap B = \emptyset$ ,  $\mathbf{P}(A), \mathbf{P}(B) > 0$

Ts:  $A \not\perp B$

(iv) Hp:  $(\Omega, \mathcal{A}, \mathbf{P})$  è uno spazio di probabilità,  $A, E \in \mathcal{A}$

Ts:  $A \perp E \iff A \perp E^c \iff A^c \perp E \iff A^c \perp E^c$

(i) significa che un evento certo o un evento impossibile è indipendente da qualsiasi evento. (ii) intuitivamente è dovuta al fatto che  $A$  implica  $B$ , quindi sapere che  $A$  si è verificato modifica il grado di fiducia nel verificarsi di  $B$ . (iii) intuitivamente è dovuta al fatto che  $A$  e  $B$  sono mutuamente esclusivi, quindi dall'informazione che uno dei due si è verificato si ricava con certezza che l'altro non si è verificato, quindi non possono essere indipendenti.

Si noti che le ipotesi di (i) e (iii) sono incompatibili: infatti, se  $\mathbf{P}(A) = 1$  e  $A \cap B = \emptyset$ , necessariamente  $\mathbf{P}(B) = 0$  per  $\sigma$ -additività.

**Dim** (i)  $A \perp E \iff \mathbf{P}(A \cap E) = \mathbf{P}(A)\mathbf{P}(E)$ : se  $\mathbf{P}(A) = 0$ ,  $\mathbf{P}(A \cap E) \leq \mathbf{P}(A) = 0$ , quindi effettivamente  $\mathbf{P}(A \cap E) = 0 = \mathbf{P}(A)\mathbf{P}(E)$ . Se  $\mathbf{P}(A) = 1$ ,  $\mathbf{P}(E) = \mathbf{P}((A \cap E) \cup (A^c \cap E)) = \mathbf{P}(A \cap E) + \mathbf{P}(A^c \cap E)$ . Ma  $\mathbf{P}(A^c \cap E) = 0$  perché  $\mathbf{P}(A^c) = 0$ , quindi effettivamente  $\mathbf{P}(A \cap E) = \mathbf{P}(A)\mathbf{P}(E) = \mathbf{P}(E)$ .

(ii)  $\mathbf{P}(A \cap B) = \mathbf{P}(A)$ , quindi non è possibile che  $\mathbf{P}(A \cap B) = \mathbf{P}(A)\mathbf{P}(B)$  a meno che  $\mathbf{P}(A) = 0$  o  $\mathbf{P}(B) = 1$ , casi esclusi dalle ipotesi.

(iii)  $\mathbf{P}(A \cap B) = \mathbf{P}(\emptyset) = 0$ , quindi, essendo  $\mathbf{P}(A) \neq 0$  e  $\mathbf{P}(B) \neq 0$ , non è possibile  $\mathbf{P}(A \cap B) = \mathbf{P}(A)\mathbf{P}(B)$ .

(iv) Se  $A \perp E$ ,  $\mathbf{P}(A \cap E) = \mathbf{P}(A)\mathbf{P}(E)$ . Inoltre, essendo  $A = (A \cap E) \cup (A \cap E^c)$ , per A2  $\mathbf{P}(A) = \mathbf{P}(A \cap E) + \mathbf{P}(A \cap E^c) = \mathbf{P}(A)\mathbf{P}(E) + \mathbf{P}(A \cap E^c)$ , dunque  $\mathbf{P}(A \cap E^c) = \mathbf{P}(A)(1 - \mathbf{P}(E)) = \mathbf{P}(A)\mathbf{P}(E^c)$ , cioè  $A, E^c$  sono indipendenti. Se si ripete la stessa dimostrazione a parti invertite per  $E, A$  si ottiene l'indipendenza di  $A^c, E$ , se si ripete di nuovo su  $A^c, E$  si ottiene l'indipendenza di  $A^c, E^c$ . ■

**Def 5.1** Dato  $(\Omega, \mathcal{A}, \mathbf{P})$  spazio di probabilità e  $I$  insieme qualsiasi, una famiglia di insiemi  $(A_i)_{i \in I} \subseteq \mathcal{A}$  si dice famiglia di eventi indipendenti se  $\forall J \subseteq I : 2 \leq |J| < +\infty$  vale  $\mathbf{P}\left(\bigcap_{j \in J} A_j\right) = \prod_{j \in J} \mathbf{P}(A_j)$ .

$J$  indicizza una sottofamiglia di  $(A_i)_{i \in I}$ ;  $\prod_{j \in J} \mathbf{P}(A_j)$  è una produttoria finita. Nella definizione non si richiede che  $I$  sia al più numerabile. Noto che se  $I = \{1, 2\}$ , si ritrova la definizione già vista di due eventi indipendenti:

infatti l'unico  $J$  che soddisfa le richieste è  $J = I$  e si ritrova  $\mathbf{P}(A_1 \cap A_2) = \mathbf{P}(A_1) \mathbf{P}(A_2)$ .

Se invece  $I = \{1, 2, 3\}$ , i  $J$  possibili sono  $\{1, 2\}, \{1, 3\}, \{2, 3\}, \{1, 2, 3\}$ : la condizione  $\mathbf{P}\left(\bigcap_{j \in J} A_j\right) = \prod_{j \in J} \mathbf{P}(A_j)$  va verificata per ogni  $J$ , cioè occorre controllare le quattro uguaglianze  $\mathbf{P}(A_1 \cap A_2) = \mathbf{P}(A_1) \mathbf{P}(A_2)$ ,  $\mathbf{P}(A_1 \cap A_3) = \mathbf{P}(A_1) \mathbf{P}(A_3)$ ,  $\mathbf{P}(A_2 \cap A_3) = \mathbf{P}(A_2) \mathbf{P}(A_3)$ ,  $\mathbf{P}(A_1 \cap A_2 \cap A_3) = \mathbf{P}(A_1) \mathbf{P}(A_2) \mathbf{P}(A_3)$ . Questa è quindi una definizione laboriosa, che difficilmente potrà essere usata in pratica.

Se una famiglia di eventi è di eventi indipendenti, allora gli eventi sono a due a due, a tre a tre, ... indipendenti. Se  $(A_i)_{i \in I} \subseteq \mathcal{A}$  è una famiglia di eventi indipendenti, ogni sottinsieme della famiglia è a sua volta una famiglia di eventi indipendenti.

1. Se  $|I| = n \in \mathbb{N}$ , il numero di  $J \subseteq I : |J| < +\infty$  e  $2 \leq |J| < +\infty$  (e quindi il numero di condizioni da verificare) è  $2^n - n - 1$ : coincide infatti con l'insieme delle parti, meno gli insiemi con un solo elemento (che sono  $n$ ), meno l'insieme vuoto.

Avrei potuto pensare anche a una definizione più semplice, ad esempio che  $(A_i)_{i \in I} \subseteq \mathcal{A}$  si dice famiglia di eventi indipendenti se vale  $\mathbf{P}\left(\bigcap_{i \in I} A_i\right) = \prod_{i \in I} \mathbf{P}(A_i)$ . Si dimostra però che la 5.1 è l'unica definizione che permette di dimostrare proprietà che sono ragionevoli da un punto di vista modellistico, ad esempio che se  $A_1, A_2, A_3$  sono indipendenti, allora  $A_1$  e  $A_2 \cup A_3$  sono indipendenti.

### Prop

Hp:  $(\Omega, \mathcal{A}, \mathbf{P})$  è uno spazio di probabilità,  $(A_i)_{i \in I} \subseteq \mathcal{A}$

è una famiglia di eventi indipendenti

Ts:  $\forall \bar{i} \in I, \forall B \in \sigma\left((A_i)_{i \in I, i \neq \bar{i}}\right), A_{\bar{i}} \perp B$

La tesi significa che ogni evento è indipendente da qualsiasi elemento della  $\sigma$ -algebra generata da tutti gli altri eventi.

**Dim** (i) Dimostro che  $A_1 \perp (A_2 \cup A_3)$ . Per definizione,  $\forall J \subseteq \{1, 2, 3\} : |J| < +\infty$  e  $2 \leq |J| < +\infty$  vale  $\mathbf{P}\left(\bigcap_{j \in J} A_j\right) = \prod_{j \in J} \mathbf{P}(A_j)$ , cioè  $\mathbf{P}(A_1 \cap A_2) = \mathbf{P}(A_1) \mathbf{P}(A_2)$ ,  $\mathbf{P}(A_1 \cap A_3) = \mathbf{P}(A_1) \mathbf{P}(A_3)$ ,  $\mathbf{P}(A_2 \cap A_3) = \mathbf{P}(A_2) \mathbf{P}(A_3)$ ,  $\mathbf{P}(A_1 \cap A_2 \cap A_3) = \mathbf{P}(A_1) \mathbf{P}(A_2) \mathbf{P}(A_3)$ . Devo mostrare che  $\mathbf{P}(A_1 \cap (A_2 \cup A_3)) = \mathbf{P}(A_1) \mathbf{P}(A_2 \cup A_3)$ . Per la proprietà distributiva dell'intersezione rispetto all'unione e per P6  $\mathbf{P}(A_1 \cap (A_2 \cup A_3)) = \mathbf{P}((A_1 \cap A_2) \cup (A_1 \cap A_3)) = \mathbf{P}(A_1 \cap A_2) + \mathbf{P}(A_1 \cap A_3) - \mathbf{P}(A_1 \cap A_2 \cap A_3)$ : usando l'ipotesi e di nuovo P6 si ottiene  $\mathbf{P}(A_1) \mathbf{P}(A_2) + \mathbf{P}(A_1) \mathbf{P}(A_3) - \mathbf{P}(A_1) \mathbf{P}(A_2) \mathbf{P}(A_3) = \mathbf{P}(A_1) (\mathbf{P}(A_2) + \mathbf{P}(A_3) - \mathbf{P}(A_2) \mathbf{P}(A_3)) = \mathbf{P}(A_1) \mathbf{P}(A_2 \cup A_3)$  perché  $\mathbf{P}(A_2 \cap A_3) = \mathbf{P}(A_2) \mathbf{P}(A_3)$ . ■

### Prop

Hp:  $(\Omega, \mathcal{A}, \mathbf{P})$  è uno spazio di probabilità,  $(A_i)_{i \in I} \subseteq \mathcal{A}$  è una famiglia di eventi

indipendenti,  $(B_i)_{i \in I} \subseteq \mathcal{A}$  è tale che  $B_i = A_i$  o  $B_i = A_i^c \forall i \in I$

Ts:  $(B_i)_{i \in I} \subseteq \mathcal{A}$  è una famiglia di eventi indipendenti

**Dim** Per ipotesi  $\forall J \subseteq I : 2 \leq |J| < +\infty$  vale  $\mathbf{P}\left(\bigcap_{j \in J} A_j\right) = \prod_{j \in J} \mathbf{P}(A_j)$ . Voglio mostrare che  $\forall J \subseteq I : 2 \leq |J| < +\infty$  vale  $\mathbf{P}\left(\bigcap_{j \in J} B_j\right) = \prod_{j \in J} \mathbf{P}(B_j)$ . Procedo per induzione sulla cardinalità  $|J| = n$ . Se  $n = 2$  vale che  $\mathbf{P}\left(\bigcap_{j \in J} B_j\right) = \prod_{j \in J} \mathbf{P}(B_j)$  perché  $B_{j_1}, B_{j_2}$  sono o  $A_{j_1}, A_{j_2}$ , che sono indipendenti perché l'indipendenza sulla famiglia implica l'indipendenza a coppie, o  $A_{j_1}, A_{j_2}^c$ , o  $A_{j_1}^c, A_{j_2}^c$  che sono due coppie indipendenti perché  $A_{j_1} \perp A_{j_2} \iff A_{j_1} \perp A_{j_1}^c \iff A_{j_1}^c \perp A_{j_1}^c$ . Suppongo vera la tesi per  $n > 2$  e mostro che vale per  $n + 1$ : suppongo quindi che, dato  $J : |J| = n$ ,  $\mathbf{P}\left(\bigcap_{j \in J} B_j\right) = \prod_{j \in J} \mathbf{P}(B_j)$ . Allora, dato  $J' : |J'| = n + 1$ ,  $\bigcap_{j \in J'} B_j = \bigcap_{j \in J} B_j \cap B_{n+1}$ . Ma essendo gli  $(A_j)_{j \in J}$  e  $A_{n+1}$  una famiglia di eventi indipendenti, vale in particolare che l'evento  $\bigcap_{j \in J} A_j$  (elemento della  $\sigma$ -algebra generata da "tutti gli altri eventi") è indipendente da  $A_{n+1}$ , e dunque anche da  $A_{n+1}^c$ , a seconda della natura di  $B_{n+1}$ . ■

## 5 Assegnazione della probabilità su $\mathcal{B}(\mathbb{R})$

$\mathcal{B}(\mathbb{R})$  è un insieme "molto grande": non c'è speranza di caratterizzare i suoi elementi in modo semplice (mentre è facile farlo e. g. con la  $\sigma$ -algebra generata da una partizione discreta di  $\Omega$ ). E' noto che  $\mathcal{B}(\mathbb{R}) = \sigma(\mathcal{C})$ ; poiché nel caso di  $\Omega$  si riesce a caratterizzare la probabilità definendola sugli elementi di  $\Omega$ , avendo come  $\sigma$ -algebra  $2^\Omega$  (che è la  $\sigma$  algebra generata dagli atomi), per analogia si potrebbe cercare di vedere se è possibile caratterizzare la probabilità definendola sugli elementi di  $\mathcal{C}$ , per poi estenderla su tutto  $\mathcal{B}(\mathbb{R})$  (ma l'estensione dev'essere unica). Questo è confermato dal seguente risultato.

**Def** Dato  $\Omega$  insieme qualsiasi, una classe di sottinsiemi di  $\Omega$   $\mathcal{C}$  si dice  $\pi$ -sistema su  $\Omega$  se  $C_1, \dots, C_n \in \mathcal{C} \implies C_1 \cap \dots \cap C_n \in \mathcal{C}$ .

Un  $\pi$ -sistema è quindi una classe di sottinsiemi di  $\Omega$  chiusa per intersezione finita. Ogni collezione di sottinsiemi disgiunti di  $\Omega$  che contenga anche l'insieme vuoto è un  $\pi$ -sistema su  $\Omega$ . Ogni  $\sigma$ -algebra su  $\Omega$  è un  $\pi$ -sistema su  $\Omega$ .

[**Def** Dato  $\Omega$ , qualsiasi, si dice classe monotona (detta anche  $\lambda$ -sistema) un sottinsieme di  $2^\Omega$   $\mathcal{M}$  tale che

**M1**  $\Omega \in \mathcal{M}$

**M2**  $A, B \in \mathcal{M}, A \subset B \implies B \setminus A \in \mathcal{M}$

**M3**  $(A_n) \in \mathcal{M}, A_n \subset A_{n+1} \implies \bigcup_{n=1}^{+\infty} A_n \in \mathcal{M}$

Si dice classe monotona generata da un sottinsieme  $\mathcal{F}$  di  $2^\Omega$ , e si indica con  $\lambda(\mathcal{F})$ , la più piccola classe monotona su  $\Omega$  che contenga  $\mathcal{C}$ .

Una classe monotona è quindi chiusa per differenze e per limiti crescenti.

Ogni  $\sigma$ -algebra è una classe monotona  $(B \setminus A = B \cap A^c)$ .

**Lemma di Dynkin (teorema delle classi monotone)**

Hp :  $\mathcal{C}$  è un  $\pi$ -sistema su  $\Omega$

Ts :  $\sigma(\mathcal{C}) = \lambda(\mathcal{C})$

]

**Teo 5.2 (criterio di Caratheodory)**

Hp:  $(\Omega, \mathcal{A})$  è uno spazio probabilizzabile,  $\mathbf{P}, \mathbf{Q}$  sono due misure di probabilità su

$(\Omega, \mathcal{A}), \mathcal{C} \subseteq \mathcal{A}$  è un  $\pi$ -sistema su  $\Omega$  tale che  $\sigma(\mathcal{C}) = \mathcal{A}, \mathbf{P}(C) = \mathbf{Q}(C) \forall C \in \mathcal{C}$

Ts:  $\mathbf{P}(A) = \mathbf{Q}(A) \forall A \in \mathcal{A}$

Il teorema afferma l'unicità dell'estensione di una probabilità da un  $\pi$ -sistema che generi  $\mathcal{A}$  ad  $\mathcal{A}$ ; in altre parole, due probabilità che coincidono in corrispondenza degli elementi di  $\mathcal{C}$  sono la stessa probabilità, vale a dire che il valore di una probabilità in corrispondenza degli elementi di  $\mathcal{C}$  la identifica univocamente su tutta  $\mathcal{A}$ . Per assegnare una probabilità a  $(\Omega, \mathcal{A})$  è dunque sufficiente assegnare la probabilità sugli elementi di un  $\pi$ -sistema.

Se  $\Omega$  è discreto,  $\mathcal{A} = 2^\Omega$  e  $\mathcal{C} = \{\emptyset, \{\omega_1\}, \{\omega_n\} \dots\}$  è l'insieme dei singoletti più l'insieme vuoto, si riottiene il fatto che stabilire una probabilità sugli atomi determina univocamente la probabilità.

1. Dato  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ ,  $\mathcal{C} = \{(a, b) : -\infty \leq a < b \leq +\infty\}$  è un  $\pi$ -sistema. Ricordiamo che  $\sigma(\mathcal{C}) = \mathcal{B}(\mathbb{R})$ .

$\tilde{\mathcal{C}} = \{(-\infty, q] : q \in \mathbb{Q}\}$  è un  $\pi$ -sistema. Ricordiamo che  $\sigma(\tilde{\mathcal{C}}) = \mathcal{B}(\mathbb{R})$ .

$\mathring{\mathcal{C}} = \{(-\infty, x] : x \in \mathbb{R}\}$  è un  $\pi$ -sistema. Vale inoltre  $\sigma(\mathring{\mathcal{C}}) = \mathcal{B}(\mathbb{R})$ . Dimostro che  $\sigma(\mathring{\mathcal{C}}) \subseteq \sigma(\tilde{\mathcal{C}})$ . Infatti, dato  $x \in \mathbb{R}$ ,  $\exists (q_n)_{n \geq 1} : q_n \in \mathbb{Q} \forall n, q_n \leq x \forall n$  e  $\lim_{n \rightarrow +\infty} q_n = x$ , perché  $\mathbb{Q}$  è denso in  $\mathbb{R}$ . Allora posso scrivere  $(-\infty, x] = \bigcup_{n=1}^{+\infty} (-\infty, q_n]$ : il lato destro appartiene a  $\sigma(\tilde{\mathcal{C}})$ , essendo ogni  $\sigma$ -algebra chiusa per unione numerabile, dunque  $(-\infty, x] \in \sigma(\tilde{\mathcal{C}})$  e  $\mathring{\mathcal{C}} \subseteq \sigma(\tilde{\mathcal{C}})$ , per cui  $\sigma(\mathring{\mathcal{C}}) \subseteq \sigma(\tilde{\mathcal{C}})$ . L'inclusione  $\sigma(\tilde{\mathcal{C}}) \subseteq \sigma(\mathring{\mathcal{C}})$  è ovvia, perché ogni numero razionale è anche reale, quindi  $\sigma(\mathring{\mathcal{C}}) = \sigma(\tilde{\mathcal{C}}) = \mathcal{B}(\mathbb{R})$ .

**Def 5.3** Dato  $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathbf{P})$  spazio di probabilità, si dice funzione di ripartizione associata a  $\mathbf{P}$ , e si indica con  $F$ , la funzione  $F : \mathbb{R} \rightarrow [0, 1] : F(x) = \mathbf{P}((-\infty, x]) \forall x \in \mathbb{R}$ .

Data  $\mathbf{P}$ , la funzione  $F$  è unica, ed è l'analoga di  $p : \Omega \rightarrow [0, 1]$  per spazi al più numerabili: è infatti definita su  $\mathbb{R}$  e non su  $\mathcal{B}(\mathbb{R})$ , il che ne riduce la complessità, e permette di risalire a  $\mathbf{P}(A) \forall A$  intervallo (e anche per altro: si vede dopo). Gli intervalli del tipo  $(-\infty, x]$  sono i generatori di  $\mathcal{B}(\mathbb{R})$ , proprio come gli  $\{\omega\}$  sono i generatori di  $2^\Omega$ : in questo caso gli atomi non sono gli eventi elementari, perché non sto generando dai singoletti, ma da  $\mathring{\mathcal{C}}$ . Per il criterio di Caratheodory è quindi ragionevole aspettarsi che  $F$  identifichi univocamente  $\mathbf{P}$ : assegnare  $F$  significa definire  $\mathbf{P}$  su un  $\pi$ -sistema che genera  $\mathcal{B}(\mathbb{R})$ .

**Teo 5.4 (proprietà della funzione di ripartizione)**

(1) (caratterizzazione della funzione di ripartizione)

Hp:  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$  spazio probabilizzabile

Ts:  $F : \mathbb{R} \rightarrow [0, 1]$  è funzione di ripartizione di una  $\mathbf{P} : \mathcal{B}(\mathbb{R}) \rightarrow [0, 1] \iff$  (A)  $F$  è monotona

non decrescente, (B)  $F$  è continua da destra, (C)  $\lim_{x \rightarrow -\infty} F(x) = 0$  e  $\lim_{x \rightarrow +\infty} F(x) = 1$

(2) (misura di intervalli con la funzione di ripartizione)

Hp:  $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathbf{P})$  spazio di probabilità,  $F$  è la funzione di ripartizione associata a  $\mathbf{P}$ ,  $x < y$

Ts: (i)  $\mathbf{P}((x, y]) = F(y) - F(x)$ , (ii)  $\mathbf{P}([x, y]) = F(y) - F(x^-)$ , (iii)  $\mathbf{P}([x, y)) = F(y^-) - F(x^-)$ ,

(iv)  $\mathbf{P}((x, y)) = F(y^-) - F(x)$ , (v)  $\mathbf{P}(\{x\}) = F(x) - F(x^-)$

(3)

Hp:  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$  spazio probabilizzabile,  $F : \mathbb{R} \rightarrow [0, 1]$  è una funzione di ripartizione di una  $\mathbf{P}$

Ts:  $F$  caratterizza  $\mathbf{P}$ , i. e.  $F_1 = F_2 \iff \mathbf{P}_1 = \mathbf{P}_2$

$F(x^-) = \lim_{y \rightarrow x^-} F(y)$ . In (1) A, B, C sono proprietà caratterizzanti di una funzione di ripartizione: è l'analogo continuo delle proprietà caratterizzanti viste per  $p : \Omega \rightarrow \mathbb{R}$  nel caso discreto. La situazione è tuttavia più complessa, nel senso che avendo  $F$  non si riesce a calcolare  $\mathbf{P}(A) \forall A \in \mathcal{A}$ ; perlomeno però si riesce a calcolare la probabilità di un qualsiasi intervallo. (2) evidenzia che in generale la probabilità di un singoletto non è nulla. (3) significa che conoscendo  $F$  il valore di  $\mathbf{P} \forall A \in \mathcal{A}$  è univocamente determinato.

**Dim** (1) Mostro l'implicazione da sinistra a destra.

(A) Voglio mostrare che se  $x < y$  allora  $F(x) \leq F(y)$ . Per definizione di funzione di ripartizione  $F(x) = \mathbf{P}((-\infty, x])$ ,  $F(y) = \mathbf{P}((-\infty, y])$ : ma se  $x < y$   $(-\infty, x] \subseteq (-\infty, y]$ , quindi per P3 (monotonia)  $\mathbf{P}((-\infty, x]) \leq \mathbf{P}((-\infty, y])$ , cioè  $F(x) \leq F(y)$ .

(B) Considero una successione  $x_n$  che decresce a  $x$ , cioè  $x_n : x_n \geq x_{n+1} \forall n$  e  $\lim_{n \rightarrow +\infty} x_n = x$ : mostro che, se  $x_n \downarrow x$ ,  $\lim_{n \rightarrow +\infty} F(x_n) = F(x)$ .  $F(x_n) = \mathbf{P}((-\infty, x_n])$ ; posso porre  $A_n = (-\infty, x_n]$ , e se  $x_n \downarrow x$   $A_n \downarrow A$ , con  $A = \bigcap_{n=1}^{+\infty} A_n = (-\infty, x]$ . Quindi  $\lim_{n \rightarrow +\infty} F(x_n) = \lim_{n \rightarrow +\infty} \mathbf{P}(A_n) = \mathbf{P}(A) = \mathbf{P}((-\infty, x]) = F(x)$  per la continuità della probabilità (2.4).

(C) Considero una successione  $x_n$  che decresce a  $-\infty$ , cioè  $x_n : x_n \leq x_{n+1} \forall n$  e  $\lim_{n \rightarrow +\infty} x_n = -\infty$ : mostro che, se  $x_n \downarrow -\infty$ ,  $\lim_{n \rightarrow +\infty} F(x_n) = 0$ .  $F(x_n) = \mathbf{P}((-\infty, x_n])$ ; posso porre  $A_n = (-\infty, x_n]$ , e se  $x_n \downarrow -\infty$   $A_n \downarrow A$ , con  $A = \bigcap_{n=1}^{+\infty} A_n = \emptyset$ . Quindi  $\lim_{n \rightarrow +\infty} F(x_n) = \lim_{n \rightarrow +\infty} \mathbf{P}(A_n) = \mathbf{P}(A) = \mathbf{P}(\emptyset) = 0$  per il teorema 2.4.

Analogamente considero una successione  $x_n$  che cresce a  $+\infty$ , cioè  $x_n : x_n \leq x_{n+1} \forall n$  e  $\lim_{n \rightarrow +\infty} x_n = +\infty$ : mostro che, se  $x_n \uparrow +\infty$ ,  $\lim_{n \rightarrow +\infty} F(x_n) = 1$ .  $F(x_n) = \mathbf{P}((-\infty, x_n])$ ; posso porre  $A_n = (-\infty, x_n]$ , e se  $x_n \uparrow +\infty$   $A_n \uparrow A$ , con  $A = \bigcup_{n=1}^{+\infty} A_n = \mathbb{R}$ . Quindi  $\lim_{n \rightarrow +\infty} F(x_n) = \lim_{n \rightarrow +\infty} \mathbf{P}(A_n) = \mathbf{P}(A) = \mathbf{P}(\Omega) = 1$  per 2.4.

(2) Sia  $x < y$ . (i)  $(x, y] = (-\infty, y] \setminus (-\infty, x]$ , quindi per P5 vale  $\mathbf{P}((x, y]) = \mathbf{P}((-\infty, y] \setminus (-\infty, x]) = \mathbf{P}((-\infty, y]) - \mathbf{P}((-\infty, x]) = \mathbf{P}((-\infty, y]) - \mathbf{P}((-\infty, x]) = F(y) - F(x)$ . Oppure, usando più semplicemente la  $\sigma$ -additività:  $(-\infty, y] = (-\infty, x] \cup (x, y]$ , quindi  $F(y) = F(x) + \mathbf{P}((x, y])$  e  $\mathbf{P}((x, y]) = F(y) - F(x)$ .

(iv)  $(x, y) = \bigcup_{n=1}^{+\infty} (x, y - \frac{1}{n}]$ : l'uguaglianza vale anche con  $\bigcup_{n=1}^{+\infty} (x, y - \frac{1}{n})$ , ma non si potrebbe usare (i). Ponendo  $A_n = (x, y - \frac{1}{n}]$ , poiché  $A_n \uparrow A$ , per la continuità della probabilità al contrario si ha che  $\mathbf{P}(\bigcup_{n=1}^{+\infty} A_n) = \lim_{n \rightarrow +\infty} \mathbf{P}(A_n)$ , ed è noto che  $\mathbf{P}((x, y - \frac{1}{n}]) = F(y - \frac{1}{n}) - F(x)$ , per cui  $\lim_{n \rightarrow +\infty} \mathbf{P}(A_n) = F(y^-) - F(x)$ .

(ii)  $[x, y] = (-\infty, y] \setminus (-\infty, x)$ . Posso scrivere  $\mathbf{P}((-\infty, x)) = \lim_{n \rightarrow +\infty} \mathbf{P}((-\infty, x - \frac{1}{n})) = \lim_{n \rightarrow +\infty} (F(x - \frac{1}{n}) - F(-n)) = F(x^-)$ . Quindi  $\mathbf{P}([x, y]) = \mathbf{P}((-\infty, y]) - \mathbf{P}((-\infty, y] \cap (-\infty, x)) = F(y) - \mathbf{P}((-\infty, x)) = F(y) - F(x^-)$ .

(iii)  $[x, y) = (-\infty, y) \setminus (-\infty, x)$ , quindi  $\mathbf{P}((-\infty, y) \setminus (-\infty, x)) = \mathbf{P}((-\infty, y)) - \mathbf{P}((-\infty, x)) = F(y^-) - F(x^-)$ .

(v)  $\{x\} = (-\infty, x] \setminus (-\infty, x)$ :  $\mathbf{P}((-\infty, x]) - \mathbf{P}((-\infty, x)) = F(x) - F(x^-)$  per quanto visto sopra.

(3) L'implicazione da destra a sinistra è ovvia. Dimostro quella da sinistra a destra:  $F_1(x) = F_2(x) \forall x \in \mathbb{R}$ , quindi  $\mathbf{P}_1, \mathbf{P}_2$  coincidono sugli elementi di  $\mathring{\mathcal{C}}$ ; essendo  $\mathring{\mathcal{C}}$  un  $\pi$ -sistema, per il criterio di Caratheodory  $\mathbf{P}_1 = \mathbf{P}_2$  su  $\sigma(\mathring{\mathcal{C}}) = \mathcal{B}(\mathbb{R})$ . ■

1. Dato  $\alpha \in \mathbb{R}$ , in  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$  definisco la probabilità massa di Dirac in  $\alpha$  come  $\mathbf{P}(A) = \begin{cases} 1 & \text{se } \alpha \in A \\ 0 & \text{se } \alpha \notin A \end{cases} \quad \forall$

$A \in \mathcal{B}(\mathbb{R})$ . Si può quindi ricavare  $F(x) = \mathbf{P}((-\infty, x]) = \begin{cases} 1 & \text{se } \alpha \leq x \\ 0 & \text{se } \alpha > x \end{cases}$ , detta funzione gradino di Heaviside.

La massa di Dirac rappresenta la situazione in cui si hanno tutte le informazioni: si è certi che l'evento  $\{\alpha\}$  si verifica.

2. Se  $F$  è una funzione di ripartizione continua in  $\mathbb{R}$ , allora - per una versione del teorema dei valori intermedi che si applica perché  $\lim_{x \rightarrow +\infty} F(x) = 1, \lim_{x \rightarrow -\infty} F(x) = 0$  -  $\forall \lambda \in (0, 1) \exists x \in \mathbb{R} : F(x) = \lambda$ .

**Def** Dato lo spazio probabilizzabile  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$  e  $T \subseteq \mathbb{R}$  discreto, si dice che  $\mathbf{P}$  è una probabilità discreta su  $\mathcal{B}(\mathbb{R})$  se  $\exists p : T \rightarrow \mathbb{R}$ , detta funzione di densità discreta, tale che  $\begin{cases} p(x) \geq 0 \forall x \in T \\ \sum_{x \in T} p(x) = 1 \end{cases}$ , e se  $\mathbf{P}$  è definita come  $\mathbf{P}(A) = \sum_{x \in A \cap T} p(x) \forall A \in \mathcal{B}(\mathbb{R})$ .

Quindi  $F(x) = \mathbf{P}((-\infty, x]) = \sum_{y \in (-\infty, x] \cap T} p(y)$ . L'assegnazione di una probabilità su  $\mathbb{R}$  comprende quindi anche il caso numerabile.

Dato  $x \in T$ , è possibile che  $p(x) = 0$ .

**Def** Dato  $T \subseteq \mathbb{R}$  al più numerabile, dato lo spazio di probabilità  $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathbf{P})$  con funzione di densità discreta  $p : T \rightarrow \mathbb{R}$ , si dice supporto della densità o della probabilità, e si indica con  $\mathcal{S}, \{x \in T : p(x) > 0\}$ .

Allora  $\forall A \in \mathcal{B}(\mathbb{R}) \mathbf{P}(A) = \sum_{x \in A \cap T} p(x) = \sum_{x \in A \cap \mathcal{S}} p(x)$ , e in particolare  $\mathbf{P}((-\infty, x]) = \sum_{y \in (-\infty, x] \cap \mathcal{S}} p(y)$ .

1. Le distribuzioni di probabilità bernoulli, binomiale, poisson introdotte nel caso di  $\Omega$  discreti ( $\{0, 1\}, \{0, \dots, n\}, \mathbb{N}$ , rispettivamente), tutti immersi in  $\mathbb{R}$ , hanno delle  $p$  che sono funzioni di densità discrete.

Ci si potrebbe chiedere perché si prende  $T$  e non direttamente  $\mathcal{S}$ .  $Bi(n, p)$  al variare di  $n$  ha come supporto  $T = \mathbb{N}$ . Se invece fisso  $n$   $Bi(n, p)$  ha come supporto  $\mathcal{S} = \{0, \dots, n\} \subsetneq T$ .

**Def** Una funzione di ripartizione  $F$  associata a  $\mathbf{P}$  si dice di puro salto se, detto  $O = \{\text{punti di salto}\}$ ,  $\sum_{x \in O} (F(x) - F(x^-)) = 1$ .

$S$  coincide con l'insieme dei punti di discontinuità di  $F$ , perché una funzione monotona definita su  $\mathbb{R}$  può avere solo punti di discontinuità a salto. Inoltre  $S$  è discreto, perché l'insieme dei punti di discontinuità di una funzione monotona definita su  $\mathbb{R}$  è al più numerabile.

### Prop 5.5 (caratterizzazione delle probabilità discrete)

Hp:  $(\mathbb{R}, \mathcal{B}(\mathbb{R})), \mathbf{P} : \mathcal{B}(\mathbb{R}) \rightarrow [0, 1]$  con funzione di ripartizione  $F$

Ts:  $\mathbf{P}$  è una probabilità discreta con densità  $p$  e supporto  $\mathcal{S} \iff F$  è di puro salto e

$$\sum_{x \in \mathcal{S}} F(x) - F(x^-) = 1; \text{ in tal caso } \mathcal{S} = \{\text{punti di discontinuità di } F\} \text{ e } p(x) = F(x) - F(x^-)$$

Questo significa che posso caratterizzare  $p$  e ricavarne il valore a partire dalle proprietà analitiche di  $F$ .

1. Se  $F$  è di puro salto ed è discontinua in  $\{x_1, x_2, x_3\}$ , allora  $\mathcal{S} = \{x_1, x_2, x_3\}$  e  $p(x_k) = F(x_k) - F(x_k^-)$ .

Analogamente posso ricavare  $F$  da  $p$ , dato che  $\mathbf{P}((-\infty, x]) = \sum_{y \in (-\infty, x] \cap T} p(y) \forall x$ .

**Def** Data  $f : \mathbb{R} \rightarrow \mathbb{R}$ , se  $f$  è Riemann integrabile,  $f(x) \geq 0 \forall x \in \mathbb{R}$  e  $\int_{-\infty}^{+\infty} f(x) dx = 1$ ,  $f$  si dice funzione di densità continua.

E' allora ben definita la funzione di  $x \int_{-\infty}^x f(t) dt$ , e si può dimostrare che  $F(x) = \int_{-\infty}^x f(t) dt$  è una funzione di ripartizione, cioè soddisfa A (perché  $f$  è non negativa), B (perché una funzione integrale è continua per il teorema fondamentale del calcolo integrale), C e identifica dunque un'unica probabilità  $\mathbf{P}$  su  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ . Poiché in tal caso  $F$  è continua,  $\mathbf{P}(\{x\}) = 0 = F(x) - F(x^-) \forall x \in \mathbb{R}$ . Le richieste su  $f$  sono la versione continua delle richieste sulla densità discreta. Se quindi  $F$  non è continua, non ammette una densità, cioè  $\exists f : \mathbb{R} \rightarrow \mathbb{R} : F(x) = \int_{-\infty}^x f(t) dt$ .

Se inoltre  $f$  è continua a tratti,  $F$  è  $C^1$  a tratti per il teorema fondamentale e  $F'(x) = f(x)$ . Quindi, se mi viene assegnata  $F$   $C^1$  a tratti, allora so che  $F$  è "di questo tipo qua"?? perché la sua derivata "fa il gioco della densità".

Si può quindi costruire un modello di probabilità su  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$  a partire da una funzione di densità, perché essa definisce  $F(x) = \int_{-\infty}^x f(t) dt$  e quindi  $\mathbf{P}$ . In base al tipo di probabilità che voglio costruire sceglierò una funzione di ripartizione di un tipo o di un altro. Si noti che non è per ora spiegato come sono fatte le  $\mathbf{P}$  per cui esista una densità continua.

$$1. \text{ Sia } f(t) = \frac{1}{b-a} I_{(a,b]}(t): f \text{ soddisfa le richieste sopra. Allora } F(x) = \int_{-\infty}^x f(t) dt = \begin{cases} 0 & \text{se } x < a \\ \frac{1}{b-a}(x-a) & \text{se } x \in [a, b] \\ 1 & \text{se } x > b \end{cases}$$

è una funzione di ripartizione, che dota  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$  della cosiddetta probabilità uniforme su  $(a, b]$ ,  $\mathcal{U}((a, b])$ .

$$2. \text{ Sia } f(t) = \lambda e^{-\lambda t} I_{(0,+\infty)}(t): f \text{ soddisfa le richieste sopra. Allora } F(x) = \int_{-\infty}^x f(t) dt = \begin{cases} 0 & \text{se } x \leq 0 \\ 1 - e^{-\lambda x} & \text{se } x \in (0, +\infty) \end{cases}$$

è una funzione di ripartizione, che dota  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$  del modello esponenziale  $\varepsilon(\lambda)$ .

$$3. \text{ Sia } f(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2}: f \text{ soddisfa le richieste sopra. Allora } F(x) = \int_{-\infty}^x f(t) dt \text{ è una funzione di ripartizione, che dota } (\mathbb{R}, \mathcal{B}(\mathbb{R})) \text{ del modello normale } \mathcal{N}(0, 1).$$

4. Se  $F$  non è continua, non ammette una densità di probabilità continua; se non è di puro salto, non ammette una densità di probabilità discreta. Se quindi  $F$  è discontinua e non è di puro salto, non ammette densità di probabilità.



## 6 Variabili aleatorie

Dato  $(\Omega, \mathcal{A}, \mathbf{P})$  modello di un esperimento aleatorio, è utile introdurre una funzione  $X : \Omega \rightarrow \mathbb{R}$ , che restituisca un numero che riassume ciò che mi interessa dell'esperimento.

1. Se lancio due dadi distinti,  $\Omega = \{\omega = (\omega_1, \omega_2) : \omega_i \in \{1, \dots, 6\} \ \forall i = 1, 2\}$ . Se sono interessata alla somma dei risultati, posso introdurre  $X : \Omega \rightarrow \mathbb{R}$ ,  $X(\omega) = \omega_1 + \omega_2$ .

E' normale allora che per ogni  $B \in \mathcal{B}(\mathbb{R})$  mi interessi la probabilità che la funzione che descrive il risultato di mio interesse assuma valori in  $B$ , dunque la probabilità che  $X$  cada in  $B$ , cioè  $\mathbf{P}((X \in B)) \ \forall B \in \mathcal{B}(\mathbb{R})$ . Affinché possa calcolare tale probabilità è necessario che  $(X \in B) \in \mathcal{A} \ \forall B \in \mathcal{B}(\mathbb{R})$ : questa condizione prende il nome di misurabilità.

**Def 6.2** Dato  $E$  insieme qualsiasi ed  $\mathcal{E}$   $\sigma$ -algebra su  $E$ , si dice spazio misurabile la coppia  $(E, \mathcal{E})$ .

$\mathcal{E}$  è ovviamente una famiglia di sottinsiemi di  $E$ . Se in particolare  $E = \Omega$ , che ha il significato modellistico di spazio campionario,  $(E, \mathcal{E})$  si dice spazio probabilizzabile.

**Def 6.3** Dati gli spazi misurabili  $(E, \mathcal{E})$ ,  $(F, \mathcal{F})$ , una funzione  $f : E \rightarrow F$  si dice funzione misurabile se  $\forall B \in \mathcal{F} \ f^{-1}(B) \in \mathcal{E}$ .

$f^{-1}(B)$  è ben definita perché il generico elemento di  $\mathcal{F}$  è un sottinsieme di  $F$ . La definizione formalizza il ragionamento fatto prima nel caso di  $(E, \mathcal{E}) = (\Omega, \mathcal{A})$ ,  $(F, \mathcal{F}) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ . In generale una funzione misurabile si indica con  $f : (E, \mathcal{E}) \rightarrow (F, \mathcal{F})$ .

Si vedrà in seguito che la misurabilità è una richiesta di regolarità di  $f$  molto debole.

**Def 6.4** Dato  $(\Omega, \mathcal{A})$  spazio probabilizzabile e  $(E, \mathcal{E})$  spazio misurabile,  $X : \Omega \rightarrow E$  si dice variabile aleatoria se  $X$  è una funzione misurabile.

A livello strettamente matematico non c'è quindi differenza tra una v.a. e una funzione misurabile. La definizione di v.a. significa che  $\forall B \in \mathcal{E} \ (X \in B) = X^{-1}(B) \in \mathcal{A}$ , cioè  $X^{-1}(B)$  è un evento e dunque - se ho fissato anche  $\mathbf{P}$  su  $(\Omega, \mathcal{A})$  - ne posso calcolare la probabilità: cioè, dato  $(\Omega, \mathcal{A}, \mathbf{P})$  spazio di probabilità, se  $X : \Omega \rightarrow E$  è una variabile aleatoria allora  $\mathbf{P}(X \in B)$  è ben definita  $\forall B \in \mathcal{E}$   $\sigma$ -algebra su  $E$ . Questo è fondamentale per poter calcolare la probabilità di eventi rappresentabili attraverso i valori che assume  $X$  (cioè eventi generati da  $X$ ), che è quello che ci interessa.

Per noi sarà sempre  $E = \mathbb{R}$ , dotato della  $\sigma$ -algebra  $\mathcal{E} = \mathcal{B}(\mathbb{R})$ , o  $E = \mathbb{R}^n$  (nel qual caso  $\mathbf{X}$  sarà detto vettore aleatorio), dotato della  $\sigma$ -algebra  $\mathcal{E} = \mathcal{B}(\mathbb{R}^n) = \mathcal{B}^n$ , che indica  $\sigma((a_1, b_1) \times (a_2, b_2) \times \dots \times (a_n, b_n) : -\infty \leq a_i < b_i \leq +\infty \ \forall i = 1, \dots, n)$ , cioè la  $\sigma$ -algebra generata dai rettangoli aperti di  $\mathbb{R}^n$ .

1. Se lancio due dadi distinti,  $\Omega = \{1, \dots, 6\}^2$ ,  $\mathcal{A} = 2^\Omega$ . Sia  $E = \mathbb{R}$ ,  $\mathcal{E} = \mathcal{B}(\mathbb{R})$ : data  $X : \Omega \rightarrow \mathbb{R}$ ,  $X(\omega) = \omega_1 + \omega_2$ ,  $X$  è una variabile aleatoria? E' vero che  $\forall B \in \mathcal{B}(\mathbb{R})$   $X^{-1}(B) \in 2^\Omega$ ? Sicuramente  $f^{-1}(B) \subseteq \Omega$ , ma allora  $f^{-1}(B) \in 2^\Omega$  e  $f$  è una variabile aleatoria.

Quindi, ogni volta che si hanno  $(\Omega, 2^\Omega), (E, \mathcal{E}), X : \Omega \rightarrow E$  è una variabile aleatoria.

2. Considero come spazi misurabili  $(\mathbb{R}, \mathcal{B}(\mathbb{R})), (\mathbb{R}, \mathcal{B}(\mathbb{R}))$  e  $X : \mathbb{R} \rightarrow \mathbb{R}$ ,  $X(\omega) = \lceil \omega \rceil$ .  $X(\Omega) = \mathbb{Z}$ .  $X$  è una variabile aleatoria? E' vero che  $\forall B \in \mathcal{B}(\mathbb{R})$   $X^{-1}(B) \in \mathcal{B}(\mathbb{R})$ ? Se  $B \in \mathcal{B}(\mathbb{R})$ ,  $X^{-1}(B) = (X \in B) = (X \in B) \cap \Omega = (X \in B) \cap (X \in \mathbb{Z})$ , per quanto osservato sopra. Ma per la proprietà 4 delle controimmagini  $(X \in B) \cap (X \in \mathbb{Z}) = (X \in (B \cap \mathbb{Z})) = \bigcup_{k \in (B \cap \mathbb{Z})} (X = k) = \bigcup_{k \in (B \cap \mathbb{Z})} (k - 1, k]$ , che è un'unione al più numerabile di intervalli. Poiché  $\mathcal{B}(\mathbb{R})$  contiene tutti gli intervalli di  $\mathbb{R}$  e anche tutte le loro unioni numerabili, essendo una  $\sigma$ -algebra,  $\bigcup_{k \in (B \cap \mathbb{Z})} (k - 1, k] \in \mathcal{B}(\mathbb{R})$  e  $X$  è una variabile aleatoria.
3. Considero come spazi misurabili  $(\mathbb{R}, \mathcal{B}(\mathbb{R})), (\mathbb{R}, \mathcal{B}(\mathbb{R}))$  e  $X : \mathbb{R} \rightarrow \mathbb{R}$ ,  $X(\omega) = \omega^2$ .  $X(\Omega) = [0, +\infty)$ . E' vero che  $\forall B \in \mathcal{B}(\mathbb{R})$   $X^{-1}(B) \in \mathcal{B}(\mathbb{R})$ ? Non riusciamo ancora a stabilirlo analiticamente, ma i criteri che seguono permettono di farlo.

### Teo 6.5 (misurabilità della composta di due misurabili)

Hp:  $(\Omega, \mathcal{A})$  spazio probabilizzabile,  $(E, \mathcal{E}), (F, \mathcal{F})$  spazi misurabili,

$X : \Omega \rightarrow E$  variabile aleatoria,  $h : E \rightarrow F$  funzione misurabile

Ts:  $Y = h \circ X : \Omega \rightarrow F$  è una variabile aleatoria

**Dim** Devo mostrare che  $\forall C \in \mathcal{F}$   $Y^{-1}(C) = (Y \in C) \in \mathcal{A}$ . Ma  $Y^{-1}(C) = (X^{-1}(h^{-1}(C))) = (X \in h^{-1}(C))$ :  $h^{-1}(C) \in \mathcal{E} \forall C \in \mathcal{F}$  perché  $h$  è misurabile, quindi, essendo  $h^{-1}(C)$  un generico sottinsieme di  $\mathcal{E}$ , si ha che  $X^{-1}(h^{-1}(C)) \in \mathcal{A}$  perché  $X$  è una variabile aleatoria, dunque  $Y$  è una variabile aleatoria. ■

### Teo (caratterizzazione della misurabilità)

Hp:  $(E, \mathcal{E}), (F, \mathcal{F})$  spazi misurabili,  $h : E \rightarrow F$ ,  $\mathcal{C} \subseteq \mathcal{F} : \sigma(\mathcal{C}) = \mathcal{F}$

Ts:  $h$  è misurabile  $\iff h^{-1}(B) \in \mathcal{E} \forall B \in \mathcal{C}$

Si può verificare la misurabilità di  $h$  controllando quindi solo le controimmagini di insiemi che formano un insieme che genera la  $\sigma$ -algebra di arrivo. 6.6 (i) esibisce un caso particolare di questo teorema.

**Dim\*** L'implicazione da sinistra a destra è ovvia perché  $h^{-1}(B) \in \mathcal{E} \forall B \in \mathcal{F}$ .

Se invece  $h^{-1}(B) \in \mathcal{E} \forall B \in \mathcal{C}$ , allora  $\mathcal{C}$  è contenuto nell'insieme dei  $B \in \mathcal{F} : h^{-1}(B) \in \mathcal{E}$ , indicato con  $\mathcal{D}$ . Quindi  $\mathcal{C} \subseteq \mathcal{D} \subseteq \mathcal{F}$ , per cui  $\sigma(\mathcal{C}) \subseteq \sigma(\mathcal{D}) \subseteq \sigma(\mathcal{F})$ , cioè  $\mathcal{F} \subseteq \sigma(\mathcal{D}) \subseteq \mathcal{F}$ , vale a dire  $\sigma(\mathcal{D}) = \mathcal{F}$ . Ma  $\mathcal{D}$  è una  $\sigma$ -algebra

su  $\mathcal{F}$  perché se  $B \in \mathcal{F}$  è tale che  $h^{-1}(B) \in \mathcal{E}$ , allora anche  $B^c \in \mathcal{F}$  è tale che  $h^{-1}(B^c) = (h^{-1}(B))^c \in \mathcal{E}$ , e sempre per le proprietà della controimmagine anche l'unione di  $B$  siffatti è in  $\mathcal{E}$ . Da questo si deduce che  $\sigma(\mathcal{D}) = \mathcal{D} = \mathcal{F}$ , cioè per ogni  $B \in \mathcal{F}$   $h^{-1}(B) \in \mathcal{E}$ , vale a dire che  $h$  è misurabile. ■

**Teo 6.6 (criteri di misurabilità)**

Hp:  $(\Omega, \mathcal{A})$  spazio probabilizzabile,  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$

(i)  $X : \Omega \rightarrow \mathbb{R}$  è una variabile aleatoria  $\iff (X \leq x) \in \mathcal{A} \forall x \in \mathbb{R}$

(ii)  $X = I_A : \Omega \rightarrow \mathbb{R}$  è una variabile aleatoria  $\iff A \in \mathcal{A}$

(iii)  $\mathbf{X} : \Omega \rightarrow \mathbb{R}^n$ ,  $\mathbf{X} = (X_1, \dots, X_n)$  è un vettore aleatorio  $\iff X_k : \Omega \rightarrow \mathbb{R}$   
è una variabile aleatoria  $\forall k = 1, \dots, n$

(iv) Hp:  $(\Omega, \mathcal{A})$  spazio probabilizzabile,  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ ,  $X_k : \Omega \rightarrow \mathbb{R}$  è una  
variabile aleatoria  $\forall k = 1, \dots, n$  e  $h : \mathbb{R}^n \rightarrow \mathbb{R}$  è misurabile

Ts:  $h \circ (X_1, \dots, X_n) : \Omega \rightarrow \mathbb{R}$  è una variabile aleatoria

(v) Hp:  $(\Omega, \mathcal{A})$  spazio probabilizzabile,  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ ,  $X, Y : \Omega \rightarrow \mathbb{R}$  sono v. a.

Ts:  $X + Y, X - Y, XY, X \vee Y, X \wedge Y, \frac{X}{Y} I_{(Y \neq 0)}$  sono misurabili

(vi) Hp:  $(\Omega, \mathcal{A})$  spazio probabilizzabile,  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ ,  $X_n : \Omega \rightarrow \mathbb{R}$  successione di v. a.

Ts :  $\sup_n X_n, \inf_n X_n, \lim_{n \rightarrow +\infty} X_n$ , se esiste, sono misurabili

(i) afferma che per valutare la misurabilità di  $X$  non occorre controllare che la controimmagine di ogni sottoinsieme di  $\mathcal{B}(\mathbb{R})$  stia in  $\mathcal{A}$ : è sufficiente controllare le controimmagini di un  $\pi$ -sistema: delle semirette, cioè gli insiemi  $(-\infty, x]$ . (iii) afferma che è possibile controllare la misurabilità componente per componente. (iv) è una generalizzazione di 6.5. (v) è ben definita perché in  $\mathbb{R}$  sono ben definite tutte le operazioni della tesi.  $\sup_n X_n$  è,  $\forall \omega \in \Omega$ ,  $\sup \{X_n(\omega) : n \in \mathbb{N}\}$ , ed è quindi una funzione di  $\omega$ .

**Dim\*** (i) è il teorema sopra nel caso particolare di  $F = \mathbb{R}$ ,  $\mathcal{F} = \mathcal{B}(\mathbb{R})$ , dato che  $\sigma(\{(-\infty, x] : x \in \mathbb{R}\}) = \mathcal{B}(\mathbb{R})$ .

(ii)  $\forall B \in \mathcal{B}(\mathbb{R})$   $I_A^{-1}(B) = (I_A \in B) \in \mathcal{A} \iff A \in \mathcal{A}$  perché  $I_A^{-1}(B)$  può essere solo  $A, A^c$  o  $\Omega$  per definizione di funzione indicatrice.

(iii) Se  $\mathbf{X} = (X_1, \dots, X_n)$  è un vettore aleatorio, allora  $\forall B \in \mathcal{B}(\mathbb{R}^n) \mathbf{X}^{-1}(B) \in \mathcal{A}$ . Voglio mostrare che  $\forall C \in \mathcal{B}(\mathbb{R}) X_1^{-1}(C) \in \mathcal{A}$ . Ma  $X_1^{-1}(C) = \mathbf{X}^{-1}(C \times \mathbb{R} \times \dots \times \mathbb{R})$ , che appartiene a  $\mathcal{A}$  per ipotesi, per cui  $X_1$  (e ogni altra componente) è misurabile.

Dimostro l'implicazione da destra a sinistra.  $\forall k = 1, \dots, n, \forall C \in \mathcal{B}(\mathbb{R}) X_k^{-1}(C) \in \mathcal{A}$ . Mostro che  $\forall B$  del tipo  $(-\infty, k_1] \times \dots \times (-\infty, k_n]$ , insieme generatore di  $\mathcal{B}(\mathbb{R}^n)$ , vale  $\mathbf{X}^{-1}(B) \in \mathcal{A}$ . Ma  $\mathbf{X}^{-1}(B) = \bigcap_{i=1}^n X_i^{-1}((-\infty, k_i])$ , che quindi appartiene ad  $\mathcal{A}$ . Allora, per il teorema sopra,  $\mathbf{X}$  è misurabile.

(v) deriva da (iv) nel caso particolare di  $n = 2$ , con  $h(x, y) = x + y, x - y, xy, x \vee y, x \wedge y, \frac{x}{y} I_{(y \neq 0)}$ .

(vi) Sappiamo che  $\lim_{n \rightarrow +\infty} X_n(\omega) = X(\omega) \forall \omega \in \Omega$ : mostro che allora  $X$  è una variabile aleatoria. Dimostro, usando (i), che  $(X \leq x)^c \in \mathcal{A} \forall x \in \mathbb{R}$ .  $(X \leq x)^c = \{\omega : X(\omega) > x\} = \{\omega : \lim_{n \rightarrow +\infty} X_n(\omega) > x\} = \bigcup_{r=1}^{+\infty} \{\omega : X_n(\omega) > x + \frac{1}{r} \text{ definitivamente per } n \rightarrow +\infty\} = \bigcup_{r=1}^{+\infty} \liminf_n \{\omega : X_n(\omega) > x + \frac{1}{r}\} = \bigcup_{r=1}^{+\infty} \bigcup_{n=1}^{+\infty} \bigcap_{k=n}^{+\infty} \{\omega : X_n(\omega) > x + \frac{1}{r}\}$  che è in  $\mathcal{A}$  perché ogni  $X_n$  è misurabile. ■

**Def 6.7** Dati gli spazi  $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n)), (\mathbb{R}^m, \mathcal{B}(\mathbb{R}^m))$ , una funzione  $h : \mathbb{R}^n \rightarrow \mathbb{R}^m$  misurabile si dice boreliana.

**Teo 6.8 (continua implica boreliana)**

Hp:  $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n)), (\mathbb{R}^m, \mathcal{B}(\mathbb{R}^m)), h : \mathbb{R}^n \rightarrow \mathbb{R}^m$  è una funzione continua

Ts:  $h$  è boreliana

**Dim\*** Mostro che  $h$  è misurabile mostrando che  $h^{-1}(B) \in \mathcal{E} \forall B \in \mathcal{C}$ , con  $\mathcal{C}$  insieme degli intervalli aperti di  $\mathbb{R}^m$ , che genera  $\mathcal{B}(\mathbb{R}^m)$ . Infatti, se  $h$  è continua, la controimmagine di ogni aperto è aperta e appartiene quindi a  $\mathcal{B}(\mathbb{R}^n)$ , cui appartengono tutti gli insiemi aperti (che si possono scrivere a partire da intervalli aperti). ■

1. Nell'esempio in cui  $X(\omega) = \omega^2$ ,  $X$  è una variabile aleatoria perché è continua, per 6.8.

2.  $X : \mathbb{R} \rightarrow \mathbb{R}^2, X(\omega) = \begin{pmatrix} \lceil \omega \rceil \\ \omega^2 \end{pmatrix}$  è un vettore aleatorio perché ogni componente è una v. a., per 6.6 (iii).

3. Se  $X : \Omega \rightarrow \mathbb{R}$  è una variabile aleatoria,  $Y = X^2 + \lceil X \rceil : \Omega \rightarrow \mathbb{R}$  è una variabile aleatoria perché somma di variabili aleatorie perché composizione di variabili aleatorie (6.5).

Dato  $(\Omega, \mathcal{A}, \mathbf{P})$  spazio di probabilità e  $(E, \mathcal{E})$  spazio misurabile, se  $X : \Omega \rightarrow E$  è una variabile aleatoria,  $\mathbf{P}(X \in B)$  al variare di  $B$  definisce una funzione.

**Def 7.1**  $(\Omega, \mathcal{A}, \mathbf{P})$  spazio di probabilità e  $(E, \mathcal{E})$  spazio misurabile, data  $X : \Omega \rightarrow E$  variabile aleatoria, si dice legge di  $X$  la funzione  $P^X : \mathcal{E} \rightarrow [0, 1], P^X(B) = \mathbf{P}(X \in B)$ .

La legge di  $X$  dunque associa a ogni elemento  $B$  della  $\sigma$ -algebra dello spazio di arrivo la probabilità che la variabile aleatoria appartenga a quell'elemento (che è un sottinsieme di  $E$ ), cioè la controimmagine di  $B$  attraverso  $X$ .  $\mathbf{P}(X \in B)$  è ben definita perché  $X$  è una variabile aleatoria, dunque  $X^{-1}(B) \in \mathcal{A}$  per ogni  $B$ .

E' naturale sperare che  $P^X$  sia una misura di probabilità su  $\mathcal{E}$ .

**Prop 7.2 (la legge di una v.a. è una probabilità)**

Hp:  $(\Omega, \mathcal{A}, \mathbf{P})$  è uno spazio di probabilità,  $(E, \mathcal{E})$  è uno spazio misurabile,

$X : \Omega \rightarrow E$  è una variabile aleatoria con legge  $P^X : \mathcal{E} \rightarrow [0, 1]$

Ts:  $P^X$  è una misura di probabilità su  $(E, \mathcal{E})$

La tesi significa che, dato uno spazio di probabilità, una variabile aleatoria tra tale spazio e un altro spazio misurabile induce su quest'ultimo una misura di probabilità, per cui lo spazio misurabile  $(E, \mathcal{E})$  diventa uno spazio di probabilità  $(E, \mathcal{E}, P^X)$ : ho costruito un nuovo modello probabilistico, che tiene conto solo degli eventi generati da  $X$ , cioè solo degli elementi di  $\mathcal{A}$  che si possono costruire come controimmagine attraverso  $X$ . Se tale spazio è noto, si può quindi lavorare su quello e ignorare lo spazio di partenza.

**Dim** Devo mostrare che  $P^X$  soddisfa gli assiomi A1, A2. E' vero che  $P^X(E) = 1$ ?  $P^X(E) = \mathbf{P}(X \in E) = \mathbf{P}(X^{-1}(E)) = \mathbf{P}(\Omega) = 1$  perché  $\mathbf{P}$  è una misura di probabilità su  $(\Omega, \mathcal{A})$ . E' vero che, data  $(B_n)_{n \geq 1}$  con  $B_n \in \mathcal{E}$   $\forall n$ ,  $B_h \cap B_k = \emptyset \forall h \neq k$ , vale  $P^X\left(\bigcup_{n=1}^{+\infty} B_n\right) = \sum_{n=1}^{+\infty} P^X(B_n)$ ? L'idea è sfruttare la definizione di  $P^X$  e poi il fatto che  $\mathbf{P}$  è una misura di probabilità.  $P^X\left(\bigcup_{n=1}^{+\infty} B_n\right) = \mathbf{P}\left(X \in \bigcup_{n=1}^{+\infty} B_n\right) = \mathbf{P}\left(X^{-1}\left(\bigcup_{n=1}^{+\infty} B_n\right)\right) = \mathbf{P}\left(\bigcup_{n=1}^{+\infty} X^{-1}(B_n)\right)$  per le proprietà della controimmagine; dato che  $\mathbf{P}$  è una misura di probabilità, essendo  $(X^{-1}(B_n))_{n \geq 1}$  una famiglia numerabile di elementi disgiunti (per la proprietà 4 della controimmagine, perché i  $B_n$  sono disgiunti) di  $\mathcal{A}$ ,  $\mathbf{P}\left(\bigcup_{n=1}^{+\infty} X^{-1}(B_n)\right) = \sum_{n=1}^{+\infty} \mathbf{P}(X^{-1}(B_n)) = \sum_{n=1}^{+\infty} P^X(B_n)$ , quindi vale A2. ■

1. L'insieme  $\{(X \in B) : B \in \mathcal{E}\} \subseteq \mathcal{A}$ , detto insieme degli eventi generati da  $X$  e indicato con  $\sigma(X)$ , è una  $\sigma$ -algebra sull'insieme  $\Omega$ . Infatti contiene  $\Omega$  (si ottiene per  $B = E$ ); è chiuso per complementazione, perché dato  $B \in \mathcal{E}$  e  $X^{-1}(B)$ ,  $(X^{-1}(B))^c$  appartiene ancora a  $\{(X \in B) : B \in \mathcal{E}\}$  perché  $(X^{-1}(B))^c = X^{-1}(B^c)$  per la proprietà 3 della controimmagine e perché  $B^c \in \mathcal{E}$ , essendo  $\mathcal{E}$  una  $\sigma$ -algebra. Inoltre, data  $(B_n)_{n \geq 1} \in \mathcal{E}$  e  $(X^{-1}(B_n))_{n \geq 1}$ ,  $\bigcup_{n=1}^{+\infty} (X^{-1}(B_n))$  appartiene ancora a  $\{(X \in B) : B \in \mathcal{E}\}$  perché  $\bigcup_{n=1}^{+\infty} (X^{-1}(B_n)) = X^{-1}\left(\bigcup_{n=1}^{+\infty} B_n\right)$  per la proprietà 4 della controimmagine e perché  $\bigcup_{n=1}^{+\infty} B_n \in \mathcal{E}$ , essendo  $\mathcal{E}$  una  $\sigma$ -algebra.

Se come spazi di partenza e di arrivo si considerano  $(\Omega, \mathcal{A}, \mathbf{P})$ ,  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$  e una variabile aleatoria  $X : \Omega \rightarrow \mathbb{R}$  con legge  $P^X$ ,  $P^X$  è una misura di probabilità su  $\mathcal{B}(\mathbb{R})$ , dunque è caratterizzata dalla sua funzione di ripartizione.

**Def 7.3** Dato  $(\Omega, \mathcal{A}, \mathbf{P})$  spazio di probabilità,  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$  spazio misurabile e  $X : \Omega \rightarrow \mathbb{R}$  variabile aleatoria con legge  $P^X$ , si dice funzione di ripartizione di  $X$  la funzione di ripartizione di  $P^X$   $F_X : \mathbb{R} \rightarrow [0, 1]$ ,  $F_X(t) = P^X((-\infty, t])$ .

Vale per definizione  $F_X(t) = P^X((-\infty, t]) = \mathbf{P}(X \in (-\infty, t]) = \mathbf{P}(X \leq t)$ .

Dicendo "funzione di ripartizione di  $X$ " si commette un abuso di notazione perché in realtà la funzione di ripartizione è associata a una probabilità: in questo caso, a  $P^X$ . Qualora esista la densità di probabilità discreta o continua associata a  $P^X$  (indicata rispettivamente con  $p_X$  o  $f_X$ ), conoscerla è equivalente a conoscere la legge  $P^X$ : si intenderà con legge, quindi, sia  $P^X$  che  $F_X$  che  $p_X/f_X$  (qualora esistano).

E' dunque possibile assegnare una misura di probabilità a  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$  specificando la funzione di ripartizione  $F_X$  di una variabile aleatoria a valori reali.

**Def** Dato  $(\Omega, \mathcal{A}, \mathbf{P})$  spazio di probabilità,  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$  spazio misurabile e  $X : \Omega \rightarrow \mathbb{R}$  variabile aleatoria con legge  $P^X$  e funzione di ripartizione  $F_X$ , dato  $\alpha \in [0, 1]$ , si dice quantile di  $X$  di ordine  $\alpha$ , e si indica con  $q_\alpha$ , ogni numero  $q_\alpha : \mathbf{P}(X \leq q_\alpha) \geq \alpha$  e  $\mathbf{P}(X < q_\alpha) \leq \alpha$ . Se  $\alpha = \frac{1}{2}$ ,  $q_\alpha$  si dice mediana.

1. Lancio due dadi equi e osservo gli esiti.  $\Omega = \{1, \dots, 6\}^2$ ,  $\mathcal{A} = 2^\Omega$ ,  $\mathbf{P}$  è la probabilità uniforme. Definisco la variabile aleatoria (si è già dimostrato che lo è)  $X : \Omega \rightarrow \mathbb{R}$ ,  $X(\omega) = \omega_1 + \omega_2 \ \forall \ \omega = (\omega_1, \omega_2) \in \Omega$ . Qual è la legge di  $X$ ?  $P^X = \mathbf{P}(X \in B) = \mathbf{P}(X^{-1}(B))$ . Dato che  $X(\Omega) = \{2, \dots, 12\} \subseteq \mathbb{R}$  è discreto,  $P^X$  è una probabilità discreta (vedi più avanti le definizioni equivalenti) e conoscerla è equivalente a conoscere  $p_X$  densità discreta,  $p_X : T \rightarrow \mathbb{R}$  con  $T = X(\Omega)$  e  $p_X(k) = P^X(\{k\}) = \mathbf{P}(X = k)$ : ma quest'ultima è nota, perché  $\mathbf{P}(X = k) = \frac{|\{(\omega_1, \omega_2) : \omega_1 + \omega_2 = k\}|}{36}$ : ( $X = k$ ) è un elemento di  $2^\Omega$ , e dato che  $\mathbf{P}$  è la probabilità uniforme si calcola come cardinalità di  $X^{-1}(\{k\})$  fratto  $|\Omega|$ . Si ricava che  $p_X(k) = \begin{cases} \frac{k-1}{36} & \text{se } k = 2, \dots, 6 \\ \frac{13-k}{36} & \text{se } k = 7, \dots, 12 \end{cases}$ . Dunque  $P^X : \mathcal{B}(\mathbb{R}) \rightarrow [0, 1]$  è in realtà una probabilità discreta con densità discreta  $p_X : T \rightarrow \mathbb{R}$  e  $P^X(B) = \sum_{k \in B \cap T} p_X(k)$ . In questo caso il supporto della densità  $p_X$  (si può anche dire supporto di  $X$ ) è  $S_X = T = \{k \in T : p_X(k) > 0\}$ . E' interessante notare che, nonostante  $p$  sullo spazio di partenza sia uniforme, la probabilità indotta da  $X$  su  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$  non è uniforme; e che  $p_X$  è individuata univocamente dallo spazio di probabilità  $(\Omega, \mathcal{A}, \mathbf{P})$  e dalla definizione di variabile aleatoria.

2. Sia  $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathbf{P})$  uno spazio di probabilità con  $\mathbf{P} = \varepsilon(\lambda)$ , sia  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$  uno spazio misurabile,  $X : \Omega \rightarrow \mathbb{R}$ ,  $X(\omega) = \lceil \omega \rceil$  (si è già dimostrato che è una v. a.). Qual è la legge di  $X$ ? In questo caso  $X(\Omega) = \mathbb{Z}$ , quindi  $P^X : \mathcal{B}(\mathbb{R}) \rightarrow [0, 1]$  è una probabilità discreta con  $p_X : T \rightarrow [0, 1]$ ,  $T = \mathbb{Z}$ . Dato  $k \in \mathbb{Z}$ ,  $p_X(k) = P^X(\{k\}) = \mathbf{P}(X = k) = \mathbf{P}((k-1, k])$  per quanto già visto su  $X(\omega) = \lceil \omega \rceil$ . Ora ricordo che  $\mathbf{P}((k-1, k]) =$

$$F(k) - F(k-1) = \int_{-\infty}^k f(t) dt - \int_{-\infty}^{k-1} f(t) dt = \int_{k-1}^k \lambda e^{-\lambda t} I_{(0,+\infty)}(t) dt = \begin{cases} 0 & \text{se } k \leq 0 \\ e^{-\lambda(k-1)} - e^{-\lambda k} & \text{se } k \geq 1 \end{cases}.$$

Dunque  $p_X(k) = \begin{cases} 0 & \text{se } k \leq 0 \\ e^{-\lambda(k-1)} (1 - e^{-\lambda}) & \text{se } k \geq 1 \end{cases}$ .  $S_X = \mathbb{N}^* = \mathbb{N} \setminus \{0\}$ : in questo caso  $S_X \subsetneq T$ . Noto che si può scrivere  $p_X(k) = (1-p)^{k-1} p$  se  $k \geq 1$ : si scrive che  $P^X \sim \text{geom}(1 - e^{-\lambda})$ .

3. Sia  $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathbf{P})$  uno spazio di probabilità con  $\mathbf{P}$  qualsiasi, sia  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$  uno spazio misurabile,  $X : \mathbb{R} \rightarrow \mathbb{R}$ ,  $X(\omega) = \omega$  (si è già dimostrato che è una v. a.). Qual è la legge di  $X$ ? In questo caso  $X(\Omega) = \mathbb{R}$ , quindi non posso usare la densità discreta per  $X$ .  $F_X(t) = P^X((-\infty, t]) = \mathbf{P}(X \in (-\infty, t]) = \mathbf{P}(\omega \in (-\infty, t]) = \mathbf{P}((-\infty, t])$  perché  $X^{-1}((-\infty, t]) = (-\infty, t]$ .  $\mathbf{P}((-\infty, t]) = F(t)$ : la legge di  $X$  ha come funzione di ripartizione la stessa di quella associata a  $\mathbf{P}$  sullo spazio di partenza, perché  $X$  è la funzione identità. In generale, prendendo come  $X$  l'identità, si replica sullo spazio di arrivo la probabilità dello spazio di partenza.

4. Sia  $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathbf{P})$  uno spazio di probabilità con  $\mathbf{P} = \varepsilon(\lambda)$ , sia  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$  uno spazio misurabile,  $X : \Omega \rightarrow \mathbb{R}$ ,  $X(\omega) = \omega^2$  (si è già dimostrato che è una variabile aleatoria). Qual è la legge di  $X$ ? In questo caso  $X(\Omega) = [0, +\infty)$ , quindi devo di nuovo usare la funzione di ripartizione di  $X$ .  $F_X(t) = P^X((-\infty, t]) = \mathbf{P}(X^{-1}((-\infty, t])) = \begin{cases} \mathbf{P}(\emptyset) & \text{se } t < 0 \\ \mathbf{P}([-\sqrt{t}, \sqrt{t}]) & \text{se } t \geq 0 \end{cases}$ . Se  $t \geq 0$ ,  $F_X(t) = \int_{-\sqrt{t}}^{\sqrt{t}} \lambda e^{-\lambda r} I_{(0,+\infty)}(r) dr = 1 - e^{-\lambda\sqrt{t}}$ , se  $t < 0$   $F_X(t) = 0$ .  $F_X$  è derivabile eccetto in 0, quindi esiste una densità continua di probabilità:  $f_X(t) = \frac{\lambda}{2\sqrt{t}} e^{-\lambda\sqrt{t}} I_{(0,+\infty)}(t)$ , che non è la densità esponenziale. Ora so tutto della legge di  $X$ : so calcolare ad esempio  $P^X((a, b)) = \mathbf{P}(X \in (a, b)) = \int_a^b f_X(t) dt$ .

5. Se  $X$  ha funzione di ripartizione  $F$  strettamente monotona, allora la variabile aleatoria  $F(X)$  ha distribuzione uniforme in  $(0, 1)$ . Infatti  $\mathbf{P}(F(X) \leq t)$  vale 0 se  $t < 0$ , 1 se  $t > 1$ ; se  $t \in (0, 1)$   $\mathbf{P}(F(X) \leq t) = \mathbf{P}(X \leq F^{-1}(t)) = F(F^{-1}(t)) = t$ , per cui  $F(X) \sim \mathcal{U}((0, 1))$ .

## 6.1 Relazione tra variabili aleatorie

**Def** Dati  $(\Omega, \mathcal{A}), (E, \mathcal{E})$  spazi misurabili e due variabili aleatorie  $X, Y : \Omega \rightarrow E$ , si dice che  $X = Y$  certamente se  $\forall \omega \in \Omega \ X(\omega) = Y(\omega)$ .

Due v. a. sono uguali certamente se sono la stessa funzione: questa definizione non aggiunge niente alla definizione già vista in analisi di uguaglianza tra funzioni; non è influenzata dal fatto che  $X, Y$  sono v.a. Dal punto di vista probabilistico, significa che si effettua un unico esperimento con due osservatori.

Però, dato che  $X, Y$  sono variabili aleatorie, a noi interessa anche la loro legge: si introduce allora una definizione di uguaglianza più debole, tra le leggi.

**Def** Dati  $(\Omega_1, \mathcal{A}_1, \mathbf{P}_1), (\Omega_2, \mathcal{A}_2, \mathbf{P}_2)$  spazi di probabilità e  $(E, \mathcal{E})$  spazio misurabile, date due variabili aleatorie  $X : \Omega_1 \rightarrow E, Y : \Omega_2 \rightarrow E$  con leggi  $P^X, P^Y$ , si dice che  $X$  e  $Y$  hanno la stessa legge, e si scrive  $P^X = P^Y$  ( $X \sim Y$ ) se  $\forall B \in \mathcal{E} P^X(B) = P^Y(B)$ .

(Quando si parla di leggi devo specificare le  $\mathbf{P}$  nelle hp della def!) Due v. a. possono quindi avere la stessa legge anche se definite su spazi misurabili diversi, perché la definizione richiede che  $P^X : \mathcal{E} \rightarrow [0, 1], P^Y : \mathcal{E} \rightarrow [0, 1]$  siano la stessa funzione: per confrontare le due leggi è sufficiente che entrambe siano definite sulla stesse  $\sigma$ -algebra.

Se  $X = Y$  certamente ovviamente  $P^X = P^Y$ : il viceversa non è vero, com'è ovvio dal fatto che è possibile che  $X, Y$  siano definite su domini diversi, pur avendo la stessa legge.

1. Lancio un dado rosso e un dado blu.  $\Omega = \{1, \dots, 6\}^2, \mathcal{A} = 2^\Omega, \mathbf{P}$  è la probabilità uniforme, lo spazio misurabile di arrivo è  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ .  $X, Y : \Omega \rightarrow \mathbb{R}$  descrivono l'esito del lancio rispettivamente sul dado rosso e su quello blu, cioè  $X(\omega) = \omega_1, Y(\omega) = \omega_2$ .  $X$  e  $Y$  non sono uguali certamente: se  $\omega = (1, 2)$ ,  $X(\omega) \neq Y(\omega)$ . Tuttavia, poiché  $X(\Omega) = Y(\Omega) = \{1, \dots, 6\} = T$ ,  $P^X, P^Y$  sono probabilità discrete, e sono la stessa funzione perché coincidono sugli atomi:  $P^X(\{k\}) = p_X(k) = \frac{1}{6} = P^Y(\{k\}) = p_Y(k) \forall k \in T$ . Quindi  $X$  e  $Y$  hanno la stessa legge.

**Def 7.4** Dato  $(\Omega, \mathcal{A}, \mathbf{P})$  spazio di probabilità,  $(E, \mathcal{E})$  spazio misurabile e  $X : \Omega \rightarrow E, Y : \Omega \rightarrow E$  variabili aleatorie, si dice che  $X$  e  $Y$  sono uguali quasi certamente, e si scrive  $X = Y$  q. c., se  $\exists A \in \mathcal{A} : \mathbf{P}(A) = 1$  e  $\forall \omega \in A X(\omega) = Y(\omega)$ .

Due v. a. si dicono quindi uguali quasi certamente se, ristrette a un evento di probabilità 1, sono la stessa funzione, cioè se coincidono su un evento quasi certo: una volta assegnato  $(\Omega, \mathcal{A}, \mathbf{P})$ , si sa quindi che alcuni eventi sono trascurabili, cioè hanno probabilità nulla. La definizione dipende dalla misura di probabilità scelta per  $(\Omega, \mathcal{A})$ .

Se  $E = \mathbb{R}$  o  $\mathbb{R}^n$ ,  $X - Y = Z$  è ancora una v. a. per 6.6 (v), quindi  $\forall B \in \mathcal{B}(\mathbb{R}) Z^{-1}(B) \in \mathcal{A}$ , e in particolare  $Z^{-1}(\{0\}) = (X - Y = 0) \in \mathcal{A}$ .  $X = Y$  q. c.  $\iff \mathbf{P}(X - Y = 0) = 1$ . Infatti se  $X = Y$  q. c.,  $\exists A \in \mathcal{A} : \mathbf{P}(A) = 1$  e  $\forall \omega \in A X(\omega) = Y(\omega)$ : allora  $A \subseteq Z^{-1}(\{0\})$ , per cui  $\mathbf{P}(Z^{-1}(\{0\})) = \mathbf{P}(X = Y) \geq \mathbf{P}(A) = 1$ , cioè  $\mathbf{P}(X = Y) = 1$ . Se invece so che  $\mathbf{P}(Z^{-1}(\{0\})) = 1$ , allora  $X, Y$  sono uguali q. c. con  $A = Z^{-1}(\{0\})$ .

In generale, per capire se due v. a. sono uguali quasi certamente è sufficiente calcolare la probabilità del più grande sottinsieme di  $\Omega$  su cui esse coincidono: è 1 se e solo se sono uguali quasi certamente.

La relazione di uguaglianza quasi certa gode della proprietà riflessiva e simmetrica. Inoltre, date tre v. a.  $X, Y, Z : \Omega \rightarrow E$ , se  $X$  e  $Y$  sono uguali quasi certamente  $\exists A \in \mathcal{A} : \mathbf{P}(A) = 1$  e  $\forall \omega \in A X(\omega) = Y(\omega)$ ;



se  $Y$  e  $Z$  sono uguali quasi certamente  $\exists B \in \mathcal{A} : \mathbf{P}(B) = 1$  e  $\forall \omega \in B \ Y(\omega) = Z(\omega)$ . Allora  $\mathbf{P}(A \cap B) = \mathbf{P}(A) + \mathbf{P}(B) - \mathbf{P}(A \cup B) = 2 - \mathbf{P}(A \cup B)$ ; poiché  $\mathbf{P}(A \cup B) \geq \mathbf{P}(A)$ ,  $\mathbf{P}(A \cup B) = 1$ , dunque  $\mathbf{P}(A \cap B) = 1$ . Allora  $\mathbf{P}(A \cap B) = 1$  e  $X(\omega) = Z(\omega) \ \forall \omega \in (A \cap B)$ , cioè anche  $X$  e  $Z$  sono uguali quasi certamente: l'uguaglianza quasi certa gode della proprietà transitiva.

1. Considero i due spazi  $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \varepsilon(\lambda)), (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ . Considero  $X_1, X_2 : \Omega \rightarrow \mathbb{R}$ ,  $X_1(\omega) = \omega$ ,  $X_2(\omega) = |\omega|$ ,  

$$X_3(\omega) = \omega I_{\mathbb{R} \setminus \mathbb{Q}}(\omega) = \begin{cases} \omega & \text{se } \omega \in \mathbb{R} \setminus \mathbb{Q} \\ 0 & \text{se } \omega \in \mathbb{Q} \end{cases} . \quad X_1 \neq X_2 \ \forall \omega \in (-\infty, 0) . \quad \text{Tuttavia, } X_1 = X_2 \text{ q. c., perché}$$

$$X_1(\omega) = X_2(\omega) \ \forall \omega \in [0, +\infty), \text{ e se } A = [0, +\infty), \ \mathbf{P}(A) = 1 - F(0) = \int_0^{+\infty} \lambda e^{\lambda t} dt = 1 .$$
Anche se  $X_1$  e  $X_2$  come funzioni non coincidono, la probabilità assegnata allo spazio misurabile fa sì che coincidano su un evento quasi certo.

Analogamente,  $X_1 \neq X_3 \ \forall \omega \in \mathbb{Q} \setminus \{0\}$ . Tuttavia,  $X_1 = X_3$  q. c., perché  $X_1(\omega) = X_3(\omega) \ \forall \omega \in \mathbb{R} \setminus \mathbb{Q}$ , e se  $A = \mathbb{R} \setminus \mathbb{Q}$ ,  $\mathbf{P}(A) = \mathbf{P}(\mathbb{R}) - \mathbf{P}(\mathbb{Q}) = 1 - \mathbf{P}(\mathbb{Q}) = 1 - \sum_{q \in \mathbb{Q}} \mathbf{P}(\{q\}) = 1$  per la  $\sigma$ -additività, essendo  $\mathbb{Q}$  discreto, e perché la probabilità di qualsiasi singoletto è nulla per probabilità che ammettano una densità continua. Anche se  $X_1$  e  $X_3$  come funzioni non coincidono, la probabilità assegnata allo spazio misurabile fa sì che coincidano su un evento quasi certo.

2. Considero i due spazi  $((0, 1), \mathcal{B}((0, 1)), \mathcal{U}([0, 1))), (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ , e  $X, Y : (0, 1) \rightarrow \mathbb{R}$ ,  $X(\omega) = \omega$ ,  $Y(\omega) = 1 - \omega$ .  
 $P^X(B) = \mathbf{P}(X \in B)$ ,  $P^Y(B) = \mathbf{P}(Y \in B)$ . Usando la f. d. r. delle variabili aleatorie,  $F_X(t) =$   

$$P^X((-\infty, t]) = \mathbf{P}(X^{-1}((-\infty, t])) = \begin{cases} 0 & \text{se } t \leq 0 \\ \mathbf{P}((0, t]) = t & \text{se } t \in (0, 1) \\ \mathbf{P}((0, 1)) = 1 & \text{se } t \geq 1 \end{cases} = F(t) \text{ (anche } P^X \text{ è uniforme). Invece}$$

$$F_Y(t) = P^Y((-\infty, t]) = \mathbf{P}(Y^{-1}((-\infty, t])) = \begin{cases} 0 & \text{se } t \leq 0 \\ \mathbf{P}([1-t, 1)) = t & \text{se } t \in (0, 1) \\ \mathbf{P}((0, 1)) = 1 & \text{se } t \geq 1 \end{cases} . \quad \text{Quindi } F_Y(t) = F_X(t) \ \forall t \text{ e}$$

$$P^X(B) = P^Y(B) \ \forall B, \text{ ma } \mathbf{P}(X = Y) = \mathbf{P}\left(\left\{\frac{1}{2}\right\}\right) = 0 \text{ perché ogni singoletto è un evento di probabilità nulla}$$
se la probabilità è continua. Quindi  $X$  e  $Y$  non sono uguali q. c., pur avendo la stessa legge.

**Prop 7.5 (uguaglianza certa, quasi certa, in legge)**

Hp:  $(\Omega, \mathcal{A}, \mathbf{P})$  è uno spazio di probabilità,  $(E, \mathcal{E})$  è uno spazio

misurabile,  $X, Y : \Omega \rightarrow E$  sono due variabili aleatorie

Ts: (i)  $X = Y \implies$  (ii)  $X = Y$  q. c.  $\implies$  (iii)  $P^X = P^Y$

Il fatto che (ii) $\nRightarrow$ (i) è mostrato dall'esempio 1 sopra con  $X_1, X_2$  o  $X_1, X_3$ . Il fatto che (iii) $\nRightarrow$ (ii) è mostrato dall'esempio 2 sopra con  $X_1, X_2$ . La proposizione mostra che l'uguaglianza quasi certa è una relazione di "forza intermedia" tra l'uguaglianza certa e l'uguaglianza tra le leggi.

**Dim** (i) $\Rightarrow$ (ii) Se  $X = Y$ , significa che  $X(\omega) = Y(\omega) \forall \omega \in \Omega$ . Se si prende  $A = \Omega$ , è evidente che  $\exists A \in \mathcal{A} : \mathbf{P}(A) = 1$ , perché  $\mathbf{P}(\Omega) = 1$ .

(ii) $\Rightarrow$ (iii) Se  $X = Y$  q. c., significa che  $\exists A \in \mathcal{A} : \mathbf{P}(A) = 1$  e  $X(\omega) = Y(\omega) \forall \omega \in A$ . Dato  $B \in \mathcal{E}$ ,  $P^X(B) = \mathbf{P}(X^{-1}(B))$ ,  $P^Y(B) = \mathbf{P}(Y^{-1}(B))$ .  $\mathbf{P}(X \in B) = \mathbf{P}((X \in B) \cap A) = \mathbf{P}((Y \in B) \cap A) = \mathbf{P}(Y \in B)$ . La penultima uguaglianza è dovuta al fatto che  $(X^{-1}(B)) \cap A = \{\omega \in \Omega : \omega \in A \wedge X(\omega) \in B\}$ : ogni elemento di questo insieme è anche in  $Y^{-1}(B) \cap A$ , perché  $X(\omega) = Y(\omega) \forall \omega \in A$ , e viceversa, dunque  $Y^{-1}(B) \cap A = X^{-1}(B) \cap A$ .

[\*In generale infatti, se  $\mathbf{P}(A) = 1$ ,  $C = (C \cap A) \cup (C \cap A^c)$ , unione finita di eventi disgiunti:  $\mathbf{P}(C) = \mathbf{P}(C \cap A) + \mathbf{P}(C \cap A^c) = \mathbf{P}(C \cap A)$  perché  $\mathbf{P}(C \cap A^c) \leq \mathbf{P}(A^c) = 1 - \mathbf{P}(A) = 0$ ] ■

**Def** Dato  $(\Omega, \mathcal{A}, \mathbf{P})$  spazio di probabilità, si dice che una proprietà  $P(\omega)$  che dipende da  $\omega \in \Omega$  vale quasi certamente se  $\exists A \subseteq \mathcal{A} : \mathbf{P}(A) = 1$  e  $P(\omega)$  vale  $\forall \omega \in A$ .

Alternativamente, si dice che  $P(\omega)$  vale quasi certamente se  $\mathbf{P}(\{\omega : \neg P(\omega)\}) = 0$ .

Se si considera il più grande sottinsieme di  $\Omega$ , che chiamo  $A_{\max}$ , in cui vale la proprietà  $p$  e si trova che  $\mathbf{P}(A_{\max}) = 0$ , allora si può concludere che  $p$  non vale quasi certamente, per monotonia della probabilità: su ogni  $A \subseteq A_{\max}$ , in cui varrà  $p$ , sarà ancora  $\mathbf{P}(A) = 0$ . Se invece si trova  $\mathbf{P}(A_{\max}) = 1$ , è possibile che esistano  $A \subset A_{\max} : \mathbf{P}(A) = 1$ .

1. Considero i due spazi  $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathcal{U}(0, 1))$ ,  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$  e la variabile aleatoria  $X : \mathbb{R} \rightarrow \mathbb{R}$ ,  $X(\omega) = \omega$ .  $X(\Omega) = \mathbb{R}$ .

Dimostro che tuttavia  $X \in [0, 1]$  quasi certamente, cioè che  $\exists A \in \mathcal{B}(\mathbb{R}) : \mathbf{P}(A) = 1$  e  $X(\omega) \in [0, 1] \forall \omega \in A$ .

Se considero  $A = (0, 1)$ , è ovvio che  $X(\omega) \in [0, 1] \forall \omega \in A$ , ma soprattutto  $\mathbf{P}(A) = \int_0^1 1 dt = 1$ . Infatti

$\mathbf{P}(X \in [0, 1]) = \mathbf{P}(\omega \in [0, 1]) = 1$ : nonostante l'immagine di  $X$  sia  $\mathbb{R}$ , i valori che  $X$  assume sono governati da  $\mathbf{P}$  perché questa determina la probabilità dei sottinsiemi di  $\Omega$ , cioè degli input della  $X$ .

## 6.2 Integrazione rispetto a una misura di probabilità

Considero  $(\Omega, \mathcal{A}, \mathbf{P})$  spazio di probabilità e  $X : \Omega \rightarrow \mathbb{R}$  variabile aleatoria reale. E' naturale voler dire qualcosa sui valori di  $X$  prima di effettuare l'esperimento, e. g. voler trovare "il centro" di  $X$ , che rispetti l'idea di valore "centrale" rispetto ai possibili valori di  $X(\omega)$  al variare di  $\omega \in \Omega$ .

Tale valore centrale, di cui si darà una definizione rigorosa nel seguito, si indica in vari modi:  $\mathbf{E}(X)$ ,  $\mu$ ,  $\int_{\Omega} X(\omega) d\mathbf{P}(\omega)$ ,  $\int_{\Omega} X(\omega) \mathbf{P}(d\omega)$ ,  $\int_{\Omega} X d\mathbf{P}$ . Si costruisce ora un oggetto matematico che rispetti l'intuizione modellistica di  $\mathbf{E}(X)$ .

#### PASSO 1

**Def 8.1** Dato  $(\Omega, \mathcal{A})$  spazio probabilizzabile e  $X : \Omega \rightarrow \mathbb{R}$  variabile aleatoria reale,  $X$  si dice semplice se assume un numero finito di valori, cioè se  $|Im(X)| = n < +\infty$ .

Questo è equivalente a dire che si può scrivere, data  $\{A_1, A_2, \dots, A_n\}$  partizione finita di  $\Omega$  e  $x_1, \dots, x_n \in \mathbb{R}$ ,  $X(\omega) = \sum_{k=1}^n x_k I_{A_k}(\omega) = x_1 I_{A_1}(\omega) + \dots + x_n I_{A_n}(\omega)$ . Se  $\omega \in A_1$ ,  $X(\omega) = x_1$ , eccetera.  $X^{-1}(\{x_k\}) = (X = x_k) = A_k$ ,  $X(\Omega) = \{x_1, \dots, x_n\}$ .

Nelle ipotesi della definizione non occorre specificare la probabilità dello spazio.

**Def 8.2** Dato  $(\Omega, \mathcal{A}, \mathbf{P})$  spazio di probabilità e  $X : \Omega \rightarrow \mathbb{R}$  variabile aleatoria reale semplice, si dice valore atteso di  $X$ , e si indica con  $\mathbf{E}(X)$ ,  $\sum_{k=1}^n x_k \mathbf{P}(X = x_k)$ .

La definizione è ben posta perché  $(X = x_k) \in \mathcal{A}$ , essendo  $X$  una v.a.  $\sum_{k=1}^n x_k \mathbf{P}(X = x_k)$  è proprio come calcolare l'integrale di Riemann di una funzione a scala (in quel caso si considera una successione di funzioni a scala), ma nell'integrale di Riemann si calcola  $\sum_{k=1}^n f(I_k) \mu(I_k)$  dove  $\mu(I_k) = x_{k+1} - x_k$ .

Il valore atteso di una v. a. r. semplice è quindi una media dei valori che essa assume, *pesata sulle probabilità dei valori*: dati i valori  $x_1, \dots, x_n$  sull'asse reale,  $\mathbf{E}(X)$  individua il centro di tale distribuzione in base alla misura di probabilità dello spazio.  $\mathbf{E}(X) \in \mathbb{R}$ ; è possibile che  $\mathbf{E}(X) \notin Im(X)$ .

1. Dato  $(\Omega, \mathcal{A}, \mathbf{P})$  spazio di probabilità e  $A \in \mathcal{A}$ ,  $X(\omega) = I_A(\omega)$  è una v. a. semplice tale che  $\mathbf{E}(X) = \mathbf{P}(A)$ .

#### Prop (proprietà del valore atteso di v. a. semplici)

Hp:  $(\Omega, \mathcal{A}, \mathbf{P})$  spazio di probabilità,  $X, Y : \Omega \rightarrow \mathbb{R}$

variabili aleatorie reali semplici,  $\alpha, \beta \in \mathbb{R}$

Ts: (i)  $\alpha X + \beta Y$  è una v. a. reale semplice

(ii) (linearità)  $\mathbf{E}(\alpha X + \beta Y) = \alpha \mathbf{E}(X) + \beta \mathbf{E}(Y)$

(iii) (monotonia) se  $X(\omega) \leq Y(\omega) \forall \omega \in \Omega$ ,  $\mathbf{E}(X) \leq \mathbf{E}(Y)$

1. Se  $Y : \Omega \rightarrow [0, +\infty]$  è una variabile aleatoria semplice e  $X : \Omega \rightarrow \mathbb{R}$ ,  $X(\omega) = 0 \leq Y(\omega)$ , allora  $X$  è una v. a. r. semplice con  $\mathbf{E}(X) = 0$ , per cui  $\mathbf{E}(Y) \geq 0$ .

#### PASSO 2

**Prop (approssimazione di una funzione)**

Hp:  $(\Omega, \mathcal{A})$  spazio misurabile,  $X : \Omega \rightarrow [0, +\infty]$  v. a. reale

Ts:  $\exists (X_n)_{n \geq 1} : (1) X_n : \Omega \rightarrow [0, +\infty]$  è semplice  $\forall n \geq 1$ ,

(2)  $\forall \omega \in \Omega \lim_{n \rightarrow +\infty} X_n(\omega) = X(\omega)$  e  $(X_n)_{n \geq 1}$  è monotona non decrescente

$X : \Omega \rightarrow [0, +\infty]$  ha come  $\sigma$ -algebra sullo spazio di arrivo quella generata dagli aperti contenuti in  $[0, +\infty]$ .  
 $(X_n)$  monotona non decrescente significa che  $X_n(\omega) \leq X_{n+1}(\omega) \forall n, \forall \omega$ : (2) afferma quindi che  $(X_n)_{n \geq 1}$  converge puntualmente a  $X$  (più avanti questa verrà chiamata convergenza certa) in maniera monotona: si scrive  $X_n \nearrow X$ .

Il senso della proposizione è che qualsiasi variabile aleatoria reale non negativa può essere scritta come limite di una successione di variabili aleatorie reali semplici. Come costruisco tale successione? Una possibilità è

$$\text{fissato } n \geq 1, \text{ al variare di } k = 0, \dots, n2^n - 1, X_n(\omega) = \begin{cases} \frac{k}{2^n} & \text{se } X(\omega) \in [\frac{k}{2^n}, \frac{k+1}{2^n}) \\ n & \text{se } X(\omega) \geq n \end{cases}$$

$$\text{Quindi, fissato } n, X_n(\omega) = \begin{cases} 0 & \text{se } X(\omega) \in [0, \frac{1}{2^n}) \\ \frac{1}{2^n} & \text{se } X(\omega) \in [\frac{1}{2^n}, \frac{2}{2^n}) \\ \frac{2}{2^n} & \text{se } X(\omega) \in [\frac{2}{2^n}, \frac{3}{2^n}) \\ \dots & \dots \\ \frac{n2^n-1}{2^n} & \text{se } X(\omega) \in [\frac{n2^n-1}{2^n}, n) \\ n & \text{se } X(\omega) \geq n \end{cases} \quad \text{è un'approssimazione per difetto di } X: \text{ quando } X$$

assume valori in  $[a, b]$ ,  $X_n$  assume valore  $a$ . Al crescere di  $n$ ,  $X_n$  approssima  $X$  sempre più finemente, perché la partizione dell'asse  $y$  è più fitta. Chiaramente  $X_n$  è una v. a. semplice:  $Im(X_n) = \{0, \frac{1}{2^n}, \dots, n\}$ , con cardinalità  $n2^n + 1$ ; per definizione  $\mathbf{E}(X_n) = \sum_{k=0}^{n2^n-1} \frac{k}{2^n} \mathbf{P}(X \in [\frac{k}{2^n}, \frac{k+1}{2^n})) + n\mathbf{P}(X \geq n)$ . La definizione data fornisce di fatto una definizione di  $X_n$  su tutto  $\Omega$ , ma ottenuta partizionando il codominio di  $X$  (cioè  $[0, +\infty]$ ), dato che  $X^{-1}(X(\Omega)) = \Omega$ . Questa è l'idea alla base dell'integrale di Lebesgue.

$$1. \text{ Se } n = 1, k = 0 \text{ o } k = 1, \text{ quindi si avranno tre tratti: } X_1(\omega) = \begin{cases} 0 & \text{se } X(\omega) \in [0, \frac{1}{2}) \\ \frac{1}{2} & \text{se } X(\omega) \in [\frac{1}{2}, 1) \\ 1 & \text{se } X(\omega) \geq 1 \end{cases}$$

Si dimostra che  $X_{n+1}(\omega) \geq X_n(\omega) \forall \omega \in \Omega, \forall n$ , quindi  $\mathbf{E}(X_{n+1}) \geq \mathbf{E}(X_n) \geq 0$  per monotonia del valore atteso: allora  $(\mathbf{E}(X_n))_{n \geq 1}$  è una successione a valori reali non negativi monotona non decrescente, che può tendere a  $L \geq 0$  o  $+\infty$  per  $n \rightarrow +\infty$ .

**Def 8.3** Dato  $(\Omega, \mathcal{A}, \mathbf{P})$  spazio di probabilità e  $X : \Omega \rightarrow [0, +\infty]$  variabile aleatoria reale non negativa, si dice valore atteso di  $X$ , e si indica con  $\mathbf{E}(X)$ ,  $\lim_{n \rightarrow +\infty} \mathbf{E}(X_n)$ .

Il valore atteso può essere  $+\infty$ . Per come sono stati costruite le  $X_n$ , vale anche  $\mathbf{E}(X) = \lim_{n \rightarrow +\infty} \left( \sum_{k=0}^{n2^n-1} \frac{k}{2^n} \mathbf{P}(X \in [\frac{k}{2^n}, \frac{k+1}{2^n}]) \right)$ . In realtà la definizione di valore atteso non dipende dalla scelta di  $(X_n)_{n \geq 1}$ , purché sia una successione di v. a. semplici che cresce a  $X$ .

Se  $X$  è una v.a.r. semplice non negativa, la 8.3 e la 8.2 sono coerenti.

Se  $\mathbf{P}(X = +\infty) > 0$ ,  $\mathbf{E}(X) = +\infty$ . Infatti  $\mathbf{E}(X) \geq \lim_{n \rightarrow +\infty} n\mathbf{P}(X \geq n)$ , ma  $X^{-1}([n, +\infty)) \supseteq (X = +\infty) = \bigcap_{n=1}^{+\infty} (X \geq n)$ , quindi  $\mathbf{P}(X \geq n) \geq \mathbf{P}(X = +\infty)$  e  $\lim_{n \rightarrow +\infty} n\mathbf{P}(X \geq n) \geq \lim_{n \rightarrow +\infty} n\mathbf{P}(X = +\infty) = +\infty$ , quindi  $\mathbf{E}(X) = +\infty$ .

PASSO 3

**Def 8.4** Data  $X : \Omega \rightarrow \mathbb{R}$ , si definisce parte positiva di  $X$ , e si indica con  $X_+$ , la funzione  $X_+ : \Omega \rightarrow [0, +\infty)$ ,  $X_+(\omega) = \max\{X(\omega), 0\} = X(\omega) \vee 0$ ; si definisce parte negativa di  $X$ , e si indica con  $X_-$ , la funzione  $X_- : \Omega \rightarrow [0, +\infty)$ ,  $X_-(\omega) = -\min\{X(\omega), 0\} = -X(\omega) \wedge 0$ .

Vale quindi  $X(\omega) = X_+(\omega) - X_-(\omega)$ , mentre  $|X(\omega)| = X_+(\omega) + X_-(\omega)$ . Questa definizione permette di generalizzare ulteriormente la definizione di valore atteso: infatti  $X_+, X_-$  sono v.a.r. non negative (infatti si possono scrivere come prodotto di  $X$  per  $I_A$  dove  $A = \{\omega \in \Omega : X(\omega) > 0\}$  e  $A = \{\omega \in \Omega : X(\omega) < 0\}$  rispettivamente, e  $A \in \mathcal{A}$  perché  $A = (X \in (0, +\infty))$  e  $X$  è una variabile aleatoria, quindi  $X_+$  è il prodotto di due variabili aleatorie), e per definire  $\mathbf{E}(X)$  ci si può ricondurre alla definizione precedente.

**Def 8.5** Dato  $(\Omega, \mathcal{A}, \mathbf{P})$  spazio di probabilità e  $X : \Omega \rightarrow \mathbb{R}$  variabile aleatoria con parte positiva  $X_+$  e parte negativa  $X_-$ , si dice che  $X$  ammette valore atteso se almeno uno tra  $\mathbf{E}(X_+)$  e  $\mathbf{E}(X_-)$  è finito. In tal caso si dice valore atteso di  $X$ , e si indica con  $\mathbf{E}(X)$ ,  $\mathbf{E}(X_+) - \mathbf{E}(X_-)$ . Se  $\mathbf{E}(X_+) = \mathbf{E}(X_-) = +\infty$ , si dice che  $X$  non ammette valore atteso.

Il valore atteso così definito appartiene a  $[-\infty, +\infty] = \mathbb{R}^*$ . Se  $X$  è una v.a.r. semplice non negativa, la 8.5, 8.3 e 8.2 sono coerenti. Se  $X$  è una v.a.r. non negativa, la 8.3 e 8.5 sono coerenti perché  $X_-(\omega) = 0$ .

**Def 8.6** Dato  $(\Omega, \mathcal{A}, \mathbf{P})$  spazio di probabilità e  $X : \Omega \rightarrow \mathbb{R}$  variabile aleatoria con parte positiva  $X_+$  e parte negativa  $X_-$ , si dice che  $X$  è integrabile, o che ammette valore atteso finito, se  $\mathbf{E}(X_+)$  e  $\mathbf{E}(X_-)$  sono finiti.

In tal caso il valore atteso di  $X$  è  $\mathbf{E}(X) = \mathbf{E}(X_+) - \mathbf{E}(X_-) \in \mathbb{R}$ .

**Def 8.7** Dato  $(\Omega, \mathcal{A}, \mathbf{P})$  spazio di probabilità, si indica con  $\mathcal{L}^1(\Omega, \mathcal{A}, \mathbf{P})$  lo spazio delle variabili aleatorie reali integrabili, cioè  $\mathcal{L}^1(\Omega, \mathcal{A}, \mathbf{P}) = \{X : \Omega \rightarrow \mathbb{R} \text{ variabili aleatorie} : X \text{ ammette valore atteso finito}\}$ .

Il vantaggio dell'astrazione nella definizione di  $\mathbf{E}(X)$  è che è possibile ricavare facilmente proprietà generali del valore atteso, senza ipotesi su  $(\Omega, \mathcal{A}, \mathbf{P})$ .

**Teo 8.8 (proprietà del valore atteso)**

Sia  $(\Omega, \mathcal{A}, \mathbf{P})$  uno spazio di probabilità.

1)  $\mathcal{L}^1(\Omega, \mathcal{A}, \mathbf{P})$  è uno spazio vettoriale

2) Hp:  $\mathbf{E} : \mathcal{L}^1(\Omega, \mathcal{A}, \mathbf{P}) \rightarrow \mathbb{R}$ ,  $X, Y \in \mathcal{L}^1$ ,  $\alpha, \beta \in \mathbb{R}$

Ts: (i) (linearità)  $\mathbf{E}(\alpha X + \beta Y) = \alpha \mathbf{E}(X) + \beta \mathbf{E}(Y)$

(ii) (positività) se  $X \geq 0$ ,  $\mathbf{E}(X) \geq 0$

1) significa che se  $X, Y \in \mathcal{L}^1$  e  $\alpha, \beta \in \mathbb{R}$ ,  $\alpha X + \beta Y \in \mathcal{L}^1$ , cioè se  $\mathbf{E}(X), \mathbf{E}(Y)$  sono finiti, allora anche  $\mathbf{E}(\alpha X + \beta Y)$  è finito (proprio come l'insieme delle funzioni Riemann integrabili è uno spazio vettoriale). Questo fa sì che l'uguaglianza  $\mathbf{E}(\alpha X + \beta Y) = \alpha \mathbf{E}(X) + \beta \mathbf{E}(Y)$  in 2) (i) sia ben posta: si può effettivamente calcolare  $\mathbf{E}(\alpha X + \beta Y)$ ; in 2) (i) sono coinvolti solo numeri reali, e la proprietà è analoga alla linearità dell'integrale di Riemann. 2) (i),(ii) si sono dimostrate sopra per v.a. semplici.

3) Hp:  $X, Y : \Omega \rightarrow \mathbb{R}$  sono v. a.,  $Y \in \mathcal{L}^1$ ,  $Y \geq X \geq 0$

Ts:  $X \in \mathcal{L}^1$ , (monotonia)  $\mathbf{E}(Y) \geq \mathbf{E}(X)$

4) Hp:  $X : \Omega \rightarrow \mathbb{R}$  è una v. a.,  $X \geq 0$ ,  $\mathbf{E}(X) = 0$

Ts:  $X = 0$  quasi certamente

5) Hp:  $X, Y : \Omega \rightarrow \mathbb{R}$  sono v. a. che ammettono valore atteso,  $X = Y$  q. c.

Ts:  $\mathbf{E}(X) = \mathbf{E}(Y)$

3) implica che una v. a. non negativa e limitata è integrabile, quindi per funzioni limitate nonnegative integrabilità e misurabilità sono equivalenti; con 7) si rimuoverà l'ipotesi di positività.

(4) non vale se si rimuove l'ipotesi  $X \geq 0$ : il "centro" può essere in 0 anche se  $X$  assume sia valori positivi che negativi con probabilità non nulla.

5) significa che, essendo  $\mathbf{P}(X = Y) = 1$ ,  $X, Y$  hanno lo stesso valore atteso, eventualmente infinito, perché differiscono su un insieme di probabilità nulla, per cui l'integrale "non lo vede". (5) implica anche che vale l'implicazione opposta in (4). Facendo la regola del valore atteso questa proprietà sarà ulteriormente generalizzata.

**Dim\*** (4) Per la disuguaglianza di Markov  $\mathbf{P}(|X| \geq a) \leq 0 \forall a > 0$ : essendo  $X \geq 0$ ,  $\mathbf{P}(X \geq a) = 0 \forall a > 0$ . La successione di eventi  $(X \geq \frac{1}{n}) = X^{-1}([\frac{1}{n}, +\infty))$  cresce a  $X^{-1}((0, +\infty))$ , per cui, per continuità della probabilità  $\mathbf{P}(X > 0) = \lim_{n \rightarrow +\infty} \mathbf{P}(X \geq \frac{1}{n}) = 0$ , cioè  $X = 0$  q. c., essendo  $X \geq 0$ . ■

6) Hp:  $X, Y : \Omega \rightarrow \mathbb{R}$  sono v. a.,  $X, Y \geq 0, \alpha, \beta \geq 0$

Ts:  $\mathbf{E}(\alpha X + \beta Y) = \alpha \mathbf{E}(X) + \beta \mathbf{E}(Y)$

7)  $X \in \mathcal{L}^1(\Omega, \mathcal{A}, \mathbf{P}) \iff |X| \in \mathcal{L}^1(\Omega, \mathcal{A}, \mathbf{P})$ , e in tal caso

$$|\mathbf{E}(X)| \leq \mathbf{E}(|X|)$$

6) è la 2), ma per l'altra categoria di v. a. per quali sappiamo che esiste il valore atteso; è possibile che qualche valore atteso sia  $+\infty$ . Si chiede  $X, Y \geq 0, \alpha, \beta \geq 0$  per evitare di avere a che fare con forme di indeterminazione, di modo che  $\mathbf{E}(\alpha X + \beta Y)$  sotto tali ipotesi può essere solo non negativo o  $+\infty$ . Comunque tutti i valori attesi certamente esistono perché  $X, Y$  sono non negative.

7) evidenzia la differenza tra l'integrale con cui è stato definito il valore atteso e l'integrale di Riemann: se  $f$  è Riemann integrabile,  $|f|$  è Riemann integrabile, ma non vale il viceversa. Vale una disuguaglianza analoga a  $|\mathbf{E}(X)| \leq \mathbf{E}(|X|)$  per l'integrale di Riemann.  $|X|$  è una variabile aleatoria perché composizione della funzione continua e quindi misurabile  $|x|$  con  $X$ .

**Dim\*** 7) Se  $X \in \mathcal{L}^1(\Omega, \mathcal{A}, \mathbf{P})$ , per definizione  $\mathbf{E}(X_+), \mathbf{E}(X_-) < +\infty$ : allora  $\mathbf{E}(|X|) = \mathbf{E}(X_+ + X_-) = \mathbf{E}(X_+) + \mathbf{E}(X_-) < +\infty$ . Se invece  $|X| \in \mathcal{L}^1(\Omega, \mathcal{A}, \mathbf{P})$ ,  $\mathbf{E}(X_+) + \mathbf{E}(X_-) < +\infty$ , ma dato che  $\mathbf{E}(X_+), \mathbf{E}(X_-)$  sono entrambi positivi non è possibile che almeno uno dei due sia infinito, cioè  $\mathbf{E}(X_+), \mathbf{E}(X_-) < +\infty$ . Allora  $X$  è integrabile.

$|\mathbf{E}(X)| = |\mathbf{E}(X_+) - \mathbf{E}(X_-)| \leq |\mathbf{E}(X_+)| + |\mathbf{E}(X_-)| = \mathbf{E}(X_+) + \mathbf{E}(X_-) = \mathbf{E}(|X|)$  per definizione, linearità, disuguaglianza triangolare e positività del valore atteso. ■

Si enunciano due teoremi che permettono di scambiare limite e valore atteso per successioni di v. a. r. nonnegative e integrabili, rispettivamente.

### **Teorema di Beppo Levi (convergenza monotona) 8.8 (8)**

Hp:  $(\Omega, \mathcal{A}, \mathbf{P})$  spazio di probabilità,  $(X_n)_{n \geq 1}$  è una successione di v.a.r.,

$X : \Omega \rightarrow \mathbb{R}$  è una v. a. tale che  $(X \geq )X_n \geq 0$  q. c.  $\forall n$  e  $X_n \nearrow X$  q. c.

Ts:  $\mathbf{E}(X) = \lim_{n \rightarrow +\infty} \mathbf{E}(X_n)$

$X_n \leq X$  superfluo.  $X_n \nearrow X$  significa che  $X_n$  converge puntualmente in modo monotono ( $X_n \leq X_{n+1} \forall n, \forall \omega$ ) a  $X$  quasi certamente, cioè che  $\exists A \in \mathcal{A} : \mathbf{P}(A) = 1$  e  $\lim_{n \rightarrow +\infty} X_n(\omega) = X(\omega) \forall \omega \in A$  in modo monotono. Questo insieme di  $\omega$  appartiene a  $\mathcal{A}$  perché  $\lim_{n \rightarrow +\infty} X_n =: Y$  è una variabile aleatoria, per cui  $\lim_{n \rightarrow +\infty} X_n(\omega) = X(\omega) \iff Y - X = 0$ .

Il teorema afferma che per una successione di v.a.r. non negative che cresca a  $X$  è possibile scambiare limite e integrale.

**Teorema di convergenza dominata 8.8 (9)**

Hp:  $(X_n)_{n \geq 1}$  è una successione di v.a.r.,  $Y \in \mathcal{L}^1(\Omega, \mathcal{A}, \mathbf{P})$

$\forall n \geq 1$  vale  $|X_n| \leq Y$  q. c. e  $X_n \rightarrow X$  q. c.

Ts:  $X_n \in \mathcal{L}^1(\Omega, \mathcal{A}, \mathbf{P}) \forall n \geq 1, X \in \mathcal{L}^1, \mathbf{E}(X) = \lim_{n \rightarrow +\infty} \mathbf{E}(X_n)$

Le ipotesi implicano  $Y \geq 0$ . Una v. a. certamente limitata è quindi in  $\mathcal{L}^1$ , perché se  $\exists c : |X(\omega)| \leq c \forall \omega \in \Omega$ ,  $Y(\omega) = c$  è integrabile (prendendo  $X_n = X \forall n$ ), allora  $|X| \in \mathcal{L}^1$  e  $X \in \mathcal{L}^1$ .

**Corollario 9.1 (ulteriori proprietà del valore atteso)**

Hp:  $(\Omega, \mathcal{A}, \mathbf{P})$  è uno spazio di probabilità,  $X : \Omega \rightarrow \mathbb{R}$  è una v.a.

Ts: (i) se  $\exists c \in \mathbb{R} : X = c$  q. c.,  $X \in \mathcal{L}^1(\Omega, \mathcal{A}, \mathbf{P})$  e  $\mathbf{E}(X) = c$

(ii) se  $\mathbf{E}(X) > 0, \mathbf{P}(X > 0) > 0$

(iii)  $\mathbf{E}(|X|) = \mathbf{E}(X_+) + \mathbf{E}(X_-)$

(iv) se  $X \in [a, b]$  q. c. con  $a, b \in \mathbb{R}, X \in \mathcal{L}^1(\Omega, \mathcal{A}, \mathbf{P})$  e  $\mathbf{E}(X) \in [a, b]$

(iv) implica che  $X$  limitata è integrabile (in generale in ogni  $L^p$ )

Non vale il viceversa né per (i) né per (ii) né per (iv). (iv) rafforza l'intuizione di  $\mathbf{E}(X)$  come "centro" dei valori assunti da  $X$ : se  $X \in [a, b]$  q. c., è ragionevole che anche il centro vi appartenga.

**Dim** (i) Se  $X = c$  q. c., allora  $|X| = |c|$  q. c. (posso prendere lo stesso  $A$ ). Allora, essendo  $|X|$  e  $|c|$  v. a. nonnegative, il loro valore atteso esiste, e per 8.8 (5) vale  $\mathbf{E}(|X|) = \mathbf{E}(|c|)$ . Ma  $|c|$  è una variabile aleatoria semplice, quindi  $\mathbf{E}(|c|) = |c| \cdot \mathbf{P}(\Omega) = |c|$ : allora  $\mathbf{E}(|X|) < +\infty$ , cioè  $|X| \in \mathcal{L}^1(\Omega, \mathcal{A}, \mathbf{P})$ , quindi per 8.8 (7) anche  $X \in \mathcal{L}^1(\Omega, \mathcal{A}, \mathbf{P})$ . Ora che so che esiste  $\mathbf{E}(X)$ , posso applicare 8.8 (5) a  $X, c$ , per cui  $\mathbf{E}(X) = c \cdot 1 = c$ .

(ii) Per definizione  $\mathbf{E}(X) = \mathbf{E}(X_+) - \mathbf{E}(X_-) > 0$ : allora  $\mathbf{E}(X_+) > \mathbf{E}(X_-) \geq 0$ , dunque  $\mathbf{E}(X_+) > 0$  e  $\mathbf{P}(X_+ > 0) > 0$  (se per assurdo fosse  $\mathbf{P}(X_+ > 0) = 0$ , si avrebbe  $\mathbf{P}(X_+ = 0) = 1$  perché  $X_+$  è non negativa, ma



questo significa che  $X_+ = 0$  q. c., perciò si avrebbe  $\mathbf{E}(X_+) = 0$ , che è assurdo), quindi  $\mathbf{P}(X > 0) > 0$  perché  $X^{-1}(0, +\infty) = X_+^{-1}(0, +\infty)$ .

(iii) E' noto che  $|X| = X_+ + X_-$ . Essendo entrambe v. a. r. nonnegative, per la linearità in 8.8 (6) si ha  $\mathbf{E}(|X|) = \mathbf{E}(X_+) + \mathbf{E}(X_-)$ .

(iv) Poiché  $X \in [a, b]$  q. c., si ha che  $|X| \leq \max\{|a|, |b|\}$ . Pongo  $c = \max\{|a|, |b|\}$ : per monotonia (8.8 (3)) si ha che  $|X| \in \mathcal{L}^1(\Omega, \mathcal{A}, \mathbf{P})$  e  $\mathbf{E}(|X|) \leq \mathbf{E}(\max\{|a|, |b|\}) = \max\{|a|, |b|\}$ . Allora per 8.8 (7)  $X \in \mathcal{L}^1(\Omega, \mathcal{A}, \mathbf{P})$  e per 8.8 (3) applicato al fatto che  $X - a \geq 0$ ,  $b - X \geq 0$ , si ha che  $\mathbf{E}(X - a) \geq 0$ ,  $\mathbf{E}(b - X) \geq 0$ , cioè, per linearità (8.8 (1)), si ha che  $\mathbf{E}(X) \in [a, b]$ . ■

La definizione di valore atteso ha una gran quantità di applicazioni.

### Teo (A1: valore atteso in spazi discreti)

Hp:  $\Omega$  discreto,  $(\Omega, 2^\Omega, \mathbf{P})$  è uno spazio di probabilità,  $\mathbf{P}$  ha densità

$p : \Omega \rightarrow [0, 1]$ ,  $X : \Omega \rightarrow [0, +\infty]$  è una v.a. o  $X \in \mathcal{L}^1(\Omega, 2^\Omega, \mathbf{P})$

$$\text{Ts: } \mathbf{E}(X) = \sum_{\omega \in \Omega} X(\omega) p(\omega)$$

Nelle ipotesi figurano le due tipologie di variabili aleatorie per le quali sappiamo per certo che esiste il valore atteso: quelle non negative e quelle integrabili. La somma nella tesi è effettivamente una somma se  $\Omega$  è finito, è una serie se  $\Omega$  è infinito numerabile. La tesi fornisce uno strumento per calcolare agevolmente il valore atteso.

**Dim** Considero il caso di  $\Omega$  finito,  $X : \Omega \rightarrow [0, +\infty]$ . Allora  $|ImX| = n \in \mathbb{N}$ ,  $ImX = \{x_1, \dots, x_n\}$ , e  $X$  è semplice, quindi si può scrivere  $X(\omega) = \sum_{k=1}^n x_k I_{A_k}(\omega) = \begin{cases} x_1 & \text{se } \omega \in A_1 \\ \dots & \dots \\ x_n & \text{se } \omega \in A_n \end{cases}$ , con  $\{A_1, \dots, A_n\}$  partizione discreta di  $\Omega$ .  $A_k = X^{-1}(\{x_k\}) = \{\omega : X(\omega) = x_k\}$ . Allora so scrivere  $\mathbf{E}(X) = \sum_{k=1}^n x_k \mathbf{P}(A_k)$ : poiché  $\Omega$  è discreto e anche  $A_k$  è discreto,  $A_k$  può essere scritto come unione al più numerabile di singoletti, per cui  $\mathbf{E}(X) = \sum_{k=1}^n x_k \mathbf{P}(A_k) = \sum_{k=1}^n x_k \sum_{\omega \in A_k} p(\omega) = \sum_{k=1}^n \sum_{\omega \in A_k} x_k p(\omega) = \sum_{k=1}^n \sum_{\omega \in A_k} X(\omega) p(\omega) = \sum_{\omega \in \bigcup_{k=1}^n A_k} X(\omega) p(\omega) = \sum_{\omega \in \Omega} X(\omega) p(\omega)$ . Il riordinamento non altera la somma perché le serie sono a termini positivi. Il caso di  $X \in \mathcal{L}^1$  si tratta analogamente.

Considero ora  $\Omega$  infinito numerabile: numero  $\Omega$  scrivendo  $\Omega = \{\omega_n : n \geq 1\}$ . Devo in qualche modo ricondurre al caso precedente: allora introduco  $\forall n \geq 1$   $X_n(\omega) = X(\omega) I_{\{\omega_1, \dots, \omega_n\}}(\omega) = \begin{cases} X(\omega_1) & \text{se } \omega = \omega_1 \\ \dots & \dots \\ X(\omega_n) & \text{se } \omega = \omega_n \\ 0 & \text{se } \omega \notin \{\omega_1, \dots, \omega_n\} \end{cases} : X_n \text{ è una va-}$

riabile aleatoria semplice, quindi per il punto precedente  $\mathbf{E}(X_n) = \sum_{\omega \in \{\omega_1, \dots, \omega_n\}} X(\omega) p(\omega) = \sum_{k=1}^n X(\omega_k) p(\omega_k)$ . Inoltre  $X_n \geq 0 \forall n$  perché  $X$  è nonnegativa per ipotesi e  $X \geq X_{n+1} \geq X_n \geq 0 \forall n, \forall \omega$ : quindi  $X_n \nearrow X$  e si può applicare il teorema di convergenza monotona, per cui  $\mathbf{E}(X) = \lim_{n \rightarrow +\infty} \mathbf{E}(X_n) = \lim_{n \rightarrow +\infty} \sum_{k=1}^n X(\omega_k) p(\omega_k)$ , che è per definizione  $\sum_{n=1}^{+\infty} X(\omega_n) p(\omega_n) = \sum_{\omega \in \Omega} X(\omega) p(\omega)$ . Se  $X \in \mathcal{L}^1$  la dimostrazione è analoga, ma termina con il teorema di convergenza dominata. ■

1. Considero lo spazio di probabilità  $(\mathbb{N}, 2^{\mathbb{N}}, \mathbf{P})$  con  $\mathbf{P}$  caratterizzata dalla densità di Poisson  $p(k) = e^{-\lambda} \frac{\lambda^k}{k!} \forall k \in \mathbb{N}$ , con  $\lambda > 1$ . Considero la v. a.  $X : \Omega \rightarrow \mathbb{R}, X(k) = k!$ .  $\mathbf{P}(X = +\infty) = 0 = 1 - \mathbf{P}(X \in \mathbb{N})$ .  $X$  è non negativa e  $\mathbf{P}$  è una densità discreta: per il teorema sopra,  $\mathbf{E}(X) = \sum_{\omega \in \Omega} X(\omega) p(\omega) = \sum_{k=0}^{+\infty} X(k) p(k) = \sum_{k=0}^{+\infty} e^{-\lambda} \lambda^k = +\infty$ : la serie diverge perché  $\lambda > 1$ , quindi  $\mathbf{E}(X) = +\infty$  nonostante  $X$  sia q. c. finita. Non è quindi vero che se  $\mathbf{P}(X = +\infty) = 0$ , allora, se  $\mathbf{E}(X)$  esiste,  $\mathbf{E}(X) < +\infty$ . Si è già dimostrato che invece vale il viceversa.

La seconda applicazione (A2) è la seguente. Dato  $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathbf{P})$  spazio di probabilità con  $\mathbf{P}$  discreta, avente densità  $p$  e supporto  $\mathcal{S} \subseteq \mathbb{R}$ , se  $X : \Omega \rightarrow [0, +\infty]$  o  $X \in \mathcal{L}^1$ , allora  $\mathbf{E}(X) = \sum_{\omega \in \mathcal{S}} X(\omega) p(\omega)$ .  $\mathbf{P}$  discreta rende quindi equivalente lavorare su  $\Omega$  continuo o discreto.

Questo implica che se ci si trova in uno dei due casi precedenti (A1 o A2), poiché  $X \in \mathcal{L}^1(\Omega, \mathcal{A}, \mathbf{P}) \iff |X| \in \mathcal{L}^1(\Omega, \mathcal{A}, \mathbf{P})$ ,  $\mathbf{E}(X)$  esiste finito  $\iff \mathbf{E}(|X|) = \sum_{\omega \in \Omega/\mathcal{S}} |X(\omega)| p(\omega)$  è finito.

**Def 9.3** Dato  $(\Omega, \mathcal{A}, \mathbf{P})$  spazio di probabilità, dato  $A \in \mathcal{A}$ , si definisce valore atteso di  $X$  ristretta ad  $A$ , e si indica con  $\int_A X d\mathbf{P}$  o  $\int_{\Omega} X I_A d\mathbf{P}$ , il valore atteso di  $X I_A$   $\mathbf{E}(X I_A)$ , qualora esso esista.

Esiste certamente se  $X I_A$  è una v. a. non negativa (sappiamo che è una v.a. per 6.6 (5)), oppure se  $X I_A \in \mathcal{L}^1(\Omega, \mathcal{A}, \mathbf{P})$ .

Si noti che, essendo  $I_A$  una v. a. semplice,  $\mathbf{E}(I_A) = \mathbf{P}(A) = \int_A 1 d\mathbf{P}$ : l'integrazione della v. a. costante 1 rispetto a una misura di probabilità restituisce la misura dell'insieme su cui si integra, cioè  $\mathbf{P}(A)$ .

**Teo 9.4 (linearità del valore atteso rispetto a somme infinite)**

(1) Hp:  $(\Omega, \mathcal{A}, \mathbf{P})$  è uno spazio di probabilità,  $(X_n)_{n \geq 1}$  è una

successione di v.a. tali che  $X_n : \Omega \rightarrow [0, +\infty] \forall n \geq 1$

$$\text{Ts: } \mathbf{E} \left( \sum_{n=1}^{+\infty} X_n \right) = \sum_{n=1}^{+\infty} \mathbf{E}(X_n)$$

(2) Hp:  $(\Omega, \mathcal{A}, \mathbf{P})$  è uno spazio di probabilità,  $(X_n)_{n \geq 1}$  è una successione

di v.a. tali che  $X_n : \Omega \rightarrow \mathbb{R} \forall n \geq 1$ ,  $\sum_{n=1}^{+\infty} \mathbf{E}(|X_n|) < +\infty$

Ts:  $\sum_{n=1}^{+\infty} X_n(\omega) < +\infty$  q. c.,  $\sum_{n=1}^{+\infty} X_n(\omega) \in^1 (\Omega, \mathcal{A}, \mathbf{P})$  e  $\mathbf{E}\left(\sum_{n=1}^{+\infty} X_n\right) = \sum_{n=1}^{+\infty} \mathbf{E}(X_n)$

Acquisirà senso quando faremo la media campionaria.

**Dim** Dimostro 1. L'idea è applicare la convergenza monotona alla successione delle somme parziali dei valori assoluti. Chiamo  $T_n = \sum_{k=1}^n |X_k|$ . Se  $T = \sum_{k=1}^{+\infty} |X_k|$ , allora  $T_n \nearrow T$ , per definizione di  $\sum_{k=1}^{+\infty}$  e perché la serie è a termini positivi ( $T_{n+1} = \sum_{k=1}^{n+1} |X_k| \geq T_n \geq 0 \forall n$ ). Allora, per il teorema di convergenza monotona e per linearità,  $\mathbf{E}(T) = \lim_{n \rightarrow +\infty} \mathbf{E}(T_n) = \sum_{n=1}^{+\infty} \mathbf{E}(|X_n|)$ : il lato destro è ben definito, quindi  $T$  ammette valore atteso, ma essendo  $X_n \geq 0 \forall n$  vale  $T = \sum_{k=1}^{+\infty} X_k$ , per cui  $\mathbf{E}\left(\sum_{k=1}^{+\infty} X_k\right) = \sum_{k=1}^{+\infty} \mathbf{E}(X_k)$ . Così si è dimostrata (1).

Dimostro 2. Applicando (1),  $\mathbf{E}\left(\sum_{n=1}^{+\infty} |X_n|\right) = \sum_{n=1}^{+\infty} \mathbf{E}(|X_n|) < +\infty$ , quindi  $\sum_{n=1}^{+\infty} |X_n| < +\infty$  q. c. (si è dimostrato definendo il valore atteso che se  $\mathbf{P}(X = +\infty) > 0$ , allora  $\mathbf{E}(X) = +\infty$ ). Per il criterio della convergenza assoluta anche  $\sum_{n=1}^{+\infty} X_n < +\infty$  q. c. Così si è dimostrata (2).1. Chiamo  $S_n = \sum_{k=1}^n X_k$  la successione delle somme parziali delle  $X_k$ .

Per le osservazioni sopra, vale  $\forall n \ 0 \leq |S_n| \leq T_n \leq T$ . Poiché per ipotesi  $T \in^1 (\Omega, \mathcal{A}, \mathbf{P})$  e  $S_n \rightarrow S = \sum_{k=1}^{+\infty} X_k$ , allora, per convergenza dominata,  $S_n, S \in^1 (\Omega, \mathcal{A}, \mathbf{P})$  ((2).2) e  $\mathbf{E}(S) = \lim_{n \rightarrow +\infty} \mathbf{E}(\sum_{k=1}^n X_k) = \lim_{n \rightarrow +\infty} \sum_{k=1}^n \mathbf{E}(X_k) = \sum_{k=1}^{+\infty} \mathbf{E}(X_k)$  ((2).3). Uso la linearità di  $\mathbf{E}$  perché le  $X_k$  sono integrabili per convergenza dominata.

[Ripetendo le stesse osservazioni sopra con le ipotesi di (2), si ha che  $\mathbf{E}(T) = \sum_{n=1}^{+\infty} \mathbf{E}(|X_n|) < +\infty$ , per cui  $T = \sum_{k=1}^{+\infty} |X_k|$  è integrabile. Ma  $0 \leq \left|\sum_{k=1}^{+\infty} X_k\right| \leq \sum_{k=1}^{+\infty} |X_k|$ , quindi per monotonia (8.8 (3)) anche  $\left|\sum_{k=1}^{+\infty} X_k\right| \in^1 (\Omega, \mathcal{A}, \mathbf{P})$  e di conseguenza  $\sum_{k=1}^{+\infty} X_k \in^1 (\Omega, \mathcal{A}, \mathbf{P})$ . ] ■

### Teo 9.5 (disuguaglianza di Markov)

Hp:  $(\Omega, \mathcal{A}, \mathbf{P})$  è uno spazio di probabilità,  $X : \Omega \rightarrow \mathbb{R}$  è una v.a.

Ts:  $\forall a > 0 \ \mathbf{P}(|X| \geq a) \leq \frac{\mathbf{E}(|X|)}{a}$

La tesi illustra il significato del valore atteso come centro dei valori assunti da  $X$ .  $\mathbf{E}(|X|)$  si può calcolare perché  $|X|$  è non negativa. La tesi afferma che una v. a. integrabile prende la "maggior parte dei suoi valori" in un insieme compatto. Se  $a \leq \mathbf{E}(|X|)$ , la disuguaglianza è inutile, perché potrebbe essere  $|X| = \mathbf{E}(|X|)$  q. c., nel qual caso  $\mathbf{P}(|X| \geq a)$  sarebbe 1.

**Dim** Se  $X \notin \mathcal{L}^1$ , allora  $|X| \notin \mathcal{L}^1$  e  $\mathbf{E}(|X|) = +\infty$ . In tal caso  $\mathbf{P}(|X| \geq a) \leq 1 \leq +\infty$ , quindi vale la disuguaglianza. Suppongo ora  $X \in \mathcal{L}^1$ , quindi  $|X| \in \mathcal{L}^1$ . L'idea è scrivere il lato sinistro della disuguaglianza come valore atteso di qualche v. a. e sfruttare la monotonia. Noto che  $\omega \in (|X| \geq a) \iff I_{(|X| \geq a)}(\omega) = 1$ , e vale la disuguaglianza  $I_{(|X| \geq a)}(\omega) \leq \frac{|X(\omega)|}{a} I_{(|X| \geq a)}(\omega)$  (si ha  $0 \leq 0$ , o  $1 \leq \frac{|X(\omega)|}{a}$  quando  $|X(\omega)| \geq a$ ). Essendo  $(|X| \geq a) \in \mathcal{A}$  perché  $|X|$  è una v. a., si ha che entrambi i lati della disuguaglianza sono v. a. Vorrei dire che il lato destro è una v. a. in  $\mathcal{L}^1$  per poter usare la monotonia: ma  $\frac{|X(\omega)|}{a} I_{(|X| \geq a)}(\omega) \leq \frac{|X(\omega)|}{a} I_{\Omega}(\omega) = \frac{|X(\omega)|}{a}$ , che è una v. a. non negativa in  $\mathcal{L}^1$ , quindi  $\frac{|X(\omega)|}{a} I_{(|X| \geq a)}(\omega) \in \mathcal{L}^1$  e per 8.8 (3) vale  $\mathbf{E}(I_{(|X| \geq a)}(\omega)) \leq \mathbf{E}\left(\frac{|X(\omega)|}{a}\right) = \frac{1}{a} \mathbf{E}(|X(\omega)|)$ . Ma  $\mathbf{E}(I_{(|X| \geq a)}(\omega)) = 1 \cdot \mathbf{P}(|X| \geq a)$  per la definizione di valore atteso per le v. a. semplici, quindi si ha la tesi. ■

(Assolutamente inutile:  $\mathbf{E}(I_{(|X| \geq a)}(\omega)) = \mathbf{P}(|X| \geq a)$  si vede con la definizione di valore atteso per v. a. non negative. Fissato  $n$ ,  $X_1(\omega) = \begin{cases} 0 & \text{se } I_{(|X| \geq a)}(\omega) = 0 \\ \frac{1}{2} & \text{se } I_{(|X| \geq a)}(\omega) = 1 \end{cases} = \begin{cases} 0 & \text{se } |X(\omega)| < a \\ \frac{1}{2} & \text{se } |X(\omega)| \geq a \end{cases}, \dots, X_n(\omega) = \begin{cases} 0 & \text{se } I_{(|X| \geq a)}(\omega) = 0 \\ \frac{2^n - 1}{2^n} & \text{se } I_{(|X| \geq a)}(\omega) = 1 \end{cases}$  (perché, per come sono descritti gli intervallini, quello in cui è previsto che la funzione assuma valore 1 è quello con  $k = 2^n - 1$ ).  $\mathbf{E}(X_n) = 0 \cdot \mathbf{P}(|X| < a) + \frac{2^n - 1}{2^n} \mathbf{P}(|X| \geq a) = (1 - \frac{1}{2^n}) \mathbf{P}(|X| \geq a)$ . Quindi  $\mathbf{E}(I_{(|X| \geq a)}(\omega)) = \lim_{n \rightarrow +\infty} (1 - \frac{1}{2^n}) \mathbf{P}(|X| \geq a) = \mathbf{P}(|X| \geq a)$ .)

### 6.2.1 Spazi $L^p$ e varianza

Si è già dimostrato che la relazione di uguaglianza quasi certa tra due v. a. è una relazione di equivalenza: si può scrivere in tal caso  $X \mathcal{R} Y$ . Se si considerano  $X, Y \in \mathcal{L}^1(\Omega, \mathcal{A}, \mathbf{P})$ , tale relazione di equivalenza induce una partizione di  $\mathcal{L}^1$  in classi di equivalenza; la classe di equivalenza di  $X \in \mathcal{L}^1$  si indica con  $[X]$ :  $[X] = \{Y \in \mathcal{L}^1(\Omega, \mathcal{A}, \mathbf{P}) : Y = X \text{ q. c.}\}$ . Se  $X = Y$  q. c.,  $P^X = P^Y$  e  $\mathbf{E}(X) = \mathbf{E}(Y)$ , quindi il valore atteso e la legge sono gli stessi per ogni v. a. nella stessa classe di equivalenza.

E' quindi ben definito l'insieme quoziente di  $\mathcal{L}^1(\Omega, \mathcal{A}, \mathbf{P})$  rispetto alla relazione  $\mathcal{R}$ , indicato con  $\mathcal{L}^1/\mathcal{R}$ : è l'insieme delle classi di equivalenza  $[X]$  indotte da  $\mathcal{R}$  in  $\mathcal{L}^1(\Omega, \mathcal{A}, \mathbf{P})$ , al variare di  $X \in \mathcal{L}^1$ , e si indica con  $L^1(\Omega, \mathcal{A}, \mathbf{P})$ .

$L^1$  è uno spazio vettoriale su  $\mathbb{R}$  con la somma  $[X_1] + [X_2] = [X_1 + X_2]$ :  $+: L^1 \times L^1 \rightarrow L^1$  è la funzione che, date due classi di equivalenza  $[X], [Y]$ , prende un qualsiasi  $X_1 \in [X]$ , un qualsiasi  $Y_1 \in [Y]$  e associa a esse la classe di equivalenza cui appartiene  $X_1 + Y_1$  (che appartiene a  $L^1$  per linearità del valore atteso). Questa non dipende dalla scelta di  $X_1$  e  $Y_1$  perché se  $X_1 = X_2$  q. c. (su  $A$ ) e  $Y_1 = Y_2$  q. c. (su  $B$ ),  $\alpha X_1 + \beta Y_1 = \alpha X_2 + \beta Y_2$  q. c. (su  $A \cap B$ :  $\mathbf{P}(A \cap B) = \mathbf{P}(A) + \mathbf{P}(B) - \mathbf{P}(A \cup B) = 2 - \mathbf{P}(A \cup B) = 1$ ). L'elemento neutro è  $[0]$ .  $\alpha[X_1] = [\alpha X_1]$ .

Noi, commettendo un abuso di notazione, scriveremo  $L^1$  invece di  $\mathcal{L}^1$  e  $X \in L^1$ , invece che  $[X] \in L^1$ .

**Def 9.6** Dato  $(\Omega, \mathcal{A}, \mathbf{P})$  spazio di probabilità e  $p \in [1, +\infty)$ , si indica con  $\mathcal{L}^p(\Omega, \mathcal{A}, \mathbf{P})$  lo spazio delle variabili aleatorie reali con  $p$ -esima potenza assoluta sommabile, cioè  $\mathcal{L}^p(\Omega, \mathcal{A}, \mathbf{P}) = \{\text{variabili aleatorie } X : \Omega \rightarrow \mathbb{R} : |X|^p \text{ ammette valore a}$

Il valore assoluto serve a far sì che  $|X|^p$  sia ben definito  $\forall p \geq 1$ . In tal caso  $\mathbf{E}(|X|^p) < +\infty$  è detto momento assoluto  $p$ -esimo associato a  $X$  (vedi es). Se  $X \in \mathcal{L}^p(\Omega, \mathcal{A}, \mathbf{P})$ , allora  $\mathbf{E}(|X|^p) < +\infty$ , cioè  $|X|^p \in \mathcal{L}^1(\Omega, \mathcal{A}, \mathbf{P})$ . In generale, l'insieme delle v. a. semplici è denso nell'insieme  $L^p$ .

**Def 9.7** Dato  $(\Omega, \mathcal{A}, \mathbf{P})$  spazio di probabilità e  $\mathcal{L}^p(\Omega, \mathcal{A}, \mathbf{P})$ , si indica con  $L^p$  l'insieme delle classi di equivalenza  $[X]$  indotte da  $\mathcal{R}$  in  $\mathcal{L}^p(\Omega, 2^\Omega, \mathbf{P})$ , al variare di  $X \in \mathcal{L}^p$ .

Se  $X = c$  q. c., allora  $X \in L^p \forall p \in [1, +\infty)$  e  $\mathbf{E}(|X|^p) = |c|^p$  (si estende banalmente 9.1 (8)).

Si vedrà che  $\forall p$   $L^p(\Omega, \mathcal{A}, \mathbf{P})$  è uno spazio vettoriale (nel seguito si dimostra che  $L^2$  lo è), che diventa normato con la norma  $(\mathbf{E}(|X|^p))^{\frac{1}{p}}$  e anche completo, dunque di Banach. La norma induce quindi la metrica  $d(X, Y) = (\mathbf{E}(|X - Y|^p))^{\frac{1}{p}}$ ;  $L^1$  con tale distanza è uno spazio metrico completo.

Se  $p = 2$  si può definire il prodotto scalare  $\langle \cdot, \cdot \rangle : L^2(\Omega, \mathcal{A}, \mathbf{P}) \times L^2(\Omega, \mathcal{A}, \mathbf{P}) \rightarrow \mathbb{R}$ ,  $\langle X, Y \rangle = \mathbf{E}(XY) = \int_{\Omega} XY d\mathbf{P}$ , in analogia a  $\langle f, g \rangle = \int_a^b f(x)g(x)dx$  (ma stavolta si integra rispetto a una misura). E' ben definito perché, come si vedrà ora, il prodotto di v. a. in  $L^2$  è in  $L^1$ . Effettivamente è positivo, bilineare e commutativo per le proprietà del valore atteso, e  $\langle X, X \rangle = 0 \iff \mathbf{E}(X^2) = 0 \iff [X] = [0]$ . Allora tale prodotto scalare induce la norma  $\sqrt{\mathbf{E}(|X|^2)}$ , e  $L^2$  è completo rispetto a tale norma: questo rende  $L^2(\Omega, \mathcal{A}, \mathbf{P})$ , l'unico tra tutti gli  $L^p$ , uno spazio di Hilbert oltre che di Banach.

**Teo 9.8 (proprietà degli spazi  $L^p$ )**

Hp:  $(\Omega, \mathcal{A}, \mathbf{P})$  spazio di probabilità

Ts: (1) se  $1 \leq p < q$ ,  $L^q \subseteq L^p \subseteq L^1$

(2)  $L^2$  è uno spazio vettoriale

(3) (disuguaglianza di Cauchy-Schwarz) se  $X, Y \in L^2$ ,

$$XY \in L^1 \text{ e } |\mathbf{E}(XY)| \leq \sqrt{\mathbf{E}(X^2)\mathbf{E}(Y^2)}$$

(1) vale sostanzialmente perché  $\mathbf{P}$  è una misura finita, cioè  $\mathbf{P}(\Omega) < +\infty$ . (3) vale perché  $\mathbf{E}(XY)$  è un prodotto scalare in  $L^2(\Omega, \mathcal{A}, \mathbf{P})$ , quindi vale la disuguaglianza solita  $|\langle X, Y \rangle| \leq \|X\| \|Y\|$ .

**Dim** (1) Vediamo che  $L^2 \subseteq L^1$ . Se  $X \in L^2$ , allora  $\mathbf{E}(X^2) < +\infty$ : mostro che  $\mathbf{E}(X) < +\infty$  usando il fatto che  $\mathbf{E}(|X|) < +\infty$ , con la monotonia applicata a un'opportuna disuguaglianza. Infatti  $|X(\omega)| < X^2(\omega) + 1 \forall \omega \in \Omega$ : essendo  $X^2 + 1 \in L^1$  per linearità ( $\mathbf{E}(X^2 + 1) = \mathbf{E}(X^2) + 1$ ), per 8.8 (3) anche  $|X| \in L^1$ , dato che  $\mathbf{E}(|X|) \leq \mathbf{E}(X^2 + 1) < +\infty$ . Si usa  $\mathbf{P}(\Omega) < +\infty$  per affermare che  $\mathbf{E}(1) = 1$ : infatti, essendo 1 una v. a. semplice,  $\mathbf{E}(1) = 1 \cdot \mathbf{P}(\Omega) = \mathbf{P}(\Omega) = 1$ .

(3) E' una conseguenza dello struttura di spazio euclideo; in ogni caso, si ridimostra. E' noto che  $|ab| \leq \frac{1}{2}a^2 + \frac{1}{2}b^2$ : quindi  $|XY| \leq \frac{1}{2}X^2 + \frac{1}{2}Y^2$ . Entrambi i lati sono nonnegativi e  $\frac{1}{2}X^2 + \frac{1}{2}Y^2 \in L^1$  perché  $L^1$  è uno spazio vettoriale, quindi per monotonia (8.8 (3)) anche  $|XY| \in L^1$  e  $XY \in L^1$ .

La seconda parte è una conseguenza dello struttura di spazio euclideo; in ogni caso, si ridimostra.  $\mathbf{E}((aX + Y)^2) = a^2\mathbf{E}(X^2) + \mathbf{E}(Y^2) + 2a\mathbf{E}(XY) \geq 0$  per positività del valore atteso, quindi, vedendo la disuguaglianza nell'incognita  $a$ , il  $\Delta$  dev'essere non positivo, per cui  $4\mathbf{E}^2(X, Y) - 4\mathbf{E}(X^2)\mathbf{E}(Y^2) \leq 0$ , cioè  $|\mathbf{E}(XY)| \leq \sqrt{\mathbf{E}(X^2)\mathbf{E}(Y^2)}$ .

(2) Date  $X, Y \in L^2$ ,  $X + Y \in L^2$  perché  $\mathbf{E}((X + Y)^2) = \mathbf{E}(X^2 + Y^2 + 2XY)$ , ma la funzione integranda è in  $L^1$  perché somma di v. a.  $L^1$ , per ipotesi e per 3, quindi anche  $\mathbf{E}((X + Y)^2)$  è un numero reale. ■

**Def 9.9** Dato  $(\Omega, \mathcal{A}, \mathbf{P})$  spazio di probabilità, data  $X \in L^2(\Omega, \mathcal{A}, \mathbf{P})$ , si definisce varianza di  $X$ , e si indica con  $var(X)$  o  $\sigma_X^2$ ,  $\mathbf{E}((X - \mathbf{E}(X))^2)$ .

$\sigma^2$  misura quanto i valori assunti da  $X$  sono dispersi attorno a  $\mathbf{E}(X)$ . La definizione è ben posta, perché  $L^2 \subseteq L^1$  e si può calcolare  $\mathbf{E}(X)$ , e perché  $X - \mathbf{E}(X) \in L^2$ , essendo  $L^2$  uno spazio vettoriale, per cui  $\sigma^2 \in \mathbb{R}$ .

**Prop 9.10 (proprietà della varianza)**

Hp:  $(\Omega, \mathcal{A}, \mathbf{P})$  spazio di probabilità,  $X \in L^2(\Omega, \mathcal{A}, \mathbf{P})$

Ts: (1)  $var(X) \geq 0$

(2)  $var(X) = 0 \iff \exists c \in \mathbb{R} : X = c \text{ q. c. } \iff \exists c : P^X \sim \delta_c$

(3) (disuguaglianza di Chebyshev)  $\mathbf{P}(|X - \mathbf{E}(X)| \geq a) \leq \frac{\sigma^2}{a^2} \forall a > 0$

(4)  $var(X) = \mathbf{E}(X^2) - \mathbf{E}^2(X)$

(5\*) se  $Y = aX + b$ ,  $var(Y) = a^2 var(X)$

(1) è intuitivamente ovvia per il fatto che un indice di dispersione con valore negativo non ha significato. (2) intuitivamente è dovuta al fatto che dispersione nulla di  $X$  significa che  $X$  assume lo stesso valore quasi ovunque. (3) illustra il significato di  $\sigma^2$  come indice di dispersione, come fa la disuguaglianza di Markov per il valore atteso: la varianza controlla la probabilità che  $X$  si allontani dal suo valore atteso.

**Dim** (1) Per definizione  $var(X) = \mathbf{E}((X - \mathbf{E}(X))^2)$ . Essendo  $(X - \mathbf{E}(X))^2 \geq 0$ , per 8.8 (2) (ii) vale  $\mathbf{E}((X - \mathbf{E}(X))^2) \geq 0$ .

(2) La seconda e la terza proposizione sono chiaramente equivalenti perché  $\exists c : P^X \sim \delta_c$  se e solo se la legge di  $X$   $P^X : \mathcal{B}(\mathbb{R}) \rightarrow [0, 1]$  è tale che  $P^X(A) = \mathbf{P}(X^{-1}(A)) = \begin{cases} 1 & \text{se } c \in A \\ 0 & \text{se } c \notin A \end{cases} \forall A \in \mathcal{B}(\mathbb{R})$ . Se quindi scelgo  $A = \{c\}$ ,  $\mathbf{P}(X = c) = 1$ , vale a dire che  $X = c$  q. c., e se  $X = c$  q. c. la legge è necessariamente quella sopra.

Dimostro che la seconda proposizione implica la prima: se  $X = c$  q. c.,  $\mathbf{E}(X) = c$ , per cui  $X - \mathbf{E}(X) = 0$  q. c. e  $(X - \mathbf{E}(X))^2 = 0$  q. c., per cui  $\mathbf{E}((X - \mathbf{E}(X))^2) = 0 = \text{var}(X)$ .

Inoltre la prima proposizione implica la seconda perché, vedendo la varianza come norma al quadrato, si ha  $\text{var}(X) = \mathbf{E}((X - \mathbf{E}(X))(X - \mathbf{E}(X))) = \langle X - \mathbf{E}(X), X - \mathbf{E}(X) \rangle = \|X - \mathbf{E}(X)\|^2 = 0$  se e solo se  $X = \mathbf{E}(X)$  q. c., cioè  $X \in [\mathbf{E}(X)]$ .

(3) Per la disuguaglianza di Markov applicata a  $Y = (X - \mathbf{E}(X))^2 = |Y|$  con valore di soglia  $a^2$ ,  $\mathbf{P}(|X - \mathbf{E}(X)| \geq a) = \mathbf{P}(|X - \mathbf{E}(X)|^2 \geq a^2) \leq \frac{\mathbf{E}(|X - \mathbf{E}(X)|^2)}{a^2} = \frac{\sigma^2}{a^2}$ .

[Devo dimostrare che  $\mathbf{P}(|X - \mathbf{E}(X)| \geq a) \leq \frac{\mathbf{E}((X - \mathbf{E}(X))^2)}{a^2}$ . Noto che  $|X - \mathbf{E}(X)| \geq a \iff (X - \mathbf{E}(X))^2 \geq a^2$ : questo significa che  $I_{(|X - \mathbf{E}(X)| \geq a)}(\omega) \leq \frac{(X - \mathbf{E}(X))^2}{a^2}$ , perché quando  $|X - \mathbf{E}(X)| < a$  vale ovviamente  $\frac{(X - \mathbf{E}(X))^2}{a^2} \geq 0$ , mentre quando  $|X - \mathbf{E}(X)| \geq a$  vale anche  $(X - \mathbf{E}(X))^2 \geq a^2$ , cioè  $\frac{(X - \mathbf{E}(X))^2}{a^2} \geq 1$ , per cui la disuguaglianza è verificata. Allora  $\mathbf{E}(I_{(|X - \mathbf{E}(X)| \geq a)}(\omega)) \leq \mathbf{E}\left(\frac{(X - \mathbf{E}(X))^2}{a^2}\right)$ , cioè  $\mathbf{P}(|X - \mathbf{E}(X)| \geq a) \leq \frac{\mathbf{E}((X - \mathbf{E}(X))^2)}{a^2}$ .]

(4) Per definizione  $\text{Var}(X) = \mathbf{E}((X - \mathbf{E}(X))^2) = \mathbf{E}(X^2 + \mathbf{E}^2(X) - 2X\mathbf{E}(X))$ , che è, per linearità e per idempotenza,  $\mathbf{E}(X^2) + \mathbf{E}^2(X) - 2\mathbf{E}(X)\mathbf{E}(X) = \mathbf{E}(X^2) - \mathbf{E}^2(X)$ .

(5)  $\text{var}(aX + b) = \mathbf{E}((aX + b - \mathbf{E}(aX + b))^2) = \mathbf{E}((aX - a\mathbf{E}(X))^2) = a^2\mathbf{E}((X - \mathbf{E}(X))^2) = a^2\text{var}(X)$ .

■

## 6.2.2 Calcolo del valore atteso di una variabile aleatoria

Sappiamo che dati gli spazi  $(\Omega, \mathcal{A}, \mathbf{P})$ ,  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$  e  $X : \Omega \rightarrow \mathbb{R}$  con legge  $P^X : \mathcal{B}(\mathbb{R}) \rightarrow [0, 1]$ ,  $P^X(B) = \mathbf{P}(X^{-1}(B))$ , allora  $P^X$  è una misura di probabilità su  $\mathcal{B}(\mathbb{R})$ , cioè  $(\mathbb{R}, \mathcal{B}(\mathbb{R}), P^X)$  è uno spazio di probabilità. Posso vedere la funzione  $\text{Id} : \mathbb{R} \rightarrow \mathbb{R}$  come una nuova variabile aleatoria, definita su  $(\mathbb{R}, \mathcal{B}(\mathbb{R}), P^X)$ , per cui se ne può calcolare il valore atteso secondo  $P^X$   $E^X(\text{Id})$ .

### Teo 10.1 (regola del valore atteso)

Hp:  $(\Omega, \mathcal{A}, \mathbf{P})$  spazio di probabilità,  $(E, \mathcal{E})$  spazio misurabile,

$X : \Omega \rightarrow E$  è una v. a.,  $h : E \rightarrow \mathbb{R}$  è una funzione misurabile

Ts: (1)  $h(X) \in L^1(\Omega, \mathcal{A}, \mathbf{P}) \iff h \in L^1(E, \mathcal{E}, P^X)$

(2) se  $h : E \rightarrow [0, +\infty]$  o  $h \in L^1(E, \mathcal{E}, P^X)$ ,  $\mathbf{E}(h(X)) = E^X(h)$

$h : E \rightarrow [0, +\infty]$  o  $h \in L^1(E, \mathcal{E}, P^X)$  fa sì che  $E^X(h)$  sia ben definito.  $h(X)$  è non negativa se  $h$  è non negativa, oppure è integrabile per (1) se  $h$  è integrabile, quindi comunque  $\mathbf{E}(h(X))$  è ben definito. (2) significa

$$\mathbf{E}(h(X)) = \int_{\Omega} h(X(\omega)) d\mathbf{P}(\omega) = \int_E h(x) dP^X(x) = E^X(h)$$

E' quindi una formula di cambio di variabile per gli integrali rispetto a una misura di probabilità (analogamente al teorema di sostituzione per gli integrali di Riemann): la sostituzione è  $X(\omega) = x$ , per cui  $\omega = X^{-1}(x)$  e  $d\mathbf{P}(\omega) = d\mathbf{P}(X = x) = dP^X(x)$ . Si può quindi calcolare il valore atteso di una v.a. con due diversi integrali, rispetto a due diverse misure di probabilità. La formula rende inoltre evidente che  $\mathbf{E}(h(X))$  non dipende dai valori che assume  $X$  al variare di  $\omega$ , ma solo da  $h$  e dalla legge di  $X$ : due v. a. con la stessa legge hanno lo stesso valore atteso, anche se sono definite su spazi di probabilità diversi. Cioè, in generale, se  $X : \Omega_1 \rightarrow E, Y : \Omega_2 \rightarrow E$  sono tali che  $P^X = P^Y$ , allora  $\mathbf{E}(X) = E^X(Id) = E^Y(Id) = \mathbf{E}(Y)$ : questa è una generalizzazione di 8.8 (5), perché se  $X = Y$  q. c. allora  $P^X = P^Y$ .

Un modo naturale di applicare il teorema, se  $X : \Omega \rightarrow \mathbb{R}$ , è prendere  $h : \mathbb{R} \rightarrow \mathbb{R}$ ,  $h(x) = Id = x$ , per cui si può calcolare  $\mathbf{E}(X)$  integrando  $x$  rispetto alla legge di  $X$ :

$$\mathbf{E}(X) = \int_{\Omega} X(\omega) d\mathbf{P}(\omega) = \int_{\mathbb{R}} x dP^X(x) = E^X(Id)$$

**Dim** (Passo 1) Dato  $B \in \mathcal{E}$ , definisco  $h : E \rightarrow \mathbb{R}$ ,  $h(x) = I_B(x)$ , che è ovviamente misurabile.  $Imh = \{0, 1\}$ . Essendo  $I_B(X(\omega))$  una v. a. semplice,  $\mathbf{E}(I_B(X)) = \int_{\Omega} I_{(X \in B)}(\omega) d\mathbf{P}(\omega) = \mathbf{P}(X \in B) = P^X(B) = \int_E I_B(x) dP^X(x) = E^X(I_B)$ . Si è dimostrata (2) per  $h(x) = I_B(x)$ , con  $h$  non negativa.

(Passo 2) Considero  $h : E \rightarrow \mathbb{R}$  v. a. semplice, per cui  $h(x) = \sum_{k=1}^n h_k I_{B_k}(x)$ , dove  $\{B_1, \dots, B_n\}$  è una partizione di  $E$ ,  $h_1, \dots, h_n \in \mathbb{R}$ . Allora anche  $h(X)$  è una v. a. semplice e per definizione  $\mathbf{E}(h(X)) = \sum_{k=1}^n h_k \mathbf{P}(X \in B_k) = \sum_{k=1}^n h_k P^X(B_k) = E^X(h)$  (si mostra direttamente con il passo 1 e la linearità di  $\mathbf{E}$ ).

(Passo 3) Considero  $h : E \rightarrow [0, +\infty]$ : so che esiste una successione  $(h_n)_{n \geq 1}$  di v. a. semplici non negative che cresce a  $h$ , cioè  $h_n \nearrow h$  ( $\lim_{n \rightarrow +\infty} h_n(x) = h(x) \forall x$ ,  $0 \leq h_n \leq h_{n+1} \leq h \forall x$ ). Per ogni  $n$ , per il passo 2, vale  $\mathbf{E}(h_n(X)) = E^X(h_n)$ . Ho una successione di v. a. non negative che cresce a  $h$ : allora anche  $h_n(X(\omega))$  cresce a  $h(X(\omega)) \forall \omega$  e per il teorema di convergenza monotona su  $(\Omega, \mathcal{A}, \mathbf{P})$   $\mathbf{E}(h(X)) = \lim_{n \rightarrow +\infty} \mathbf{E}(h_n(X)) = \lim_{n \rightarrow +\infty} E^X(h_n)$ . Sempre per convergenza monotona applicata a  $h_n \nearrow h$  in  $(E, \mathcal{E}, P^X)$ , vale  $E^X(h) = \lim_{n \rightarrow +\infty} E^X(h_n)$ , per cui  $\mathbf{E}(h(X)) = E^X(h)$ . Così ho dimostrato che per  $h : E \rightarrow [0, +\infty]$  vale  $\mathbf{E}(h(X)) = E^X(h)$ .

E' noto che  $h(X) \in L^1(\Omega, \mathcal{A}, \mathbf{P}) \iff |h(X)| \in L^1(\Omega, \mathcal{A}, \mathbf{P})$ : ma  $|h(X)|$  è una v. a. non negativa, quindi - essendo per il passo 3  $\mathbf{E}(|h(X)|) = E^X(|h|) - \mathbf{E}(h(X)) \in L^1(\Omega, \mathcal{A}, \mathbf{P}) \iff |h| \in L^1(E, \mathcal{E}, P^X)$ , e di nuovo vale  $|h| \in L^1(E, \mathcal{E}, P^X) \iff h \in L^1(E, \mathcal{E}, P^X)$ : la catena di equivalenze scritte dà quindi  $h(X) \in L^1(\Omega, \mathcal{A}, \mathbf{P}) \iff h \in L^1(E, \mathcal{E}, P^X)$ , e così si è dimostrata (1).

(Passo 4) Devo dimostrare che per  $h \in L^1(E, \mathcal{E}, P^X)$  vale  $\mathbf{E}(h(X)) = E^X(h)$ . Considero  $h \in L^1(E, \mathcal{E}, P^X)$ :  $h(x) = h_+(x) - h_-(x)$ . Allora  $h(X) = (h(X))_+ - (h(X))_- = h_+(X) - h_-(X)$ , e per  $h_+, h_-$ , essendo v. a.



non negative, per il passo 3 vale  $\mathbf{E}(h_+(X)) = E^X(h_+)$ ,  $\mathbf{E}(h_-(X)) = E^X(h_-)$ , quindi per linearità (8.8 (6)) vale  $\mathbf{E}(h(X)) = \mathbf{E}(h_+(X) - h_-(X)) = \mathbf{E}(h_+(X)) - \mathbf{E}(h_-(X)) = E^X(h_+) - E^X(h_-) = E^X(h)$ . ■

Teo

Hp:  $X : \Omega \rightarrow \mathbb{R}$  è una v.a. con legge  $P^X$ ,  $h : \mathbb{R} \rightarrow [0, +\infty)$

è misurabile,  $h(X) \in L^1$ ,  $\lambda(B) := \int_B h(x) dP^X(x)$

Ts:  $\lambda$  è una misura

## 7 Variabili aleatorie discrete

**Def 10.2** Dato  $(\Omega, \mathcal{A}, \mathbf{P})$  spazio di probabilità e  $X : \Omega \rightarrow \mathbb{R}$  variabile aleatoria con legge  $P^X : \mathcal{B}(\mathbb{R}) \rightarrow [0, 1]$ ,  $X$  si dice discreta se  $P^X$  è una probabilità discreta, cioè se  $\exists T \subseteq \mathbb{R}$  discreto e  $p_X : T \rightarrow [0, 1]$ , detta densità discreta, tali che  $\forall B \in \mathcal{B}(\mathbb{R})$   $P^X(B) = \sum_{x \in T \cap B} p_X(x)$ . Si dice supporto della densità  $p_X$  o supporto di  $X$ , e si indica con  $\mathcal{S}_X$ ,  $\{x \in T : p_X(x) > 0\}$ .

(Questo implica  $\sum_{x \in T} p_X(x) = 1$ )  $\mathcal{S}_X \subseteq T$  è ovviamente discreto. La definizione è poco utilizzabile in pratica: per questo si usano alcune definizioni equivalenti.

**Teo (definizioni equivalenti di v. a. discreta)**

Hp:  $(\Omega, \mathcal{A}, \mathbf{P})$  spazio di probabilità,  $X : \Omega \rightarrow \mathbb{R}$  v. a. con legge  $P^X : \mathcal{B}(\mathbb{R}) \rightarrow [0, 1]$

e funzione di ripartizione  $F^X$

Ts:  $X$  è discreta con  $p_X : T \rightarrow [0, 1]$  avente supporto  $\mathcal{S}_X \iff (1) \sum_{x \in \mathcal{S}_X} F^X(x) - F^X(x^-) = 1$

e  $p_X(x) = F^X(x) - F^X(x^-) \quad \forall x \in T \iff (2) \exists T' \subseteq \mathbb{R} \text{ discreto} : \mathbf{P}(X \in T') = 1$

$\iff (3) \exists \tilde{X} \text{ con } \text{Im}(\tilde{X}) \subseteq T : X = \tilde{X} \text{ q. c.}$

$\mathcal{S}_X$  è anche l'insieme dei punti di discontinuità di  $F$ . (2) significa che è possibile che  $\exists \omega \in \Omega : X(\omega) \notin T'$ , ma comunque  $X$  ha quasi certamente immagine discreta.

Come applico la regola del valore atteso per una v. a. discreta?

**Prop 10.3: regola del valore atteso nel caso discreto**

Hp:  $(\Omega, \mathcal{A}, \mathbf{P})$  spazio di probabilità,  $X : \Omega \rightarrow \mathbb{R}$  v. a. discreta con legge

$P^X : \mathcal{B}(\mathbb{R}) \rightarrow [0, 1]$ , densità  $p_X : T \rightarrow [0, 1]$  e supporto  $\mathcal{S}_X$ ,  $h : \mathbb{R} \rightarrow [0, +\infty]$

misurabile o  $h \in L^1(\mathbb{R}, \mathcal{B}(\mathbb{R}), P^X)$

$$\text{Ts: } \mathbf{E}(h(X)) = E^X(h) = \sum_{x \in \mathcal{S}_X} h(x) p_X(x)$$

**Dim** Per la regola del valore atteso  $\mathbf{E}(h(X)) = \int_{\Omega} h(X(\omega)) d\mathbf{P}(\omega) = \int_{\mathbb{R}} h(x) dP^x(x)$ . Per l'applicazione A2, essendo  $(\mathbb{R}, \mathcal{B}(\mathbb{R}), P^X)$  uno spazio di probabilità con  $P^X$  discreta e  $h : \mathbb{R} \rightarrow [0, +\infty]$  misurabile o  $h \in L^1$ , allora  $E^X(h) = \sum_{x \in \mathcal{S}_X} h(x) p_X(x)$ , da cui la tesi. ■

Nelle ipotesi della proposizione, essa ha varie conseguenze.

Nel caso in cui  $h : \mathbb{R} \rightarrow \mathbb{R}$  sia una funzione misurabile qualsiasi, dato che  $h \in L^1(\mathbb{R}, \mathcal{B}(\mathbb{R}), P^X) \iff |h| \in L^1(\mathbb{R}, \mathcal{B}(\mathbb{R}), P^X)$ , la proposizione permette di dedurre che  $h \in L^1(\mathbb{R}, \mathcal{B}(\mathbb{R}), P^X) \iff \sum_{x \in \mathcal{S}_X} |h(x)| p_X(x) < +\infty$ , cioè se e solo se  $\sum_{x \in \mathcal{S}} h(x) p_X(x)$  converge assolutamente: sto convertendo integrabilità in convergenza assoluta di una serie.

Se in particolare  $h(x) = Id(x) = x$ ,  $\mathbf{E}(X) = \sum_{x \in \mathcal{S}_X} x p_X(x)$ : quindi  $X \in L^1(\Omega, \mathcal{A}, \mathbf{P}) \iff \sum_{x \in \mathcal{S}_X} x p_X(x) < +\infty$ . Questo implica che, se  $\mathcal{S}_X = \mathbb{N}$ ,  $\mathbf{E}(X) = \sum_{k=1}^{+\infty} k p_X(k) = \sum_{k=1}^{+\infty} k \mathbf{P}(X = k) = \mathbf{P}(X = 1) + 2\mathbf{P}(X = 2) + 3\mathbf{P}(X = 3) + \dots = \mathbf{P}(X = 1) + \mathbf{P}(X = 2) + \mathbf{P}(X = 3) + \dots + \mathbf{P}(X = 2) + \mathbf{P}(X = 3) + \dots + \mathbf{P}(X = 3) + \dots = \mathbf{P}(X \geq 1) + \mathbf{P}(X \geq 2) + \mathbf{P}(X \geq 3) + \dots$ . Vale quindi

$$\mathbf{E}(X) = \sum_{k=1}^{+\infty} k p_X(k) = \sum_{k=1}^{+\infty} \mathbf{P}(X \geq k)$$

La proposizione può inoltre essere sfruttata per il calcolo della varianza, ponendo  $h(X) = (X - \mathbf{E}(X))^2$ :  $Var(X) = \mathbf{E}((X - \mathbf{E}(X))^2) = E^X(h) = \sum_{x \in \mathcal{S}_X} (x - \mu)^2 p_X(x)$ .

Si può inoltre valutare se  $X \in L^p$ , ponendo  $h(X) = |X|^p$ :  $X \in L^p \iff \sum_{x \in \mathcal{S}_X} |x|^p p_X(x) < +\infty$ .

**E1** Considero lo spazio di probabilità  $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathbf{P})$  di partenza con  $\mathbf{P}$  esponenziale di parametro  $\lambda > 0$ , e lo spazio misurabile  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ . Considero  $X : \mathbb{R} \rightarrow \mathbb{R}$ ,  $X(\omega) = \lceil \omega \rceil$ : voglio calcolare  $\mathbf{E}(X)$ .  $ImX = \mathbb{Z}$  è discreta, quindi per la definizione equivalente  $P^X$  è una probabilità discreta. Si è già dimostrato che  $P^X$  ha densità geometrica di parametro  $1 - e^{-\lambda}$ , con supporto  $\mathcal{S}_X = \mathbb{N} \setminus \{0\}$ . Per la regola sopra con  $h = Id$   $\mathbf{E}(X) = \sum_{k \in \mathbb{N} \setminus \{0\}} k p_X(k) = \sum_{k=1}^{+\infty} k (1 - e^{-\lambda}) (e^{-\lambda})^{k-1}$ . Si può calcolare  $\sum_{k=1}^{+\infty} k (e^{-\lambda})^{k-1}$  vedendolo come  $\sum_{k=1}^{+\infty} k q^{k-1} = \sum_{k=1}^{+\infty} \frac{d}{dq} (q^k) = \frac{d}{dq} \left( \frac{1}{1-q} - 1 \right) = \frac{1}{(1-q)^2}$ , quindi  $\sum_{k=1}^{+\infty} k (1 - e^{-\lambda}) (e^{-\lambda})^{k-1} = \frac{1}{1 - e^{-\lambda}}$ , cioè il valore atteso di  $X$  con distribuzione geometrica di parametro  $p$  è  $\frac{1}{p}$ .

Avrei anche potuto calcolare  $\mathbf{E}(X) = \int_{\Omega} X(\omega) d\mathbf{P}(\omega) = \int_{\Omega} (XI_{(0,+\infty)})(\omega) d\mathbf{P}(\omega)$ , per il fatto che  $\mathbf{P}(X > 0) = 1$ , cioè  $\mathbf{P}(X = X_{(0,+\infty)}) = 1$ , che è uguale, in base a com'è fatta  $X$ ,  $\int_{\Omega} \sum_{k=1}^{+\infty} kI_{(k-1,k]}(\omega) d\mathbf{P}(\omega) = \mathbf{E}\left(\sum_{k=1}^{+\infty} kI_{(k-1,k]}(\omega)\right) = \sum_{k=1}^{+\infty} \mathbf{E}(kI_{(k-1,k]}(\omega))$  per 9.4, che è uguale a  $\sum_{k=1}^{+\infty} \int_{\Omega} kI_{(k-1,k]}(\omega) d\mathbf{P}(\omega) = \sum_{k=1}^{+\infty} k\mathbf{P}((k-1, k]) = \sum_{k=1}^{+\infty} k \int_{k-1}^k \lambda e^{-\lambda\omega} d\omega = \sum_{k=1}^{+\infty} k e^{-\lambda k} (e^{\lambda} - 1) = \frac{1}{1-e^{-\lambda}}$ , dato che  $kI_{(k-1,k]}(\omega)$ , fissato  $k$ , è una v. a. semplice, e utilizzando la densità associata a  $\mathbf{P}$ . Il calcolo è molto più laborioso nel secondo modo, perché non si è sfruttato il fatto che  $X$  è discreta.

$Var(X) = E^X((X - \mu)^2) = \sum_{k \in \mathcal{S}_X} (k - \mu)^2 p_X(k) = \sum_{k=1}^{+\infty} (k - \mu)^2 p(1-p)^{k-1}$ : la calcolo usando la somma di una serie di una derivata.

**E2** Dato  $(\Omega, \mathcal{A}, \mathbf{P})$  spazio di probabilità e  $T = \{a_1, \dots, a_n\}$  insieme finito con  $a_i \in \mathbb{R}, a_i \neq a_j, X : \Omega \rightarrow \mathbb{R}$  si dice variabile aleatoria uniforme discreta, e si scrive  $X \sim \mathcal{U}(\{a_1, \dots, a_n\})$  se la sua legge ha densità discreta  $p_X : T \rightarrow [0, 1], p_X(a_k) = \frac{1}{n} \forall k = 1, \dots, n$ . Vale quindi  $\mathcal{S}_X = \{a_1, \dots, a_n\}$ . Se in particolare  $T = \{1, \dots, n\}$ , si scrive  $X \sim \mathcal{U}\{1, \dots, n\}$  e vale  $\mathbf{E}(X) = E^X(Id) = \sum_{k \in \mathcal{S}_X} k p_X(k) = \sum_{k=1}^n \frac{k}{n} = \frac{1}{n} \frac{n(n+1)}{2} = \frac{n+1}{2}$ .

$$var(X) = \sum_{k \in \mathcal{S}_X} (k - \mu)^2 p_X(k) = \sum_{k=1}^n (k - \mu)^2 \frac{1}{n} = \frac{1}{n} (\sum_{k=1}^n k^2 + \mu^2 n - 2\mu \sum_{k=1}^n k) = \frac{1}{n} \left( \frac{n(n+1)(2n+1)}{6} + \left(\frac{n+1}{2}\right)^2 n - (n+1) \frac{(n+1)(2n+1)}{6} - \frac{(n+1)^2}{4} \right) = \frac{n^2-1}{12}.$$

**E3** Considero  $(\Omega, \mathcal{A}, \mathbf{P})$  spazio di probabilità qualsiasi e  $X : \Omega \rightarrow \mathbb{R}$  v. a. con legge  $P^X$  discreta, avente densità  $p_X(k) = \frac{6}{\pi^2} \frac{1}{k^2}$  e  $\mathcal{S}_X = \mathbb{N}^*$  ( $\frac{6}{\pi^2}$  è la costante di normalizzazione necessaria affinché  $\sum_{k=1}^{+\infty} p_X(k) = 1$ ).  $X$  è non negativa e q. c. finita:  $\mathbf{P}(X < +\infty) = \mathbf{P}(X \in \mathbb{N}^*) = \sum_{k=1}^{+\infty} p_X(k) = 1$ , ma  $\mathbf{E}(X) = E^X(Id) = \sum_{k=1}^{+\infty} k \frac{6}{\pi^2} \frac{1}{k^2} = +\infty$ , quindi  $X$  non è integrabile, come già visto per  $X(k) = k!$ .

**E4** Dato  $(\Omega, \mathcal{A}, \mathbf{P})$  spazio di probabilità e  $A \in \mathcal{A} : \mathbf{P}(A) = p, X : \Omega \rightarrow \mathbb{R}$  si dice variabile aleatoria di Bernoulli di parametro  $p$ , e si scrive  $X \sim B(p)$ , se  $X(\omega) = I_A(\omega)$ .  $X$  esprime quindi l'essersi realizzato o meno dell'evento  $A$ . E' una v. a. semplice, quindi discreta su  $ImX = T = \{0, 1\}$ , con densità discreta  $p_X(1) = p, p_X(0) = 1-p$ .  $\mathbf{E}(X) = p, var(X) = p^2(1-p) + (1-p)^2 p = p(1-p)$ .

**E5** Dato  $(\Omega, \mathcal{A}, \mathbf{P})$  spazio di probabilità e una successione finita di eventi indipendenti  $\{A_1, \dots, A_n\} \subseteq \mathcal{A}$  tali che  $\mathbf{P}(A_i) = p \forall i = 1, \dots, n, X : \Omega \rightarrow \mathbb{R}$  si dice variabile aleatoria binomiale di parametri  $n, p$ , e si scrive  $X \sim bin(n, p)$ , se  $X(\omega) = I_{A_1}(\omega) + \dots + I_{A_n}(\omega)$ .  $X$  conta quindi quanti eventi si sono realizzati tra  $A_1, \dots, A_n$ . Si può pensare che  $\forall i A_i$  rappresenti un evento relativo alla  $i$ -esima di  $n$  prove realizzate in sequenza. Se  $n = 1$ , si ottiene una v. a. di Bernoulli.  $X$  è una v. a. semplice, quindi discreta su  $ImX = T = \{0, 1, \dots, n\}$ . La densità discreta di  $X$  è, dato  $I \subseteq \{1, \dots, n\} : |I| = k, p_X(k) = \mathbf{P}(X = k) =$

$\sum_{I:|I|=k} \mathbf{P} \left( \bigcap_{i \in I} A_i \cap \bigcap_{k \in I^c} A_k^c \right) = \binom{n}{k} p^k (1-p)^{n-k}$ .  $\mathbf{E}(X) = np$ . Poiché  $\{A_1, \dots, A_n\}$  sono una famiglia di eventi indipendenti, le. v. a.  $X_i(\omega) = I_{A_i}(\omega)$  sono tutte indipendenti, perché generano solo gli eventi  $A_i, \Omega, \emptyset$ . Ne segue che  $\text{var} \left( \sum_{i=1}^n X_i(\omega) \right) = \sum_{i=1}^n \text{var}(X_i) = np(1-p)$ .

**E6** Dato  $(\{0, 1\}^{\mathbb{N}}, 2^{\Omega}, \mathbf{P})$  spazio di probabilità e una successione finita di eventi indipendenti  $\{A_1, \dots, A_n\} \subseteq \mathcal{A} : \mathbf{P}(A_n) = p \forall n$ , data  $X : \Omega \rightarrow \mathbb{R}$  binomiale di parametri  $n, p$  con  $X(\omega) = X_1(\omega) + \dots + X_n(\omega) = I_{A_1}(\omega) + \dots + I_{A_n}(\omega)$ ,  $Z(\omega) = \min \{n \in \mathbb{N} : X_n(\omega) = 1\}$  si dice variabile aleatoria geometrica di parametro  $p$ , e si scrive  $Z \sim \text{geom}(p)$ . Vigge la convenzione  $\min \emptyset = -\infty$ .  $Z$  conta il numero di prove di Bernoulli necessarie per il primo successo, ed è discreta su  $\text{Im} Z = T = \mathbb{N} \setminus \{0\}$ , con densità discreta  $\mathbf{P}(Z = k) = P \left( \bigcap_{i=1}^{k-1} A_i^c \cap A_k \right) = (1-p)^{k-1} p = p_Z(k)$ , per ipotesi di indipendenza.  $\mathbf{E}(Z) = \sum_{k=1}^{+\infty} k(1-p)^{k-1} p = p \sum_{k=1}^{+\infty} \frac{d}{dq} q^k = p \frac{d}{dq} \left( \frac{q}{1-q} \right) = p \frac{1}{(1-q)^2} = \frac{1}{p}$ .  $\text{var}(Z) = \mathbf{E}(Z^2) - \mathbf{E}^2(Z)$ :  $\mathbf{E}(Z^2) = \sum_{k=1}^{+\infty} k^2 (1-p)^{k-1} p = p \sum_{k=1}^{+\infty} k^2 q^{k-1} = p \sum_{k=1}^{+\infty} \frac{d}{dq} (kq^k) = p \frac{d}{dq} \sum_{k=1}^{+\infty} kq^k$ . Ora  $\sum_{k=1}^{+\infty} kq^k = q \sum_{k=1}^{+\infty} \frac{d}{dq} (q^k) = q \frac{d}{dq} \left( \frac{q}{1-q} \right) = \frac{q}{(1-q)^2}$ . Quindi  $\mathbf{E}(Z^2) = p \left( \frac{(1-q)^2 + 2q(1-q)}{(1-q)^4} \right) = p \left( \frac{p^2 + 2p(1-p)}{p^4} \right) = \frac{2-p}{p^2}$  e  $\text{var}(Z) = \frac{2-p}{p^2} - \frac{1}{p^2} = \frac{1-p}{p^2}$ .

Le mode di  $P^X$  sono  $k = \arg \max_{k \in T} \mathbf{P}(X = k) = 1$ .

La funzione di ripartizione della legge di  $Z$  è  $F_Z(t) = P(Z \leq t) = \sum_{\substack{k \in \mathbb{N}^* \\ k \leq t}} \mathbf{P}(Z = k) = \begin{cases} 0 & \text{se } t < 1 \\ \sum_{k=1}^{\lfloor t \rfloor} \mathbf{P}(Z = k) & \text{se } t \geq 1 \end{cases}$   
 $p \sum_{k=1}^{\lfloor t \rfloor} (1-p)^{k-1} = p \frac{1-(1-p)^{\lfloor t \rfloor}}{1-(1-p)} = 1 - (1-p)^{\lfloor t \rfloor}$ , per cui  $F_Z(t) = \left( 1 - (1-p)^{\lfloor t \rfloor} \right) I_{[1, +\infty)}(t)$ .

La geometrica gode della proprietà di assenza di memoria:  $\forall i, j \in \mathbb{N}, \mathbf{P}(Z > i+j | Z > i) = \mathbf{P}(Z > j)$ . Infatti  $\mathbf{P}(Z > i+j | Z > i) = \frac{\mathbf{P}(Z > i+j, Z > i)}{\mathbf{P}(Z > i)} = \frac{\mathbf{P}(Z > i+j)}{\mathbf{P}(Z > i)} = \frac{(1-p)^{i+j}}{(1-p)^i} = (1-p)^j = \mathbf{P}(Z > j)$ . Vale anche il viceversa: se  $Z$  è una v. a. con densità discreta su  $T = \mathbb{N}^*$  tale che  $\forall i, j \in \mathbb{N}, \mathbf{P}(Z > i+j | Z > i) = \mathbf{P}(Z > j)$ , allora  $Z \sim \text{geom}(p)$ , con  $p = \mathbf{P}(Z = 1)$ . Infatti, l'assenza di memoria significa che  $\mathbf{P}(Z > i+j | Z > i) = \frac{\mathbf{P}(Z > i+j)}{\mathbf{P}(Z > i)} = \mathbf{P}(Z > j)$ . Questo è equivalente a dire che  $\mathbf{P}(Z > 2) = \mathbf{P}(Z > 1) \mathbf{P}(Z > 1) = [\mathbf{P}(Z > 1)]^2$ , mentre  $\mathbf{P}(Z > 3) = \mathbf{P}(Z > 2+1) = \mathbf{P}(Z > 2) \mathbf{P}(Z > 1) = [\mathbf{P}(Z > 1)]^3$ , per cui in generale  $\forall s \in \mathbb{N}^* \mathbf{P}(Z > s) = [\mathbf{P}(Z > 1)]^s$ . Avendo  $Z$  supporto  $\mathbb{N}^*$ ,  $\mathbf{P}(Z > 1) = 1 - \mathbf{P}(Z = 1) =: 1 - p$ , cioè  $\mathbf{P}(Z > s) = (1-p)^s$ , e la  $F_Z(s) = 1 - (1-p)^s$ , che caratterizza la legge di una v. a. geometrica.

**E7** Dato  $(\Omega, \mathcal{A}, \mathbf{P})$  spazio di probabilità, una successione infinita di eventi indipendenti  $\{A_1, \dots\} \subseteq \mathcal{A}$  tali che  $\mathbf{P}(A_i) = p \forall i$  e  $X_i(\omega) = I_{A_i}(\omega) \forall i$ ,  $X : \Omega \rightarrow \mathbb{R}$  si dice variabile aleatoria binomiale negativa di parametri  $n, p$ , e si scrive  $X \sim NB(n, p)$ , se  $X(\omega) = \min \left\{ k \in \mathbb{N} : \sum_{i=1}^k X_i(\omega) = n \right\}$ :  $X$  rappresenta il numero di prove necessarie per osservare  $n$  successi.  $X$  è discreta con  $\text{Im} X = \mathbb{N}^*$ .  $\mathbf{P}(X = k) = \binom{k-1}{n-1} p^n (1-p)^{k-n}$ .  $X$  può anche essere vista come somma di  $n$  geometriche indipendenti.

**E7** Dato  $(\Omega, \mathcal{A}, \mathbf{P})$  spazio di probabilità e  $\lambda > 0$ ,  $X : \Omega \rightarrow \mathbb{R}$  si dice variabile aleatoria di Poisson di parametro  $\lambda$ , e si scrive  $X \sim \text{poiss}(\lambda)$ , se, dato  $T = \mathbb{N}$ , la legge di  $X$  è  $p_X : T \rightarrow [0, 1]$ ,  $p_X(k) = e^{-\lambda} \frac{\lambda^k}{k!}$ .  $\mathbf{E}(X) = \sum_{k=0}^n k \cdot e^{-\lambda} \frac{\lambda^k}{k!} = \sum_{k=1}^n k e^{-\lambda} \frac{\lambda^k}{k!} = \lambda e^{-\lambda} \sum_{k=1}^n \frac{\lambda^{k-1}}{(k-1)!} = \lambda$ , nota la serie di Taylor di  $e^\lambda$ .  $\text{var}(X) = \mathbf{E}(X^2) - \lambda^2$ .  $\mathbf{E}(X^2) = \sum_{k=0}^n k^2 \cdot e^{-\lambda} \frac{\lambda^k}{k!} = e^{-\lambda} \left( \sum_{k=0}^n k(k-1) \frac{\lambda^k}{k!} + \sum_{k=0}^n k \frac{\lambda^k}{k!} \right) = e^{-\lambda} \left( \lambda^2 \sum_{k=2}^n \frac{\lambda^{k-2}}{(k-2)!} + \lambda \sum_{k=1}^n \frac{\lambda^{k-1}}{(k-1)!} \right) = \lambda^2 + \lambda$ , per cui  $\text{var}(X) = \lambda$ .

Per massimizzare  $\mathbf{P}(X = k)$  in  $\lambda$ , si calcola  $\frac{d}{d\lambda} \left( e^{-\lambda} \frac{\lambda^k}{k!} \right) = e^{-\lambda} \left( \frac{\lambda^{k-1}}{(k-1)!} - \frac{\lambda^k}{k!} \right) = e^{-\lambda} \lambda^{k-1} \left( \frac{1}{(k-1)!} - \frac{\lambda}{k!} \right)$ , che è nulla se e solo se  $\lambda = k$ .

Se  $X \sim \text{poiss}(\lambda_1)$ ,  $Y \sim \text{poiss}(\lambda_2)$  e  $X \perp Y$ , allora  $X + Y \sim \text{poiss}(\lambda_1 + \lambda_2)$ . Infatti  $\mathbf{P}(X + Y = k) = \sum_{i=0}^k \mathbf{P}(X = i, Y = k - i) = \sum_{i=0}^k e^{-\lambda_1} \frac{\lambda_1^i}{i!} e^{-\lambda_2} \frac{\lambda_2^{k-i}}{(k-i)!} = e^{-(\lambda_1 + \lambda_2)} \frac{1}{k!} \sum_{i=0}^k \binom{k}{i} \lambda_1^i \lambda_2^{k-i} = e^{-(\lambda_1 + \lambda_2)} \frac{(\lambda_1 + \lambda_2)^k}{k!}$ .

## 8 Variabili aleatorie assolutamente continue

Si è visto che esistono misure di probabilità  $\mathbf{P}$  su  $\mathcal{B}(\mathbb{R}) : \exists f \geq 0 : \int_{-\infty}^{+\infty} f(x) dx = 1$  e  $F(x) = \int_{-\infty}^x f(x) dx$ , per cui  $F$  è continua e  $\mathbf{P}(\{x\}) = 0$ . Voglio caratterizzare la classe delle  $f$  così fatte e trovare una formula per calcolare  $\mathbf{E}(h(X)) = E^X(h)$  quando la legge di  $X$  ammette una tale  $f$ .

**Def 11.1** Si dice misura di Lebesgue su  $\mathcal{B}(\mathbb{R})$  una funzione  $m : \mathcal{B}(\mathbb{R}) \rightarrow [0, +\infty]$  con le seguenti proprietà:

**M1**  $m((a, b]) = b - a \forall a, b : -\infty < a < b < +\infty$

**M2**  $\sigma$ -additività:  $\forall (B_n)_{n \geq 1} : B_n \in \mathcal{B}(\mathbb{R}) \forall n$  e  $B_k \cap B_h = \emptyset \forall k \neq h$ , vale  $m\left(\bigcup_{n=1}^{+\infty} B_n\right) = \sum_{n=1}^{+\infty} m(B_n)$

M1 è ben posta perché ogni intervallo appartiene a  $\mathcal{B}(\mathbb{R})$ ; M2 è ben posta perché  $\mathcal{B}(\mathbb{R})$  è chiusa rispetto all'unione numerabile. Si noti che  $m$  ha lo stesso dominio di  $\mathbf{P}$ , ma può valere anche  $+\infty$ ; ha le stesse proprietà di  $\mathbf{P}$ , eccetto quelle legate al fatto che  $\mathbf{P}(\Omega) = 1$ . M2 differenzia la misura di Lebesgue dalla misura di Peano-Jordan: ad esempio la funzione di Dirichlet avrebbe misura nulla se Peano-Jordan avesse M2.

$\mathbf{P}$  è una misura finita e positiva: studiamo i casi in cui è assolutamente continua rispetto alla misura di Lebesgue.

1. Dato l'intervallo  $(a, b) : b - a = 1$ ,  $m : \mathcal{B}((a, b)) \rightarrow [0, +\infty]$  con le proprietà M1 e M2 è una misura di probabilità sullo spazio misurabile  $((a, b), \mathcal{B}((a, b)))$ , perché  $m((a, b)) = 1$ . Tale misura di probabilità è detta uniforme sull'intervallo  $(a, b)$ .

**Teo 11.2 (esistenza e unicità della misura di Lebesgue)**

Hp:  $m : \mathcal{B}(\mathbb{R}) \rightarrow [0, +\infty]$  è la misura di Lebesgue

Ts:  $m$  esiste ed è unica

Enunciamo alcune proprietà della misura di Lebesgue, la cui dimostrazione è identica a quella già vista per le analoghe proprietà di una misura di probabilità.

**Teo (proprietà della misura di Lebesgue)**

Hp:  $m : \mathcal{B}(\mathbb{R}) \rightarrow [0, +\infty]$  è la misura di Lebesgue

Ts: (P1)  $m(\emptyset) = 0$

(P2)  $\forall \{B_1, \dots, B_n\} : B_n \in \mathcal{B}(\mathbb{R}) \ \forall n$  e  $B_k \cap B_h = \emptyset \ \forall k \neq h$ , vale  $m\left(\bigcup_{n=1}^{+\infty} B_n\right) = \sum_{n=1}^{+\infty} m(B_n)$

**Teo (proprietà della misura di Lebesgue)**

Hp:  $m : \mathcal{B}(\mathbb{R}) \rightarrow [0, +\infty]$  è la misura di Lebesgue

Ts: (P3) (monotonia) se  $A, A' \in \mathcal{B}(\mathbb{R})$  e  $A \subseteq A'$ ,  $m(A) \leq m(A')$

(P4) (differenza) se  $A, A' \in \mathcal{A}$ ,  $m(A \setminus A') = m(A) - m(A' \cap A)$

(P5) (unione) se  $A, A' \in \mathcal{A}$ ,  $m(A \cup A') = m(A) + m(A') - m(A \cap A')$

(P6) se  $(A_n)_{n \geq 1}$  è tale che  $A_n \in \mathcal{B}(\mathbb{R}) \ \forall n$ ,  $m\left(\bigcup_{n=1}^{+\infty} A_n\right) \leq \sum_{n=1}^{+\infty} m(A_n)$

(P7) disuguaglianza di Bonferroni: se  $\{A_1, \dots, A_n\}$  è tale che  $A_n \in \mathcal{B}(\mathbb{R})$

$\forall n \geq 1$ ,  $m(A_1 \cap \dots \cap A_n) \geq m(A_1) + \dots + m(A_n) - (n-1) \ \forall n \geq 2$

**Teo 2.4 (continuità della misura)**

Hp:  $m : \mathcal{B}(\mathbb{R}) \rightarrow [0, +\infty]$  è la misura di Lebesgue,  $(A_n)_{n \geq 1} : A_n \in \mathcal{B}(\mathbb{R}) \ \forall n$ ,  $A_n \uparrow A$

Ts: (P9)  $m\left(\bigcup_{n=1}^{+\infty} A_n\right) = \lim_{n \rightarrow +\infty} m(A_n)$

Hp:  $m : \mathcal{B}(\mathbb{R}) \rightarrow [0, +\infty]$  è la misura di Lebesgue,  $(A_n)_{n \geq 1} : A_n \in \mathcal{B}(\mathbb{R}) \ \forall n$ ,  $A_n \downarrow A$

Ts: (P10)  $m\left(\bigcap_{n=1}^{+\infty} A_n\right) = \lim_{n \rightarrow +\infty} m(A_n)$

## Proprietà ulteriori

Hp:  $m : \mathcal{B}(\mathbb{R}) \rightarrow [0, +\infty]$  è la misura di Lebesgue,  $a, b \in \mathbb{R}$

$$\text{Ts: (i) } m(\mathbb{R}) = +\infty$$

$$\text{(ii) } m((-\infty, a]) = +\infty = m([b, +\infty))$$

$$\text{(iii) } m(\{a\}) = 0$$

$$\text{(iv) } m(\mathbb{Q}) = 0$$

$$\text{(v) } m((a, b]) = m((a, b)) = m([a, b)) = m([a, b])$$

(i) rende evidente che la misura di Lebesgue non è una probabilità:  $\mathbf{P}$  e  $m$  sono due diverse misure sullo spazio misurabile  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ .

**Dim** (i)  $\mathbb{R} = \bigcup_{n=1}^{+\infty} (1-n, 1+n]$ . Per la continuità della probabilità (i)  $m(\mathbb{R}) = m\left(\bigcup_{n=1}^{+\infty} (1-n, 1+n]\right) = \lim_{n \rightarrow +\infty} m((1-n, 1+n]) = \lim_{n \rightarrow +\infty} 2n = +\infty$ .

(ii)  $(-\infty, a] = \bigcup_{n=1}^{+\infty} (a-n, a]$ , quindi per la continuità della probabilità  $m((-\infty, a]) = \lim_{n \rightarrow +\infty} m((a-n, a]) = +\infty$ .

(iii)  $\{a\} = \bigcap_{n=1}^{+\infty} (a - \frac{1}{n}, a + \frac{1}{n}]$ , perciò  $m(\{a\}) = \lim_{n \rightarrow +\infty} m((a - \frac{1}{n}, a + \frac{1}{n}]) = \lim_{n \rightarrow +\infty} \frac{2}{n} = 0$ .

(iv)  $\mathbb{Q} = \bigcup_{n=1}^{+\infty} \{q_n\}$ , dove  $q_n \in \mathbb{Q}$  (si può prendere e. g. la successione di Cantor per numerare  $\mathbb{Q}$ ): allora per  $\sigma$ -additività, per (iii)  $m(\mathbb{Q}) = \sum_{n=1}^{+\infty} m(\{q_n\}) = 0$ .

(v) Ciascun intervallo può essere scritto a partire da  $(a, b]$  aggiungendo o togliendo uno o più estremi: dalla  $\sigma$ -additività e da (iii) segue la tesi. ■

**Def** Data la misura di Lebesgue  $m : \mathcal{B}(\mathbb{R}) \rightarrow [0, +\infty]$ , si dice che una proprietà dipendente da  $x \in \mathbb{R}$  vale quasi ovunque se vale  $\forall x \in \mathbb{R}$  tranne al più in un insieme  $A \in \mathcal{B}(\mathbb{R})$  tale che  $m(A) = 0$ . In particolare, si dice che  $f, g : \mathbb{R} \rightarrow \mathbb{R}$  misurabili sono uguali quasi ovunque se  $A = \{x \in \mathbb{R} : f(x) \neq g(x)\}$  è tale che  $m(A) = 0$ .

## 8.1 Integrazione rispetto a una misura

La teoria dell'integrazione per funzioni  $h : (\mathbb{R}, \mathcal{B}(\mathbb{R}), m) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$  è analoga a quella già vista per variabili aleatorie: si definisce l'integrale rispetto alla misura di Lebesgue, detto integrale di Lebesgue, in maniera identica a quella già vista per l'integrale rispetto a una misura di probabilità.

Considero la misura di Lebesgue  $m : \mathcal{B}(\mathbb{R}) \rightarrow [0, +\infty]$  e la funzione  $f : \mathbb{R} \rightarrow \mathbb{R}$ . Si indica l'integrale di  $f$  rispetto alla misura di Lebesgue con  $\int_{\mathbb{R}} f(x) m(dx)$  o  $\int_{\mathbb{R}} f(x) dm(x)$  o  $\int_{\mathbb{R}} f(x) dm$ . Si definisce come segue.

[PASSO 1

**Def** Data la misura di Lebesgue  $m : \mathcal{B}(\mathbb{R}) \rightarrow [0, +\infty]$  e la funzione  $f : \mathbb{R} \rightarrow \mathbb{R}$  misurabile,  $f$  si dice semplice se assume un numero finito di valori, cioè se  $|Imf| = n < +\infty$ .

Questo è equivalente a dire che si può scrivere, data  $\{A_1, A_2, \dots, A_n\}$  partizione finita di  $\mathbb{R}$  e  $a_1, \dots, a_n \in \mathbb{R}$ ,  $f(x) = \sum_{k=1}^n a_k I_{A_k}(x) = a_1 I_{A_1}(x) + \dots + a_n I_{A_n}(x)$ . Se  $x \in A_1$ ,  $f(x) = a_1$ , eccetera.  $f^{-1}(\{a_k\}) = A_k$ ,  $f(\mathbb{R}) = Imf = \{a_1, \dots, a_n\}$ .

**Def** Data la misura di Lebesgue  $m : \mathcal{B}(\mathbb{R}) \rightarrow [0, +\infty]$  e  $f : \mathbb{R} \rightarrow \mathbb{R}$  funzione misurabile semplice, si dice integrale di  $f$  secondo la misura  $m$ , e si indica con  $\int_{\mathbb{R}} f(x) dm$ ,  $\sum_{k=1}^n a_k m(A_k)$ .

Serve che  $f$  sia misurabile affinché  $A_k \in \mathcal{B}(\mathbb{R})$ .

**Prop**

Hp:  $m : \mathcal{B}(\mathbb{R}) \rightarrow [0, +\infty]$  è la misura di Lebesgue,  $f, g : \mathbb{R} \rightarrow \mathbb{R}$

sono funzioni semplici misurabili,  $\alpha, \beta \in \mathbb{R}$

Ts: (i)  $\alpha f + \beta g$  è una funzione semplice

$$(ii) \text{ (linearità)} \quad \int_{\mathbb{R}} (\alpha f + \beta g) dm = \alpha \int_{\mathbb{R}} f dm + \beta \int_{\mathbb{R}} g dm$$

$$(iii) \text{ (monotonia)} \text{ se } f(x) \leq g(x) \quad \forall x \in \mathbb{R}, \quad \int_{\mathbb{R}} f dm \leq \int_{\mathbb{R}} g dm$$

PASSO 2

**Prop (approssimazione di una funzione)**

Hp:  $m : \mathcal{B}(\mathbb{R}) \rightarrow [0, +\infty]$  è la misura di Lebesgue,  $f : \mathbb{R} \rightarrow [0, +\infty]$

è misurabile

Ts:  $\exists (f_n)_{n \geq 1} : (1) f_n : \mathbb{R} \rightarrow [0, +\infty)$  è semplice  $\forall n \geq 1$ ,

(2)  $\forall x \in \mathbb{R} \quad \lim_{n \rightarrow +\infty} f_n(x) = f(x)$  e  $f_n$  è monotona non decrescente

$f_n$  monotona non decrescente significa che  $f_n(x) \leq f_{n+1}(x) \quad \forall n, \forall x$ : (2) afferma quindi che  $(f_n)_{n \geq 1}$  converge puntualmente a  $f$  in maniera monotona: si scrive  $f_n \nearrow f$ .

Come costruisco tale successione? Una possibilità è

$$\text{fissato } n \geq 1, \text{ al variare di } k = 0, \dots, n2^n - 1, f_n(x) = \begin{cases} \frac{k}{2^n} & \text{se } f(x) \in [\frac{k}{2^n}, \frac{k+1}{2^n}] \\ n & \text{se } f(x) \geq n \end{cases}$$

Quindi, fissato  $n$ ,  $f_n(x)$  è un'approssimazione per difetto di  $f$ : quando  $f$  assume valori in  $[a, b]$ ,  $f_n$  assume valore  $a$ . Al crescere di  $n$ ,  $f_n$  approssima  $f$  sempre più finemente, perché la partizione dell'asse  $y$  è più fitta. Chia-



ramente  $f_n$  è semplice:  $Im(f_n) = \{0, \frac{1}{2^n}, \dots, n\}$ , con cardinalità  $n2^n$ ;  $\int_{\mathbb{R}} f_n dm = \sum_{k=0}^{n2^n-1} \frac{k}{2^n} m(f^{-1}([\frac{k}{2^n}, \frac{k+1}{2^n}])) + nm(f^{-1}[n, +\infty))$ .

Si dimostra che  $f_{n+1}(x) \geq f_n(x) \forall x \in \mathbb{R}, \forall n$ , quindi  $\int_{\mathbb{R}} f_{n+1} dm \geq \int_{\mathbb{R}} f_n dm \geq 0$ : allora  $(\int_{\mathbb{R}} f_n dm)_{n \geq 1}$  è una successione a valori reali non negativi monotona non decrescente, che può tendere a  $L \geq 0$  o  $+\infty$  per  $n \rightarrow +\infty$ .

**Def** Data la misura di Lebesgue  $m : \mathcal{B}(\mathbb{R}) \rightarrow [0, +\infty]$  e  $f : \mathbb{R} \rightarrow [0, +\infty]$ , si dice integrale di  $f$  rispetto alla misura di Lebesgue, e si indica con  $\int_{\mathbb{R}} f dm$ ,  $\lim_{n \rightarrow +\infty} \int_{\mathbb{R}} f_n dm$ .

L'integrale può essere  $+\infty$ . La definizione non dipende dalla scelta di  $(X_n)_{n \geq 1}$ , purché sia una successione di funzioni semplici che cresce a  $f$ .

### PASSO 3

E' noto che  $f(x) = f_+(x) - f_-(x)$ .

**Def** Data la misura di Lebesgue  $m : \mathcal{B}(\mathbb{R}) \rightarrow [0, +\infty]$  e  $f : \mathbb{R} \rightarrow \mathbb{R}$  con parte positiva  $f_+$  e parte negativa  $f_-$ , si dice che  $f$  ammette integrale di Lebesgue se almeno uno tra  $\int_{\mathbb{R}} f_+ dm$  e  $\int_{\mathbb{R}} f_- dm$  è finito. In tal caso si dice integrale di Lebesgue di  $f$ , e si indica con  $\int_{\mathbb{R}} f dm$ ,  $\int_{\mathbb{R}} f_+ dm - \int_{\mathbb{R}} f_- dm$ . Se  $\int_{\mathbb{R}} f_+ dm = \int_{\mathbb{R}} f_- dm = +\infty$ , si dice che  $f$  non ammette valore atteso.

**Def** Data la misura di Lebesgue  $m : \mathcal{B}(\mathbb{R}) \rightarrow [0, +\infty]$  e  $f : \mathbb{R} \rightarrow \mathbb{R}$  con parte positiva  $f_+$  e parte negativa  $f_-$ , si dice che  $f$  è integrabile secondo Lebesgue se  $\int_{\mathbb{R}} f_+ dm$  e  $\int_{\mathbb{R}} f_- dm$  sono finiti.

In tal caso l'integrale di Lebesgue di  $f$  è  $\int_{\mathbb{R}} f dm = \int_{\mathbb{R}} f_+ dm - \int_{\mathbb{R}} f_- dm \in \mathbb{R}$ .

Non servono ulteriori definizioni per trattare quelli che per Riemann erano integrali impropri.

**Def** Data la misura di Lebesgue  $m : \mathcal{B}(\mathbb{R}) \rightarrow [0, +\infty]$ , si indica con  $\mathcal{L}^1(m)$  lo spazio delle funzioni  $h : \mathbb{R} \rightarrow \mathbb{R}$  integrabili secondo Lebesgue, cioè  $\mathcal{L}^1(m) = \{h : \mathbb{R} \rightarrow \mathbb{R} \text{ boreliane tali che } h \text{ ammette integrale di Lebesgue finito}\}$ .

### Teo (proprietà dell'integrale di Lebesgue)

1)  $\mathcal{L}^1(m)$  è uno spazio vettoriale

2) Hp:  $f, g \in \mathcal{L}^1, \alpha, \beta \in \mathbb{R}$

$$\text{Ts: (i) (linearità) } \int_{\mathbb{R}} (\alpha f + \beta g) dm = \alpha \int_{\mathbb{R}} f dm + \beta \int_{\mathbb{R}} g dm$$

$$(ii) \text{ (positività) se } f \geq 0, \int_{\mathbb{R}} f dm \geq 0$$

$$3) \text{ Hp: } f, g : \mathbb{R} \rightarrow \mathbb{R}, g \in \mathcal{L}^1(m), g \geq f \geq 0$$

$$\text{Ts: } f \in \mathcal{L}^1, (\text{monotonia}) \int_{\mathbb{R}} g dm \geq \int_{\mathbb{R}} f dm$$

$$4) \text{ Hp: } f : \mathbb{R} \rightarrow \mathbb{R}, f \geq 0, \int_{\mathbb{R}} f dm = 0$$

$$\text{Ts: } f = 0 \text{ quasi ovunque}$$

$$5) \text{ Hp: } f, g : \mathbb{R} \rightarrow \mathbb{R} \text{ ammettono integrale di Lebesgue, } f = g \text{ q. o.}$$

$$\text{Ts: } \int_{\mathbb{R}} g dm = \int_{\mathbb{R}} f dm$$

1. La funzione di Dirichlet è una funzione semplice e  $I_{\mathbb{Q}} = 0$  q. o. perché  $m(\mathbb{Q}) = 0$ , dunque  $\int_{\mathbb{R}} I_{\mathbb{Q}} dm = 0$ .

$$6) \text{ Hp: } f, g : \mathbb{R} \rightarrow \mathbb{R}, f, g \geq 0, \alpha, \beta \geq 0$$

$$\text{Ts: } \int_{\mathbb{R}} (\alpha f + \beta g) dm = \alpha \int_{\mathbb{R}} f dm + \beta \int_{\mathbb{R}} g dm$$

$$7) f \in \mathcal{L}^1(m) \iff |f| \in \mathcal{L}^1(m), \text{ e in tal caso}$$

$$\left| \int_{\mathbb{R}} f dm \right| \leq \int_{\mathbb{R}} |f| dm$$

7) evidenzia la differenza tra l'integrale di Lebesgue e l'integrale di Riemann: se  $f$  è Riemann integrabile,  $|f|$  è Riemann integrabile, ma non vale il viceversa. Vale una disuguaglianza analoga a  $|\mathbf{E}(X)| \leq \mathbf{E}(|X|)$  per l'integrale di Riemann.

### **Teorema di Beppo Levi (convergenza monotona)**

$$\text{Hp: } (f_n)_{n \geq 1} \text{ è una successione di funzioni misurabili, } f : \mathbb{R} \rightarrow \mathbb{R}$$

$$\text{è una funzione misurabile t. c. } f_n \nearrow f \text{ q. c.}$$

$$\text{Ts: } \int_{\mathbb{R}} f dm = \lim_{n \rightarrow +\infty} \int_{\mathbb{R}} f_n dm$$

Il teorema afferma che per una successione di funzioni misurabili non negative che cresca a  $f$  è possibile scambiare limite e integrale.

### **Teorema di convergenza dominata**

$$\text{Hp: } (f_n)_{n \geq 1} \text{ è una successione di funzioni misurabili, } g \in \mathcal{L}^1(m)$$

$$\forall n \geq 1 \text{ vale } |f_n| \leq g \text{ q. o. e } \lim_{n \rightarrow +\infty} f_n(x) = f(x) \text{ q. o.}$$

$$\text{Ts: } \int_{\mathbb{R}} f dm = \lim_{n \rightarrow +\infty} \int_{\mathbb{R}} f_n dm$$

Se  $\exists c \in \mathbb{R} : f = c$  q. o., non è vero che  $f \in \mathcal{L}^1(m)$ , perché  $\int_{\mathbb{R}} f dm = \int_{\mathbb{R}} c dm = c \int_{\mathbb{R}} dm = cm(\mathbb{R}) = +\infty$ , essendo 1 una funzione semplice.]

In generale non è vero se  $f$  è limitata, allora ha integrale di Lebesgue finito.

Inoltre  $1 \leq p < q$  non implica  $L^q \subseteq L^p \subseteq L^1$ , perché  $m(\mathbb{R})$  non è un numero reale.

**Def** Data la misura di Lebesgue  $m : \mathcal{B}(\mathbb{R}) \rightarrow [0, +\infty]$  e  $h : \mathbb{R} \rightarrow \mathbb{R}$ , dato  $A \in \mathcal{B}(\mathbb{R})$ , si definisce  $\int_A h dm$  come  $\int_{\mathbb{R}} h I_A dm$ .

Si può inoltre notare  $\int_a^b 1 dx = b - a$ : l'integrale di Riemann su un certo intervallo limitato coincide con la misura di Lebesgue di tale intervallo.

**Prop 11.3 (confronto tra Riemann e Lebesgue)**

Hp:  $h : [a, b] \rightarrow \mathbb{R}$  è limitata e Riemann-integrabile

Ts:  $h$  è Lebesgue-integrabile e  $\int_a^b h(x) dx = \int_{[a,b]} h dm = \int_{\mathbb{R}} h I_{[a,b]} dm$

Questo significa che l'insieme delle funzioni Riemann integrabili su un intervallo compatto sono un sottinsieme delle funzioni Lebesgue integrabili sullo stesso intervallo. E' quindi molto utile sfruttare l'uguaglianza con l'integrale di Lebesgue perché per esso valgono i teoremi di passaggio al limite sotto il segno di integrale.

1.  $h(x) = \frac{1}{x^2} I_{[1,+\infty)}(x)$ :  $h \geq 0$ , quindi  $\exists \int_{\mathbb{R}} h dm$ . Voglio sapere se è finito. Uso il teorema di convergenza monotona mediante le funzioni troncate:  $h_n(x) = \frac{1}{x^2} I_{[1,n]}(x)$ ,  $n \geq 1$ .  $h_n \geq 0 \forall n$ ,  $h(x) \geq h_{n+1}(x) \geq h_n(x) \forall x$  e  $\lim_{n \rightarrow +\infty} h_n(x) = h(x) \forall x$ . Allora  $\int_{\mathbb{R}} h dm = \lim_{n \rightarrow +\infty} \int_{\mathbb{R}} h_n dm = \lim_{n \rightarrow +\infty} \int_{\mathbb{R}} \frac{1}{x^2} I_{[1,n]}(x) dm = \lim_{n \rightarrow +\infty} \int_{[1,n]} \frac{1}{x^2} dm$ . Ma su  $[1, n]$  l'integrale di Lebesgue coincide con quello di Riemann, quindi quanto scritto coincide con  $\lim_{n \rightarrow +\infty} \int_1^n \frac{1}{x^2} dx = 1$ .

Il ragionamento può essere generalizzato a  $h(x) = f(x) I_{[a,+\infty)}(x)$  con  $f \geq 0$ :  $h \geq 0$ , quindi  $\exists \int_{\mathbb{R}} h dm$ . Voglio sapere se è finito. Uso il teorema di convergenza monotona mediante le funzioni troncate:  $h_n(x) = f(x) I_{[a,n]}(x)$ ,  $n \geq [a]$ .  $h_n \geq 0 \forall n$ ,  $h(x) \geq h_{n+1}(x) \geq h_n(x) \forall x$  e  $\lim_{n \rightarrow +\infty} h_n(x) = h(x) \forall x$ . Allora  $\int_{\mathbb{R}} h dm = \lim_{n \rightarrow +\infty} \int_{\mathbb{R}} h_n dm = \lim_{n \rightarrow +\infty} \int_{\mathbb{R}} f(x) I_{[a,n]}(x) dm = \lim_{n \rightarrow +\infty} \int_{[a,n]} f(x) dm$ . Ma su  $[1, n]$  l'integrale di Lebesgue coincide con quello di Riemann, quindi quanto scritto coincide con  $\lim_{n \rightarrow +\infty} \int_1^n f(x) dx$ .

Si può dire che  $f \sim g$  se  $\int_{\mathbb{R}} f dm = \int_{\mathbb{R}} g dm$ : è una relazione di equivalenza che induce una partizione di  $\mathcal{L}^1(m)$  in classi di equivalenza  $[f]$ .

**Def 11.4** Dato lo spazio di probabilità  $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathbf{P})$ , con  $F$  funzione di ripartizione di  $\mathbf{P}$ ,  $\mathbf{P}$  è detta misura di probabilità assolutamente continua se  $\exists f : \mathbb{R} \rightarrow \mathbb{R}$  boreliana tale che  $f(x) \geq 0 \forall x$  e  $F(x) = \int_{(-\infty, x]} f dm \forall x \in \mathbb{R}$ . In tal caso  $f$  si dice densità continua di  $F$  o di  $\mathbf{P}$  e  $F$  si dice assolutamente continua.

Esiste anche una definizione epsilon-delta per definire l'assoluta continuità di una funzione  $F$ . Nel seguito si scriverà  $\int_{-\infty}^x f(x) dx$ , integrale di Riemann. Il nome di densità continua per  $f$  non fa riferimento a proprietà analitiche di  $f$ , cioè non significa che  $f$  sia continua, ma solo che  $\mathbf{P}$  è assolutamente continua. Se  $\mathbf{P}$  è assolutamente continua, è continua  $F$ , per il teorema fondamentale, per cui  $\mathbf{P}(\{x\}) = F(x) - F(x^-) = 0$ .

Se  $\mathbf{P}$  è assolutamente continua con densità  $f$ ,  $f$  è integrabile secondo Lebesgue perché misurabile e nonnegativa. Appartiene a  $L^1(m)$ ? Data  $f_n(x) = f(x) I_{(-\infty, n]}(x)$ , per convergenza monotona ( $f_n \nearrow f$ ) vale  $\int_{\mathbb{R}} f dm = \lim_{n \rightarrow +\infty} \int_{\mathbb{R}} f_n dm = \lim_{n \rightarrow +\infty} \int_{\mathbb{R}} f(x) I_{(-\infty, n]}(x) dm = \lim_{n \rightarrow +\infty} \int_{(-\infty, n]} f(x) dm = \lim_{n \rightarrow +\infty} \int_{-\infty}^n f(x) dx = \lim_{n \rightarrow +\infty} F(n) = 1$  per le proprietà di  $F$  in quanto funzione di ripartizione.

### Teo 11.5 (caratterizzazione della densità)

Hp:  $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathbf{P})$  è uno spazio di probabilità,  $f : \mathbb{R} \rightarrow \mathbb{R}$

Ts: (i)  $\mathbf{P}$  è assolutamente continua con densità  $f \iff f$  è boreliana,

$$f(x) \geq 0 \forall x \text{ e } \int_{\mathbb{R}} f dm = 1$$

(ii) se  $\mathbf{P}$  è assolutamente continua con densità  $f$ ,  $\mathbf{P}(A) = \int_A f dm \forall A \in \mathcal{B}(\mathbb{R})$

(iii) se  $f_1, f_2$  sono densità per  $\mathbf{P}$  assolutamente continua, allora  $f_1 = f_2$  q.o.

In (ii) vale anche l'implicazione inversa. (iii) significa che  $\mathbf{P}$  assolutamente continua caratterizza  $[f]$ , che è un insieme di funzioni uguali a meno di insiemi di misura nulla. Si scrive  $f = \frac{d\mathbf{P}}{dm}$ , con la derivata di Radon-Nikodym.

**Dim\*** (ii) Definisco  $\tilde{\mathbf{P}} : \mathcal{B}(\mathbb{R}) \rightarrow [0, 1]$ ,  $\tilde{\mathbf{P}}(A) = \int_A f dm \forall A \in \mathcal{B}(\mathbb{R})$ .  $\tilde{\mathbf{P}}$  è una misura di probabilità perché  $\tilde{\mathbf{P}}(\mathbb{R}) = 1$  e si può dimostrare la  $\sigma$ -additività. Allora la funzione di ripartizione di  $\tilde{\mathbf{P}}$  è  $\tilde{F}(x) = \tilde{\mathbf{P}}((-\infty, x]) = \int_{-\infty}^x f(t) dt$ , che coincide con  $F(x)$  per definizione di probabilità assolutamente continua. Poiché la funzione di ripartizione caratterizza la  $\mathbf{P}$ , vale  $\tilde{\mathbf{P}} = \mathbf{P}$ , ed effettivamente  $\mathbf{P}(A) = \int_A f dm \forall A \in \mathcal{B}(\mathbb{R})$ .

(iii) Considero  $f_1, f_2$  densità per  $\mathbf{P}$  assolutamente continua. Dimostro che  $m(f_1 \neq f_2) = 0$ , mostrando che  $m(f_1 > f_2) = 0$  e  $m(f_1 < f_2) = 0$ . Dato  $\varepsilon > 0$ , definisco  $A = \{x : f_1(x) + \varepsilon < f_2(x)\} \subseteq \{x : f_1(x) < f_2(x)\}$  (dato  $\varepsilon$ , in  $A$  prendo gli  $x$  che permettono di tenere una distanza di almeno  $\varepsilon$  da  $f_2$ ). Se  $\varepsilon \rightarrow 0$ ,  $A \uparrow \{x : f_1(x) < f_2(x)\}$ . Allora, integrando in  $dm$   $f_1 + \varepsilon$  e  $f_2$ , si ottiene  $\int_A (f_1 + \varepsilon) dm \leq \int_A f_2 dm$ , cioè, per linearità,  $\mathbf{P}(A) + \varepsilon m(A) \leq \mathbf{P}(A)$ . Essendo  $\varepsilon > 0$ , questo implica  $m(A) = 0$  e quindi  $m(\{x : f_1(x) < f_2(x)\}) = 0$ . Analogamente, se  $A = \{x : f_1(x) < f_2(x) + \varepsilon\}$ , si mostra che  $m(A) = 0$  e che  $m(\{x : f_1(x) > f_2(x)\}) = 0$ , per cui  $m(f_1 \neq f_2) = m(\{x : f_1(x) < f_2(x)\}) + m(\{x : f_1(x) > f_2(x)\}) = 0$ . ■

Data  $F$ , come capisco se esiste una densità? Se  $F$  è di puro salto, è la funzione di ripartizione di una  $\mathbf{P}$  discreta. Se  $F$  è continua e  $C^1$  a tratti, allora si può dimostrare che la probabilità che essa caratterizza è assolutamente

continua: una sua densità è  $f(t) = \begin{cases} F'(x) & \text{se esiste} \\ 0 & \text{altrimenti} \end{cases} : C^{(0)} \text{ a tratti, è nonnegativa perché } F \text{ è monotona non}$   
decescente, ha integrale 1 per il teorema fondamentale; inoltre è misurabile perché continua a tratti (quindi ogni  
controimmagine può essere scritta come unione di controimmagini, ciascuna relativa a un tratto in cui  $f$  è continua  
e quindi misurabile, ed eventualmente l'insieme dei punti di discontinuità), ed è una densità di  $\mathbf{P}$ .

Quanto detto sulla densità implica anche che se  $\mathbf{P}$  è assolutamente continua con densità  $f$ , non è possibile che  
 $\exists \alpha > 0 : \lim_{x \rightarrow +\infty} f(x) = \alpha$ . Infatti, se così fosse, preso  $\varepsilon = \frac{\alpha}{2}$ , esisterebbe  $\nu : x > \nu$  implica  $\frac{\alpha}{2} < f(x) < \frac{3}{2}\alpha$ .  
In tal caso si avrebbe quindi  $\int_{-\infty}^{+\infty} f(x) dx = 1 \geq \int_{\nu}^{+\infty} f(x) dx \geq \int_{\nu}^{+\infty} \frac{\alpha}{2} dx = +\infty$ , che è assurdo. In generale, se  
esiste  $\lim_{x \rightarrow +\infty} f(x)$ , allora dev'essere  $\lim_{x \rightarrow +\infty} f(x) = 0$ .

**Def 11.6** Dato  $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathbf{P})$  spazio di probabilità e  $X : \Omega \rightarrow \mathbb{R}$  variabile aleatoria,  $X$  si dice assolutamente  
continua se la sua legge  $P^X : \mathcal{B}(\mathbb{R}) \rightarrow [0, 1]$  è assolutamente continua; in tal caso la densità continua di  $P^X$  si  
indica con  $f_X$ .

$f_X$  è spesso detta, con abuso, densità di  $X$ . L'essere assolutamente continua di  $X$  non dice niente sulla sua  
regolarità.

$X$  è assolutamente continua e  $X \in [a, b]$  q. c.  $\iff \exists f_X : f_X(x) = 0$  q. o. su  $[a, b]^c$ .

Infatti, se  $X \in [a, b]$  q. c. e  $X$  è assolutamente continua,  $\mathbf{P}(X \in A) = \mathbf{P}(X \in A \cap [a, b])$  (perché la probabilità  
dell'intersezione tra un evento qualsiasi e uno quasi certo è uguale alla probabilità dell'evento qualsiasi) e si ottiene  
 $\int_{A \cap [a, b]} f_X dm = \int_{A \cap [a, b]} f_X(t) dt = \int_A f_X(t) I_{[a, b]}(t) dt$ : se quindi  $A = [a, b]^c$   $\mathbf{P}(X \in A) = \int_A f_X dm = 0$ , cioè  
 $f_X \geq 0$  è nulla quasi ovunque in  $[a, b]^c$ .

Se invece  $\exists f_X : f_X(x) = 0$  q. o. su  $[a, b]^c$ ,  $X$  è assolutamente continua, e  $\mathbf{P}(X \in [a, b]) = \int_{[a, b]} f_X(t) dt = 1$   
perché  $\int_{[a, b]^c} f_X(t) dt = 0$  e  $\int_{\mathbb{R}} f_X(t) dt = \int_{[a, b]} f_X(t) dt + \int_{[a, b]^c} f_X(t) dt = 1$ .

Dato  $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathbf{P})$  e  $X : \Omega \rightarrow \mathbb{R}$  v. a. assolutamente continua con  $f_X, F_X$ , considero  $h : \mathbb{R} \rightarrow \mathbb{R}$   $C^1$  con  
 $h'(x) \neq 0 \forall x$  e mi chiedo quale sia  $F_Y$ , con  $Y = h(X)$ , se  $Y$  sia assolutamente continua, e, se sì, quale sia  $f_Y$ .

$F_Y(t) = \mathbf{P}(h(X) \leq t) = \mathbf{P}(X \leq h^{-1}(t)) = F_X(h^{-1}(t))$ , che è  $C^1$  grazie alle ipotesi poste. Vale quindi  
 $f_Y(t) = F'_Y(t) = |g'(t)| f_X(g(t))$ , posto  $h^{-1}(t) = g(t)$ .

1. Dato  $(\Omega, \mathcal{A}, \mathbf{P})$  e  $X : \Omega \rightarrow \mathbb{R}$  v. a. assolutamente continua con densità  $f_X$ , se  $\exists \mu : f_X(\mu + x) = f_X(\mu - x)$   
 $\forall x \in \mathbb{R}$ , cioè simmetrica rispetto a  $x = \mu$ , allora  $\mu$  è una mediana, cioè  $P(X \leq \mu) \geq \frac{1}{2}$  e  $P(X < \mu) \leq \frac{1}{2}$ .

Infatti  $P(X \leq \mu) = \int_{-\infty}^{\mu} f_X(t) dt$ , ma per 11.5  $\int_{\mathbb{R}} f_X(t) dt = 1 = \int_{-\infty}^{\mu} f_X(t) dt + \int_{\mu}^{+\infty} f_X(t) dt$ . Con il  
cambio di variabile  $z = t - \mu$   $\int_{-\infty}^{\mu} f_X(t) dt = \int_{-\infty}^0 f_X(z + \mu) dz$ , con quello  $t = \mu - w$   $\int_{\mu}^{+\infty} f_X(t) dt =$

$-\int_0^{-\infty} f_X(\mu - w) dw = \int_{-\infty}^0 f_X(\mu - w) dw = \int_{-\infty}^0 f_X(\mu - z) dz$ . Dato che  $f_X(z + \mu) = f_X(\mu - z) \forall z \in \mathbb{R}$ ,  
 $1 = 2 \int_{-\infty}^0 f_X(z + \mu) dz$ , cioè  $\int_{-\infty}^{\mu} f_X(t) dt = \frac{1}{2} = \mathbf{P}(X \leq \mu)$ .

Se inoltre  $X \in L^1(\Omega, \mathcal{A}, \mathbf{P})$ ,  $\mathbf{E}(X) = \mu$ . Infatti  $X - \mu$  e  $\mu - X$  hanno la stessa legge, perché  $\mathbf{P}(X - \mu \leq t) = \mathbf{P}(X \leq t + \mu) = \int_{-\infty}^{t+\mu} f_X(z) dz$ , mentre  $\mathbf{P}(\mu - X \leq t) = \mathbf{P}(X \geq \mu - t) = \int_{\mu-t}^{+\infty} f_X(z) dz$ , che coincide con  $\mathbf{P}(X - \mu \leq t)$ , come si può verificare con un cambio di variabile analogo a quello sopra. Poiché due v. a. con stessa legge hanno lo stesso valore atteso, vale  $\mathbf{E}(X - \mu) = \mathbf{E}(\mu - X)$ , cioè, per linearità,  $\mathbf{E}(X) = \mu$ .

Come calcolo  $\mathbf{E}(X)$  quando  $X$  è assolutamente continua?

**Teo 11.7 (regola del valore atteso nel caso assolutamente continuo)**

Hp:  $(\Omega, \mathcal{A}, \mathbf{P})$  è uno spazio di probabilità,  $X : \Omega \rightarrow \mathbb{R}$  è una v.a. assolutamente continua, con legge  $P^X$  di densità  $f_X$ ;  $h : \mathbb{R} \rightarrow \mathbb{R}$  è una funzione boreliana

Ts: (1)  $h \in L^1(\mathbb{R}, \mathcal{B}(\mathbb{R}), P^X) \iff hf_X \in L^1(\mathbb{R}, \mathcal{B}(\mathbb{R}), m)$

(2) se  $h \in L^1(\mathbb{R}, \mathcal{B}(\mathbb{R}), P^X)$  o  $h : \mathbb{R} \rightarrow [0, +\infty]$ ,  $\mathbf{E}(h(X)) = \int_{\mathbb{R}} hf_X dm$

Pur somigliando alla regola già vista, questo teorema richiederebbe una nuova dimostrazione, perché  $m$  non è una misura di probabilità. La regola già vista permetteva di calcolare il valore atteso non con un integrale in  $d\mathbf{P}$  ma con uno in  $dP^X$ ; questa permette di calcolarlo non in  $dP^X$  ma in  $dm$ , mediante il cambio di variabile  $dP^X = f_X dm$  (già visto con l'uguaglianza  $\int_A d\mathbf{P} = \int_A f_X dm$ ).

(1) significa che  $\int_{\mathbb{R}} |h| f_X dm < +\infty$ , poiché  $hf_X \in L^1(m) \iff |hf_X| \in L^1(m)$ , ma  $f_X \geq 0$ . In (2), per definizione  $\mathbf{E}(h(X)) = \int_{\Omega} h(X(\omega)) d\mathbf{P}(\omega)$ ; per la regola del valore atteso  $\mathbf{E}(h(X)) = \int_{\mathbb{R}} h(x) dP^X(x)$ . Inoltre  $\int_{\mathbb{R}} hf_X dm = \int_{\mathbb{R}} h(x) f_X(x) dx$ .

Il teorema con  $h(x) = x$  implica che se  $X$  è una v.a. assolutamente continua, allora  $\mathbf{E}(X) = \int_{\mathbb{R}} xf_X dm = \int_{\mathbb{R}} X(\omega) f(\omega) d\omega$  e  $X \in L^1(\Omega, \mathcal{A}, \mathbf{P}) \iff \int_{\mathbb{R}} |x| f_X(x) dx < +\infty$ , mentre  $var(X) = \int_{\mathbb{R}} (x - \mu)^2 f_X(x) dx$ .

**Prop (calcolo del valore atteso con la funzione di ripartizione)**

Hp:  $X$  è una v. a. quasi certamente non negativa e assolutamente

continua con densità  $f$  e funzione di ripartizione  $F$

Ts:  $\mathbf{E}(X) = \int_0^{+\infty} (1 - F(x)) dx$

**Dim**  $\int_0^{+\infty} (1 - F(x)) dx = \int_0^{+\infty} \mathbf{P}(X \geq x) dx = \int_0^{+\infty} \left( \int_x^{+\infty} f(t) dt \right) dx$ . Questo integrale doppio è fatto su un dominio normale rispetto all'asse  $x$  (la sezione del primo quadrante che sta sopra la bisettrice) che può essere visto

come normale rispetto all'asse  $t$ : per cui  $\int_0^{+\infty} \left( \int_x^{+\infty} f(t) dt \right) dx = \int_0^{+\infty} \left( \int_0^t f(t) dx \right) dt = \int_0^{+\infty} t f(t) dt = \mathbf{E}(X)$ .

■

Dato  $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathbf{P})$  con  $\mathbf{P}$  assolutamente continua con densità  $f$ , dato  $A \in \mathcal{B}(\mathbb{R}) : m(A) = 0$ , allora  $\mathbf{P}(A) = 0$ . Infatti  $\mathbf{P}(A) = \mathbf{E}(I_A) = \int_{\Omega} I_A(\omega) d\mathbf{P}(\omega) = \int_{\Omega} I_A(\omega) f(\omega) dm = 0$  perché  $I_A(\omega) f(\omega)$  è 0 quasi ovunque (in realtà questa è proprio la def di misura assolutamente continua).

1.  $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathbf{P})$  con  $\mathbf{P}$  che segue la distribuzione esponenziale  $\varepsilon(\lambda)$ ,  $\lambda > 0$ .  $X(\omega) = \omega^2$ : calcolo  $\mathbf{E}(X)$ .

Abbiamo già visto che in tal caso  $P^X$  ha densità  $f_X(x) = \frac{\lambda e^{-\lambda\sqrt{x}}}{2\sqrt{x}} I_{(0,+\infty)}(x)$ . Dunque per 11.7 applicata a  $P^X$  assolutamente continua  $\mathbf{E}(X) = \int_{\mathbb{R}} x f_X(x) dx = \int_0^{+\infty} \sqrt{x} \frac{\lambda e^{-\lambda\sqrt{x}}}{2} dx = \lambda \int_0^{+\infty} t^2 e^{-\lambda t} dt = \frac{2}{\lambda^2}$ . Posso anche utilizzare  $\mathbf{E}(X) = \int_{\Omega} X(\omega) d\mathbf{P}(\omega) = \int_{\mathbb{R}} X(\omega) f(\omega) d\omega$  perché anche  $\mathbf{P}$  è assolutamente continua con densità  $f$ : si ottiene allora  $\int_0^{+\infty} \omega^2 \lambda e^{-\lambda\omega} d\omega = \frac{2}{\lambda^2}$ .

2.  $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathbf{P})$ . Data  $X : \mathbb{R} \rightarrow \mathbb{R}$  v. a.,  $X$  si dice variabile aleatoria uniforme in  $(a, b)$ , e si scrive  $X \sim \text{unif}((a, b))$ , se ha densità  $f_X(x) = \frac{1}{b-a} I_{[a,b]}(x)$ : calcolo  $\mathbf{E}(X) = \int_a^b \frac{x}{b-a} dx = \frac{b+a}{2}$ , mentre  $\text{var}(X) = \mathbf{E}\left((X - \mu)^2\right) = \int_a^b (x - \mu)^2 \frac{1}{b-a} dx = \frac{(b-a)^2}{12}$ , usando 11.7 con  $h(x) = \left(x - \frac{b+a}{2}\right)^2$ .

3.  $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathbf{P})$ . Data  $X : \mathbb{R} \rightarrow \mathbb{R}$  v. a.,  $X$  si dice variabile aleatoria di Cauchy di parametro  $\alpha$ , e si scrive  $X \sim \text{cauchy}(\alpha)$ , se ha densità  $f_X(x) = \frac{1}{\pi(1+(x-\alpha)^2)}$  (simmetrica rispetto alla retta verticale  $x = \alpha$ ), dove  $\frac{1}{\pi}$  è una costante di normalizzazione.  $X \sim \text{cauchy}(0)$  è detta  $t$  di Student a un grado di libertà. La funzione di ripartizione di  $X$  è  $F_X(x) = \int_{-\infty}^x \frac{1}{\pi(1+(t-\alpha)^2)} dt = \frac{1}{\pi} [\arctan(t-\alpha)]_{-\infty}^x = \frac{1}{\pi} \arctan(x-\alpha) + \frac{1}{2}$ . Per quanto dimostrato in precedenza sulla simmetria (se  $\mathbf{E}(X) \in \mathbb{R}$  e  $f_X$  è simmetrica rispetto a  $x = \alpha$ ,  $\mathbf{E}(X) = \alpha$  e coincide con la mediana),  $\alpha$  è la mediana della legge di  $X$ , cioè tale che  $\mathbf{P}(X \leq \alpha) = \frac{1}{2}$ ; è evidente anche dal fatto che  $P^X((-\infty, \alpha]) = F_X(\alpha) = \frac{1}{2}$ .  $\mathbf{E}(X) = \frac{1}{\pi} \int_{-\infty}^{+\infty} x \frac{1}{1+(x-\alpha)^2} dx$ :  $\int \frac{x}{1+(x-\alpha)^2} dx = \frac{1}{2} \int \frac{2x-2\alpha}{1+(x-\alpha)^2} dx + \int \frac{\alpha}{1+(x-\alpha)^2} dx = \frac{1}{2} \ln(1+(x-\alpha)^2) + \alpha \arctan(x-\alpha)$ , quindi  $\mathbf{E}(X)$  non esiste.

! Per verificarlo con  $\alpha = 0$ , posso alternativamente calcolare  $\mathbf{E}(|X|) = \int_{-\infty}^{+\infty} \frac{|x|}{\pi(x^2+1)} dx = \int_{-\infty}^0 \frac{-x}{\pi(x^2+1)} dx + \int_0^{+\infty} \frac{x}{\pi(x^2+1)} dx = +\infty$ :  $X \notin L^1$ , ma potrebbe essere  $\mathbf{E}(X) = +\infty$ . Uso la definizione di valore atteso:  $\mathbf{E}(X_+) = \mathbf{E}(X I_{(0,+\infty)}(X))$ , cioè, con  $h(x) = x I_{(0,+\infty)}(x)$ ,  $\int_{-\infty}^{+\infty} x I_{(0,+\infty)}(x) \frac{1}{\pi(x^2+1)} dx = \int_0^{+\infty} \frac{x}{\pi(x^2+1)} dx = +\infty$  e per simmetria  $\mathbf{E}(X_-) = +\infty$ , dunque  $X$  non ammette valore atteso, pur essendo quasi certamente finita.

Dato che  $X \notin L^1$ , sicuramente  $X \notin L^2$ .  $\mathbf{E}(X^2) = \int_{-\infty}^{+\infty} x^2 \frac{1}{\pi(1+(x-\alpha)^2)} dx = +\infty$  perché per  $x \rightarrow +\infty$  la funzione integranda è asintotica a  $k$ , quindi non è integrabile in senso improprio per  $x \rightarrow +\infty$ , e per simmetria neanche per  $x \rightarrow -\infty$ . (Oppure: integrale di una funzione nonnegativa, se non è finito è  $+\infty$ ).

4. Dato  $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathbf{P})$ ,  $X : \mathbb{R} \rightarrow \mathbb{R}$  si dice variabile aleatoria esponenziale di parametro  $\lambda > 0$ , e si scrive  $X \sim \varepsilon(\lambda)$ , se  $X$  è assolutamente continua con densità  $f_X(x) = \lambda e^{-\lambda x} I_{[0, +\infty)}(x)$ . La sua funzione di ripartizione è  $F_X(t) = (1 - e^{-\lambda t}) I_{[0, +\infty)}(t)$ .  $\mathbf{E}(X) = \int_0^{+\infty} \lambda x e^{-\lambda x} dx = [-x e^{-\lambda x}]_0^{+\infty} + \int_0^{+\infty} e^{-\lambda x} dx = \frac{1}{\lambda}$ .  $\mathbf{E}(X^2) = \int_0^{+\infty} \lambda x^2 e^{-\lambda x} dx = [-x^2 e^{-\lambda x}]_0^{+\infty} + \int_0^{+\infty} 2x e^{-\lambda x} dx = \frac{2}{\lambda^2}$ , quindi  $\text{var}(X) = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}$ .

Il quantile inferiore di ordine  $\alpha$ , cioè  $t_\alpha : P^X((-\infty, t_\alpha]) = F^X(t_\alpha) = \alpha$ , si trova imponendo che  $1 - e^{-\lambda t_\alpha} = \alpha \iff t_\alpha = \frac{\ln(\alpha-1)}{\lambda}$ . La moda di  $X$  è  $\arg \max_{[0, +\infty)} f_X(x) = 0$ .

Ogni v. a. esponenziale gode di una proprietà detta assenza di memoria: significa che  $\mathbf{P}(X > s + t | X > t) = \mathbf{P}(X > s)$ , cioè, se e. g.  $X$  descrive la vita di un oggetto, non tiene conto dell'usura. Infatti  $\mathbf{P}(X > s + t | X > t) = \frac{\mathbf{P}(X > s+t, X > t)}{\mathbf{P}(X > t)} = \frac{\mathbf{P}(X > s+t)}{\mathbf{P}(X > t)} = \frac{1 - F_X(s+t)}{1 - F_X(t)} = \frac{e^{-\lambda s} e^{-\lambda t}}{e^{-\lambda t}} = e^{-\lambda s} = \mathbf{P}(X > s)$ .

Viceversa, se  $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathbf{P})$  è uno spazio di probabilità e  $X : \mathbb{R} \rightarrow \mathbb{R}$  è una v. a. assolutamente continua con supporto in  $(0, +\infty)$  e tale che  $\mathbf{P}(X > s + t | X > t) = \mathbf{P}(X > s) \forall s, t > 0$ , allora  $X \sim \varepsilon(\lambda)$ , con  $\lambda = -\ln(\mathbf{P}(X > 1))$ . Si richiede infatti che  $\frac{\mathbf{P}(X > s+t)}{\mathbf{P}(X > t)} = \mathbf{P}(X > s)$ , cioè, posto  $G(x) = \mathbf{P}(X > x)$ , che  $G(s + t) = G(s)G(t)$ . Vale evidentemente  $G(2t) = G^2(t)$ ,  $G(3t) = G^3(t)$ , ...,  $G(nt) = G^n(t)$ . Analogamente per numeri razionali  $G(\frac{m}{n}t) = G^{\frac{m}{n}}(t)$ , e quindi per i numeri reali  $G(xt) = G^x(t) \forall x$ . Nel caso particolare di  $t = 1$  si ha  $G(x) = G^x(1)$ : vale  $G(x) = (G(1))^x = e^{x \ln G(1)}$ . Si può allora porre, dato che  $G(1) = \mathbf{P}(X > 1) \in [0, 1]$ ,  $\ln G(1) = -\lambda$ .

5.  $\int_0^{+\infty} x^{z-1} e^{-x} dx$  è un numero reale se  $1 - z < 1$ , cioè  $z > 0$ : è in tal caso ben definita la funzione  $\Gamma : (0, +\infty) \rightarrow \mathbb{R}$ ,  $\Gamma(z) = \int_0^{+\infty} x^{z-1} e^{-x} dx$ .

Vale  $\Gamma(z + 1) = \int_0^{+\infty} x^z e^{-x} dx = [-x^z e^{-x}]_0^{+\infty} + \int_0^{+\infty} z x^{z-1} e^{-x} dx = z \int_0^{+\infty} x^{z-1} e^{-x} dx = z \Gamma(z)$ . Inoltre  $\Gamma(1) = 1$ : da questo si deduce che se  $z \in \mathbb{N}$   $\Gamma(z) = \int_0^{+\infty} x^{z-1} e^{-x} dx = (z-1)!$ . [Si può verificare direttamente integrando per parti:  $\int x^{z-1} e^{-x} dx = -x^{z-1} e^{-x} + \int (z-1) x^{z-2} e^{-x} dx$ ; analogamente  $\int x^{z-2} e^{-x} dx = -x^{z-2} e^{-x} + \int (z-2) x^{z-3} e^{-x} dx$ . Iterando il procedimento si ottiene che, se  $z$  è naturale,  $\int x^{z-1} e^{-x} dx = -x^{z-1} e^{-x} - (z-1) x^{z-2} e^{-x} - (z-1)(z-2) x^{z-3} e^{-x} - \dots - (z-1)(z-2) \dots 1 e^{-x}$ , con  $z$  addendi. Quindi  $\Gamma(z) = \int_0^{+\infty} x^{z-1} e^{-x} dx = (z-1)!$ , perché tutti gli altri termini si annullano.] Si può anche dimostrare che  $\Gamma(\frac{1}{2}) = \sqrt{\pi}$ . La somma di gamma di parametri  $\alpha, \lambda$  è una gamma (dim con la f caratteristica)

Dato  $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathbf{P})$ ,  $X : \mathbb{R} \rightarrow \mathbb{R}$  si dice variabile aleatoria gamma, e si scrive  $X \sim \Gamma(\alpha, \lambda)$ , se  $X$  è assolutamente continua con densità  $f_X(x) = c x^{\alpha-1} e^{-\lambda x} I_{(0, +\infty)}(x)$ , con costante di normalizzazione  $c = \frac{\lambda^\alpha}{\Gamma(\alpha)}$  e  $\alpha, \lambda > 0$ .

$X \sim \Gamma(1, \lambda)$  è una v. a. esponenziale di parametro  $\lambda$ ,  $X \sim \Gamma(\frac{n}{2}, \frac{1}{2})$  è detta variabile aleatoria chi-quadro a



$n \in \mathbb{N}$  gradi di libertà.  $f_X(x) = cx^{\alpha-1}e^{-\lambda x}I_{(0,+\infty)}(x)$  è effettivamente una densità perché è nonnegativa e  $\int_0^{+\infty} x^{\alpha-1}e^{-\lambda x}dx = \frac{\Gamma(\alpha)}{\lambda^\alpha} \forall \alpha$ .

Per trovare la moda di  $f$ , se  $\alpha \geq 1$ , si calcola  $\frac{df_X(x)}{dx} = (\alpha-1)x^{\alpha-2}e^{-\lambda x} - \lambda x^{\alpha-1}e^{-\lambda x} = x^{\alpha-2}e^{-\lambda x}(\alpha-1-\lambda x) = 0 \iff x = \frac{\alpha-1}{\lambda}$ : la moda è  $\arg \max_{(0,+\infty)} f_X(x) = \frac{\alpha-1}{\lambda}$ . Se invece  $\alpha < 1$   $f_X$  è decrescente in  $(0,+\infty)$  e ha massimo in 0, che non costituisce una moda.

In generale  $\mathbf{E}(X) = \int_0^{+\infty} \frac{\lambda^\alpha}{\Gamma(\alpha)} x^\alpha e^{-\lambda x} dx$ : lo riscrivo per far saltare fuori la densità di una v. a.  $X \sim \Gamma(\alpha+1, \lambda)$ , ottenendo quindi  $\frac{\Gamma(\alpha+1)}{\Gamma(\alpha)} \frac{1}{\lambda} \int_0^{+\infty} \frac{\lambda^{\alpha+1}}{\Gamma(\alpha+1)} x^\alpha e^{-\lambda x} dx = \frac{\Gamma(\alpha+1)}{\lambda \Gamma(\alpha)} = \frac{\alpha}{\lambda}$ . In generale, dato  $\beta : \alpha + \beta > 0$  (per poter calcolare  $\Gamma(\alpha + \beta)$ ),  $\mathbf{E}(X^\beta) = \int_0^{+\infty} \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1+\beta} e^{-\lambda x} dx = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)} \frac{1}{\lambda^\beta} \int_0^{+\infty} \frac{\lambda^{\alpha+\beta}}{\Gamma(\alpha+\beta)} x^{\alpha+\beta-1} e^{-\lambda x} dx = \frac{\Gamma(\alpha+\beta)}{\lambda^\beta \Gamma(\alpha)}$ .

Invece, posto  $\mu = \frac{\alpha}{\lambda}$ ,  $\text{var}(X) = \mathbf{E}(X^2) - \mu^2$ . Per quanto visto sopra  $\mathbf{E}(X^2) = \frac{\Gamma(\alpha+2)}{\lambda^2 \Gamma(\alpha)} = \frac{\alpha(\alpha+1)}{\lambda^2}$ , quindi  $\text{var}(X) = \frac{\alpha(\alpha+1)}{\lambda^2} - \frac{\alpha^2}{\lambda^2} = \frac{\alpha}{\lambda^2}$ . Per  $\alpha = 1$  si ottengono valore atteso e varianza dell'esponenziale.

Se  $c > 0$ ,  $Y = cX \sim \Gamma(\alpha, \frac{\lambda}{c})$ . Infatti  $\mathbf{P}(Y \leq t) = \mathbf{P}(X \leq \frac{t}{c}) = F_X(\frac{t}{c}) = \int_0^{\frac{t}{c}} \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} dx$ . Con il cambio di variabile  $w = xc$  si ottiene  $\int_0^{\frac{t}{c}} \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} dx = \int_0^t \frac{1}{\Gamma(\alpha)} (\frac{\lambda}{c})^\alpha w^{\alpha-1} e^{-\frac{\lambda}{c} w} dw$ , cioè  $F_Y(t) = \int_0^t f_Y(x) dx$  con  $f_Y(x) = \frac{1}{\Gamma(\alpha)} (\frac{\lambda}{c})^\alpha x^{\alpha-1} e^{-\frac{\lambda}{c} x}$ :  $Y$  è una gamma di parametri  $\alpha$  e  $\frac{\lambda}{c}$ .

Se  $\alpha = n \in \mathbb{N}$ , la funzione di ripartizione è  $F(t) = \left(1 - \sum_{k=0}^{n-1} e^{-\lambda t} \frac{(\lambda t)^k}{k!}\right) I_{(0,+\infty)}(t)$ . Infatti  $\mathbf{P}(X \leq t) = \int_{-\infty}^t \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} dx = \frac{\lambda^\alpha}{(\alpha-1)!} \int_{-\infty}^t x^{\alpha-1} e^{-\lambda x} dx$ .  $\int x^{\alpha-1} e^{-\lambda x} dx = -\frac{1}{\lambda} x^{\alpha-1} e^{-\lambda x} + \frac{1}{\lambda} \int (\alpha-1) x^{\alpha-2} e^{-\lambda x} dx$ ; analogamente  $\int x^{\alpha-2} e^{-\lambda x} dx = -\frac{1}{\lambda} x^{\alpha-2} e^{-\lambda x} + \frac{1}{\lambda} \int (\alpha-2) x^{\alpha-3} e^{-\lambda x} dx$ . Iterando il procedimento si ottiene che, dato che  $\alpha \in \mathbb{N}$ ,  $\int x^{\alpha-1} e^{-\lambda x} dx = -\frac{1}{\lambda} x^{\alpha-1} e^{-\lambda x} - \frac{\alpha-1}{\lambda^2} x^{\alpha-2} e^{-\lambda x} - \frac{(\alpha-1)(\alpha-2)}{\lambda^3} x^{\alpha-3} e^{-\lambda x} - \dots - \frac{(\alpha-1)(\alpha-2)\dots 1}{\lambda^\alpha} e^{-\lambda x}$ , con  $\alpha$  addendi. Quindi  $\mathbf{P}(X \geq t) = \frac{\lambda^\alpha}{(\alpha-1)!} \int_t^{+\infty} x^{\alpha-1} e^{-\lambda x} dx = \frac{\lambda^\alpha}{(\alpha-1)!} \left[ \sum_{k=0}^{\alpha-1} e^{-\lambda t} \frac{(\alpha-1)!}{\lambda^{\alpha-k}} \frac{t^k}{k!} \right] = \sum_{k=0}^{\alpha-1} e^{-\lambda t} \frac{(\lambda t)^k}{k!}$  e  $\mathbf{P}(X \leq t) = \left(1 - \sum_{k=0}^{\alpha-1} e^{-\lambda t} \frac{(\lambda t)^k}{k!}\right) I_{(0,+\infty)}(t)$ .

6. Dato  $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathbf{P})$ ,  $X : \mathbb{R} \rightarrow \mathbb{R}$  si dice variabile aleatoria normale di media  $\mu$  e varianza  $\sigma^2$ , e si scrive  $X \sim \mathcal{N}(\mu, \sigma^2)$ , se  $X$  è assolutamente continua con densità  $f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$ .

Con la sostituzione  $t = x - \mu$ ,  $\mathbf{E}(X) = \int_{-\infty}^{+\infty} x \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2} dx = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{+\infty} (t + \mu) e^{-\frac{t^2}{2\sigma^2}} dt$ :  $\int_{-\infty}^{+\infty} t e^{-\frac{t^2}{2\sigma^2}} dt = 0$  perché  $t e^{-\frac{t^2}{2\sigma^2}}$  è una funzione dispari integrabile in  $\mathbb{R}$ , mentre  $\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{t^2}{2\sigma^2}} dt = 1$  perché è l'integrale in  $\mathbb{R}$  di una densità di probabilità con  $\mu = 0$ . Dunque  $\mathbf{E}(X) = \mu$ . Con la sostituzione  $x - \mu = t$   $\text{var}(X) = \int_{-\infty}^{+\infty} (x - \mu)^2 \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2} dx = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{+\infty} t^2 e^{-\frac{t^2}{2\sigma^2}} dt$ .  $\text{var}(X) = \frac{2}{\sqrt{2\pi\sigma^2}} \int_0^{+\infty} t^2 e^{-\frac{t^2}{2\sigma^2}} dt$ .

Con la sostituzione  $t^2 = z$  si ha  $\int_0^{+\infty} t^2 e^{-\frac{t^2}{2\sigma^2}} dt = \int_0^{+\infty} \frac{1}{2} \sqrt{z} e^{-\frac{1}{2\sigma^2} z} dz$ . Usando la densità gamma,  $\int_0^{+\infty} \sqrt{z} e^{-\frac{1}{2\sigma^2} z} dz = \Gamma\left(\frac{3}{2}\right) (2\sigma^2)^{\frac{3}{2}} \int_0^{+\infty} \frac{(\frac{1}{2\sigma^2})^{\frac{3}{2}}}{\Gamma(\frac{3}{2})} \sqrt{z} e^{-\frac{1}{2\sigma^2} z} dz = \Gamma\left(\frac{3}{2}\right) (2\sigma^2)^{\frac{3}{2}}$ . Quindi  $\text{var}(X) = \frac{2}{\sqrt{2\pi\sigma^2}} \frac{1}{2} \Gamma\left(\frac{3}{2}\right) (2\sigma^2)^{\frac{3}{2}} = \frac{1}{\sqrt{2\pi}} \frac{\sqrt{\pi}}{2} 2\sqrt{2}\sigma^2 = \sigma^2$ .

Considero  $Y = |X|$ . Poiché  $F_X$ , se  $X \sim \mathcal{N}(\mu, \sigma^2)$ , è  $C^\infty$ , anche  $Y$  è assolutamente continua, avendo funzione di ripartizione  $F_Y(t) = (F_X(t) - F_X(-t))I_{(0,+\infty)}(t)$ , e quindi  $f_Y(t) = (f_X(t) + f_X(-t))I_{(0,+\infty)}(t) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{t^2+\mu^2}{2\sigma^2}} \left( e^{\frac{\mu t}{\sigma^2}} + e^{-\frac{\mu t}{\sigma^2}} \right) I_{(0,+\infty)}(t)$ .  $\mathbf{E}(Y) = \int_0^{+\infty} t f_X(t) dt + \int_0^{+\infty} t f_X(-t) dt = \frac{1}{\sqrt{2\pi\sigma^2}} \left( \int_0^{+\infty} t e^{-\frac{1}{2\sigma^2}(t-\mu)^2} dt + \int_0^{+\infty} t e^{-\frac{1}{2\sigma^2}(t+\mu)^2} dt \right)$  ??

7. Dato  $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathbf{P})$ ,  $Y : \mathbb{R} \rightarrow \mathbb{R}$  si dice variabile aleatoria lognormale di media  $\mu$  e varianza  $\sigma^2$ , e si scrive  $Y \sim \log \mathcal{N}(\mu, \sigma^2)$ , se  $Y = e^X$  con  $X \sim \mathcal{N}(\mu, \sigma^2)$ .

$F_Y(t) = \mathbf{P}(Y \leq t) = \mathbf{P}(X \leq \ln t) = \int_{-\infty}^{\ln t} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2} dx$ , cioè, con il cambio di variabile  $z = e^x$ ,  $\int_0^t \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(\ln z - \mu)^2} \frac{1}{z} dz$ , cioè  $Y$  è assolutamente continua con densità  $f_Y(z) = \frac{1}{z\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(\ln z - \mu)^2} I_{(0,+\infty)}(z)$ .

$\mathbf{E}(Y^k) = \mathbf{E}(e^{kX}) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{+\infty} e^{kx} e^{-\frac{1}{2\sigma^2}(x-\mu)^2} dx$ : con la sostituzione  $\frac{x-\mu}{\sigma} = t$ ,  $e^{k\mu} \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{k\sigma t} e^{-\frac{1}{2}t^2} dt$ .

Scrivo  $-\frac{1}{2}t^2 + k\sigma t = -\frac{1}{2}(t - k\sigma)^2 + \frac{1}{2}k^2\sigma^2$ : allora ottengo  $e^{k\mu} \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(t-k\sigma)^2} e^{\frac{1}{2}k^2\sigma^2} dt = e^{\frac{1}{2}k^2\sigma^2 + k\mu}$ , perché all'interno c'è l'integrale della densità di una normale di valore atteso  $k\sigma$  e varianza 1.

Quindi  $\text{var}(Y) = \mathbf{E}(Y^2) - \mathbf{E}^2(Y) = e^{2\sigma^2+2\mu} - e^{\sigma^2+2\mu} = e^{2\mu+\sigma^2} (e^{\sigma^2} - 1)$ .

8. Dato  $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathbf{P})$  e  $T : \mathbb{R} \rightarrow \mathbb{R}$  variabile aleatoria, si definisce intensità di guasto di  $T$   $h(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t < T < t + \Delta t | T > t)}{\Delta t} = \lim_{\Delta t \rightarrow 0} \frac{P(t < T < t + \Delta t)}{\Delta t P(T > t)} = \lim_{\Delta t \rightarrow 0} \frac{F_T(t + \Delta t) - F_T(t)}{\Delta t} \frac{1}{1 - F_T(t)} = \frac{F'_T(t)}{1 - F_T(t)}$ . Se inoltre  $T$  è assolutamente continua,  $h(t) = \frac{f_T(t)}{1 - F_T(t)}$ . L'intensità di guasto, se  $T$  rappresenta la durata di un macchinario, esprime quanto è imminente un guasto, dato che  $T$  è arrivata al valore  $t$ . Si può dimostrare che  $h$  caratterizza  $F_T$ , cioè  $F_T(t) = 1 - e^{-\int_0^t h(s) ds}$ . Se  $T \sim \varepsilon(\lambda)$ ,  $h(t) = \frac{\lambda e^{-\lambda t}}{e^{-\lambda t}} = \lambda$  è costante, dato che l'esponenziale gode dell'assenza di memoria.

9. Dato  $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathbf{P})$ ,  $X : \mathbb{R} \rightarrow \mathbb{R}$  si dice variabile aleatoria Weibull di parametri  $\alpha, \lambda$ , e si scrive  $X \sim weib(\alpha, \lambda)$  se la sua legge ha funzione di ripartizione  $F_X(t) = (1 - e^{-\lambda t^\alpha}) I_{[0,+\infty)}(t)$ , o equivalentemente densità  $f_X(t) = \lambda \alpha t^{\alpha-1} e^{-\lambda t^\alpha} I_{(0,+\infty)}(t)$ .  $X \sim weib(1, \lambda)$  è un'esponenziale. Calcolo  $\mathbf{E}(X) = \lambda \alpha \int_0^{+\infty} t^\alpha e^{-\lambda t^\alpha} dt$  con il cambio di variabile  $t^\alpha = x$ : si ottiene  $\lambda \alpha \int_0^{+\infty} \frac{1}{\alpha} x^{\frac{1}{\alpha}-1} x e^{-\lambda x} dx = \lambda \int_0^{+\infty} x^{\frac{1}{\alpha}} e^{-\lambda x} dx = \frac{\Gamma(\frac{1}{\alpha}+1)}{\lambda^{\frac{1}{\alpha}}} \int_0^{+\infty} \frac{\lambda^{\frac{1}{\alpha}+1}}{\Gamma(\frac{1}{\alpha}+1)} x^{\frac{1}{\alpha}} e^{-\lambda x} dx = \frac{\Gamma(\frac{1}{\alpha}+1)}{\lambda^{\frac{1}{\alpha}}} = \frac{\Gamma(\frac{1}{\alpha})}{\alpha \lambda^{\frac{1}{\alpha}}}$ . L'intensità di guasto di  $T$  è, supponendo  $t > 0$ ,  $h(t) = \frac{\lambda \alpha t^{\alpha-1} e^{-\lambda t^\alpha}}{e^{-\lambda t^\alpha}} = \lambda \alpha t^{\alpha-1}$ . Se  $\alpha > 1$   $h(t)$  è crescente in  $t$ , se  $\alpha \in (0, 1)$  è decrescente: un guasto è più probabile all'inizio.

10. Dato  $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathbf{P})$ ,  $X : \mathbb{R} \rightarrow \mathbb{R}$  si dice variabile aleatoria beta di parametri  $\alpha, \beta$ , e si scrive  $X \sim \beta(\alpha, \beta)$ , se è assolutamente continua con densità  $f_X(t) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} u^{\alpha-1} (1-u)^{\beta-1} I_{(0,1)}(u)$ .

## 9 Variabili aleatorie indipendenti e spazi prodotto

**Def 12.1** Dato  $(\Omega, \mathcal{A}, \mathbf{P})$  spazio di probabilità,  $(E, \mathcal{E}), (F, \mathcal{F})$  spazi misurabili,  $X : \Omega \rightarrow E, Y : \Omega \rightarrow F$  variabili aleatorie, si dice che  $X$  e  $Y$  sono variabili aleatorie indipendenti se  $\forall A \in \mathcal{E}, \forall B \in \mathcal{F}$  vale  $\mathbf{P}((X \in A) \cap (Y \in B)) = \mathbf{P}(X \in A) \mathbf{P}(Y \in B)$ .

Si scrive spesso  $\mathbf{P}((X \in A) \cap (Y \in B)) = \mathbf{P}(X \in A, Y \in B)$ . Si noti che  $X, Y$  devono avere lo stesso dominio. La definizione significa che gli tutti gli eventi generati da  $X$ , cioè  $X^{-1}(A)$  al variare di  $A \in \mathcal{E}$ , sono indipendenti da tutti gli eventi generati da  $Y$ , cioè  $Y^{-1}(B)$  al variare di  $B \in \mathcal{F}$ .

1. Lancio due volte una moneta:  $\Omega = \{t, c\}^2$ ,  $\mathcal{A} = 2^\Omega$ . Definisco  $C_1 = \{\omega \in \Omega : \omega_1 = c\} = \{(c, c), (c, t)\}$ ,  $C_2 = \{\omega \in \Omega : \omega_2 = c\} = \{(c, c), (t, c)\}$ ,  $T_1 = C_1^c$ ,  $T_2 = C_2^c$ . Essendo  $\{T_1, C_1\}, \{T_2, C_2\}$  una partizione di  $\Omega$ , per assegnare  $\mathbf{P}$  posso stabilire  $\mathbf{P}(T_1), \mathbf{P}(T_2)$  decidendo che i due lanci sono indipendenti, per 4.1, oppure posso assegnare  $\mathbf{P}$  su  $\mathcal{C} = \{T_1, T_2, T_1 \cap T_2\}$ , dato che  $\sigma(\mathcal{C}) = 2^\Omega$ , per 5.2. Stabilisco  $\mathbf{P}(T_1) = \mathbf{P}(T_2) = p$  e  $\mathbf{P}(T_1 \cap T_2) = \mathbf{P}(T_1) \mathbf{P}(T_2)$ , cioè la moneta non è necessariamente equa: quindi  $\mathbf{P}(C_1) = 1 - p$ . Definisco  $X(\omega) = \begin{cases} 1 & \text{se } \omega_1 = c \\ 0 & \text{altrimenti} \end{cases}$ ,  $Y(\omega) = \begin{cases} 1 & \text{se } \omega_2 = c \\ 0 & \text{altrimenti} \end{cases}$  :  $(X = 1) = C_1, (X = 0) = T_1, (Y = 1) = C_2, (Y = 0) =$

$T_2$ . Mi chiedo se  $X \perp Y$ , con  $\mathcal{E} = \mathcal{F} = \mathcal{B}(\mathbb{R})$ . Ma  $(X \in A)$  può essere solo  $\begin{cases} \Omega \\ \emptyset \\ (X = 0) \\ (X = 1) \end{cases} \quad \forall A \in \mathcal{E}$ , e

analogamente  $(Y \in B)$  solo  $\begin{cases} \Omega \\ \emptyset \\ (Y = 0) \\ (Y = 1) \end{cases}$ . Se almeno uno tra  $X^{-1}(A), Y^{-1}(B)$  è  $\Omega$  o  $\emptyset$ , vale l'indipendenza tra eventi. Se  $A = \{0\}, B = \{0\}$ ,  $\mathbf{P}(X = 0, Y = 0) = \mathbf{P}(T_1 \cap T_2) = \mathbf{P}(T_1) \mathbf{P}(T_2) = \mathbf{P}(X = 0) \mathbf{P}(Y = 0)$ .

Analogamente  $\mathbf{P}(X = 1, Y = 1) = \mathbf{P}(T_1^c \cap T_2^c) = \mathbf{P}(C_1 \cap C_2) = \mathbf{P}(C_1) \mathbf{P}(C_2)$  perché  $T_1 \perp T_2 \iff T_1^c \perp T_2^c$ , e poiché  $T_1 \perp T_2 \iff T_1 \perp T_2^c \iff T_1^c \perp T_2^c$ , vale anche  $\mathbf{P}(X = 0, Y = 1) = \mathbf{P}(X = 0) \mathbf{P}(Y = 1)$  e  $\mathbf{P}(X = 1, Y = 0) = \mathbf{P}(X = 1) \mathbf{P}(Y = 0)$ . Dunque  $X, Y$  sono indipendenti, e per verificarlo è sufficiente controllare che  $\mathbf{P}(X = 0, Y = 0) = \mathbf{P}(X = 0) \mathbf{P}(Y = 0)$ .

Siano  $Z = X + Y, W = X - Y, V = |X - Y|$ . Mi chiedo se  $Z \perp W, Z \perp V$ . Mostro che se  $V = h(Z)$ , non è possibile  $V \perp Z$ .

2. Considero un'urna con  $N \geq 2$  palline nere e  $B \geq 2$  bianche, e per  $k = 1, 2$   $X_k(\omega) = \begin{cases} 1 & \text{se alla } k\text{-esima estrazione esce } N \\ 0 & \text{altrimenti} \end{cases}$ .

Estraggo con reimmissione: allora  $X_1 \perp X_2$ , perché  $\mathbf{P}(X_1 = 1, X_2 = 1) = \frac{N}{N+B} \cdot \frac{N}{N+B} = \mathbf{P}(X_1 = 1)\mathbf{P}(X_2 = 1)$ ,

quindi  $X_1 \perp X_2$ , come sopra. Se invece estraggo senza reimmissione,  $\mathbf{P}(X_1 = 1, X_2 = 1) = \frac{N}{N+B} \cdot \frac{N}{N+B-1} \neq$

$\mathbf{P}(X_1 = 1)\mathbf{P}(X_2 = 1)$ .

3. Siano  $X_1, \dots, X_n$  v. a. indipendenti,  $X_M = \max\{X_1, \dots, X_n\}$ ,  $X_m = \min\{X_1, \dots, X_n\}$ .  $F_{X_M}(t) = \mathbf{P}(\max\{X_1, \dots, X_n\} \leq t) = \mathbf{P}(X_1 \leq t, \dots, X_n \leq t) = F_{\mathbf{X}}((t, t, \dots, t))$ . Usando l'ipotesi d'indipendenza,  $\mathbf{P}(X_1 \leq t, \dots, X_n \leq t) = \mathbf{P}(X_1 \leq t) \dots \mathbf{P}(X_n \leq t) = F_{X_1}(t) \dots F_{X_n}(t)$ . Se le  $X_i$  sono inoltre identicamente distribuite, si ottiene  $F_{X_M}(t) = F_{X_i}^n(t)$ .  $F_{X_m}(t) = \mathbf{P}(\min\{X_1, \dots, X_n\} \leq t) = 1 - \mathbf{P}(X_1 > t, \dots, X_n > t)$ , cioè, per indipendenza,  $1 - \mathbf{P}(X_1 > t) \dots \mathbf{P}(X_n > t) = 1 - (1 - F_{X_1}(t)) \dots (1 - F_{X_n}(t))$ . Se le  $X_i$  sono inoltre identicamente distribuite, si ottiene  $F_{X_m}(t) = 1 - (1 - F_{X_i}(t))^n$ .

Se quindi  $X_i \sim \varepsilon(\lambda_i) \forall i = 1, \dots, n$ , allora  $X_m \sim \varepsilon(\lambda_1 + \dots + \lambda_n)$ . Infatti  $F_{X_m}(t) = 1 - (1 - F_{X_1}(t)) \dots (1 - F_{X_n}(t)) = 1 - e^{-(\lambda_1 + \dots + \lambda_n)t}$ , che è quindi assolutamente continua.

La definizione data è abbastanza difficile da usare.

### Teo 12.2 (condizioni equivalenti di indipendenza)

Hp:  $(\Omega, \mathcal{A}, \mathbf{P})$  spazio di probabilità,  $(E, \mathcal{E}), (F, \mathcal{F})$  spazi misurabili,  $X : \Omega \rightarrow E, Y : \Omega \rightarrow F$  v. a.

Ts: (1)  $X \perp Y \iff (2) \forall h : (E, \mathcal{E}) \rightarrow (E', \mathcal{E}'), \forall g : (F, \mathcal{F}) \rightarrow (F', \mathcal{F}')$  misurabili

$h(X) \perp g(Y) \iff (3) \forall h, g$  nelle classi 1, 2, 3  $\mathbf{E}(h(X)g(Y)) = \mathbf{E}(h(X))\mathbf{E}(g(Y))$

dove la classe 1 è l'insieme delle  $h : E \rightarrow \mathbb{R}, g : F \rightarrow \mathbb{R}$  misurabili con  $h, g \geq 0$ ; la classe 2 è l'insieme delle  $h : E \rightarrow \mathbb{R}, g : F \rightarrow \mathbb{R}$  misurabili e limitate; la classe 3 è l'insieme delle  $h : E \rightarrow \mathbb{R}, g : F \rightarrow \mathbb{R}$  continue e limitate. Le tre classi sono insiemi di funzioni per cui sono ben definiti i valori attesi nella tesi. (3) implica che se  $X \perp Y$ ,  $X = \tilde{X}$  q. c. e  $Y = \tilde{Y}$  q. c., allora  $\tilde{X} \perp \tilde{Y}$ : infatti, se  $\forall h, g$  nelle classi 1, 2, 3  $\mathbf{E}(h(X)g(Y)) = \mathbf{E}(h(X))\mathbf{E}(g(Y))$ , vale anche  $\mathbf{E}(h(\tilde{X})g(\tilde{Y})) = \mathbf{E}(h(\tilde{X}))\mathbf{E}(g(\tilde{Y})) \forall h, g$  nelle classi 1, 2, 3 (il valore atteso è lo stesso per v. a. uguali q. c.), quindi  $\tilde{X} \perp \tilde{Y}$ .

Il teorema non semplifica particolarmente la verifica dell'indipendenza, ma permette di dedurre proprietà complicate a partire dall'indipendenza. La terza classe è contenuta nella seconda: si usa la terza classe per verificare l'indipendenza, la seconda conoscendo l'indipendenza e per dimostrare altre cose.

**Dim\*** Dimostro che (1) implica (2). Se  $\forall C \in \mathcal{E}, \forall D \in \mathcal{F}$  vale  $\mathbf{P}((X \in C) \cap (Y \in D)) = \mathbf{P}(X \in C)\mathbf{P}(Y \in D)$ , allora è anche vero che  $\forall A \in \mathcal{E}', \forall B \in \mathcal{F}'$  vale  $\mathbf{P}((h(X) \in A) \cap (g(Y) \in B)) = \mathbf{P}(h(X) \in A)\mathbf{P}(g(Y) \in B)$ . In-

fatti  $(h(X) \in A) \cap (g(Y) \in B) = (h(X))^{-1}(A) \cap (g(Y))^{-1}(B) = X^{-1}(h^{-1}(A)) \cap Y^{-1}(g^{-1}(B))$ , ma  $h^{-1}(A) \in \mathcal{E}$ ,  $g^{-1}(B) \in \mathcal{F}$  perché  $h, g$  sono misurabili, quindi si può porre  $h^{-1}(A) = C, g^{-1}(B) = D$ , e vale  $\mathbf{P}((X \in C) \cap (Y \in D)) = \mathbf{P}(X \in C) \mathbf{P}(Y \in D)$ , cioè  $\mathbf{P}((h(X) \in A) \cap (g(Y) \in B)) = \mathbf{P}(h(X) \in A) \mathbf{P}(g(Y) \in B)$ . Vedi FT per 2-3 ■

E' noto che  $X, Y \in L^1(\Omega, \mathcal{A}, \mathbf{P}) \iff \mathbf{E}(|X|) < +\infty, \mathbf{E}(|Y|) < +\infty$ .

**Corollario (il prodotto di v. a. integrabili indipendenti è integrabile)**

Hp:  $(\Omega, \mathcal{A}, \mathbf{P})$  spazio di probabilità,  $X, Y : \Omega \rightarrow \mathbb{R}$

v. a.,  $X, Y \in L^1(\Omega, \mathcal{A}, \mathbf{P})$  e  $X \perp Y$

Ts:  $XY \in L^1(\Omega, \mathcal{A}, \mathbf{P})$

( $X, Y$  devono avere lo stesso codominio) Era già noto che  $X, Y \in L^2(\Omega, \mathcal{A}, \mathbf{P})$  implica  $XY \in L^1(\Omega, \mathcal{A}, \mathbf{P})$ : l'indipendenza permette di richiedere meno regolarità su  $X, Y$ .

**Dim** E' noto che  $XY \in L^1(\Omega, \mathcal{A}, \mathbf{P}) \iff |XY| \in L^1(\Omega, \mathcal{A}, \mathbf{P})$ . Ma  $|XY| = |X||Y|$ : si applica allora 12.2, cioè la prima proposizione,  $X \perp Y$ , implica la terza con  $h(x) = |x|, g(y) = |y|$  (appartengono alla classe 1), per cui  $\mathbf{E}(|X||Y|) = \mathbf{E}(|X|) \mathbf{E}(|Y|)$ . Il lato destro è finito perché  $X, Y \in L^1(\Omega, \mathcal{A}, \mathbf{P})$ , quindi anche  $\mathbf{E}(XY) < +\infty$ . ■

**Corollario 12.3 (il valore atteso si fattorizza per v. a. indipendenti)**

Hp:  $(\Omega, \mathcal{A}, \mathbf{P})$  spazio di probabilità,

$X, Y : \Omega \rightarrow \mathbb{R}$  v. a.,  $X, Y \in L^1, X \perp Y$

Ts:  $\mathbf{E}(XY) = \mathbf{E}(X) \mathbf{E}(Y)$

Per quanto visto sopra  $\mathbf{E}(XY)$  esiste finito.

**Dim** Si è dimostrato sopra che  $XY \in L^1$ . Poiché  $h, g$  funzioni identità non rientrano in nessuna delle 3 classi sopra, per scrivere  $\mathbf{E}(XY)$  usando il teorema 12.2 occorre ricorrere a una successione di funzioni troncate. De-

finisco per  $n \geq 1$  la funzione reale  $h_n(t) = \begin{cases} -n & \text{se } t < -n \\ t & \text{se } |t| \leq n \\ n & \text{se } t > n \end{cases}$ , che è limitata  $\forall n$  ( $|h_n(t)| \leq n$ ); la successione

di funzioni  $\{h_n\}$  converge puntualmente alla funzione identità  $h(t) = t$ .  $h_n(X), h_n(Y)$  sono due successioni di funzioni che crescono a  $X, Y$  rispettivamente:  $h_n(X(\omega)) \rightarrow X(\omega) \forall \omega$ .  $\mathbf{E}(h_n(X) h_n(Y)) = \mathbf{E}(h_n(X)) \mathbf{E}(h_n(Y))$  per 12.2, essendo le  $h_n$  limitate. Cerco di usare il teorema di convergenza dominata, per poter affermare che  $\mathbf{E}(XY) = \lim_{n \rightarrow +\infty} \mathbf{E}(h_n(X) h_n(Y))$ : cerco quindi una v.a. nonnegativa e integrabile che maggiori la successione in valore assoluto  $\forall n$ . Ma  $|h_n(X) h_n(Y)| = |h_n(X)| |h_n(Y)| \leq |X| |Y| \forall n$ .  $|XY|$  è integrabile per il corollario sopra, quindi per convergenza dominata  $\mathbf{E}(XY) = \lim_{n \rightarrow +\infty} \mathbf{E}(h_n(X) h_n(Y)) = \lim_{n \rightarrow +\infty} \mathbf{E}(h_n(X)) \mathbf{E}(h_n(Y)) =$

$\mathbf{E}(X)\mathbf{E}(Y)$ , usando di nuovo la convergenza dominata al contrario nell'altro verso. ■

=====

Dato  $(\Omega, \mathcal{A}, \mathbf{P})$  e  $X : (\Omega, \mathcal{A}) \rightarrow (E, \mathcal{E}), Y : (\Omega, \mathcal{A}) \rightarrow (F, \mathcal{F}), (X(\omega), Y(\omega)) \in E \times F \forall \omega \in \Omega$ . Per poter valutare se la funzione  $(X(\omega), Y(\omega))$  è una variabile aleatoria - che, nel caso in cui abbia come componenti altre  $n$  variabili aleatorie, prende il nome di vettore aleatorio - occorre avere una  $\sigma$ -algebra su  $E \times F$ , che è naturale voler definire a partire da  $\mathcal{E}$  e  $\mathcal{F}$ .

C'è un secondo problema. Date  $X : (\Omega_1, \mathcal{A}_1) \rightarrow (E, \mathcal{E}), Y : (\Omega_2, \mathcal{A}_2) \rightarrow (F, \mathcal{F})$ , non si può usare la definizione di indipendenza tra  $X, Y$ , perché queste non sono definite sullo stesso  $\Omega$ ; può però capitare di avere comunque un pregiudizio di indipendenza. Si cerca allora di definire entrambe su uno spazio comune, per poterle confrontare: la cosa naturale è considerare  $\Omega = \Omega_1 \times \Omega_2$  e definire la v. a.  $(X_1, X_2) : \Omega \rightarrow E \times F, \begin{pmatrix} X_1(\omega_1, \omega_2) \\ X_2(\omega_1, \omega_2) \end{pmatrix} = \begin{pmatrix} X(\omega_1) \\ Y(\omega_2) \end{pmatrix}$ . Si ripresenta quindi il problema di scegliere una  $\sigma$ -algebra su  $\Omega$ .

Questa necessità prende la forma della seguente domanda: dati due spazi misurabili  $(\Omega_1, \mathcal{A}_1), (\Omega_2, \mathcal{A}_2)$ ,  $\mathcal{A}_1 \times \mathcal{A}_2 = \{A \subseteq \Omega_1 \times \Omega_2 : A = A_1 \times A_2 \text{ al variare di } A_1 \in \mathcal{A}_1, A_2 \in \mathcal{A}_2\}$  è una  $\sigma$ -algebra? La risposta è no, in generale.

1. Considero  $(\Omega_1, \mathcal{A}_1), (\Omega_2, \mathcal{A}_2)$  spazi misurabili, con  $\mathcal{A}_1 = \{\Omega_1, \emptyset\}, \mathcal{A}_2 = \{\Omega_2, \emptyset\}$ .  $\mathcal{A}_1 \times \mathcal{A}_2 = \{\Omega_1 \times \Omega_2, \emptyset\}$ : in questo caso il prodotto è una  $\sigma$ -algebra.

Il fatto che in generale  $\mathcal{A}_1 \times \mathcal{A}_2$  non sia una  $\sigma$ -algebra è evidente se si considerano  $\Omega_1 = \Omega_2 = \mathbb{R}$ , con  $\mathcal{A}_1 = \mathcal{A}_2 = \mathcal{B}(\mathbb{R})$ . Dato  $\mathcal{A}_1 \times \mathcal{A}_2$ , se si prendono  $A = A_1 \times A_2, B = B_1 \times B_2 \subseteq \mathbb{R}^2$  rettangoli non coincidenti, la loro unione non appartiene a  $\mathcal{A}_1 \times \mathcal{A}_2$  (che è formata da rettangoli o unioni di rettangoli), che quindi non è chiusa per unione finita e dunque nemmeno numerabile: perciò non è una  $\sigma$ -algebra. Questo porta alla seguente definizione.

**Def** Dati  $(\Omega_1, \mathcal{A}_1), (\Omega_2, \mathcal{A}_2)$  spazi misurabili, si definisce  $\sigma$ -algebra prodotto tra  $\mathcal{A}_1$  e  $\mathcal{A}_2$ , e si indica con  $\mathcal{A}_1 \otimes \mathcal{A}_2, \sigma(\mathcal{A}_1 \times \mathcal{A}_2)$ .

**Teo 12.5 (v. a. su spazi prodotto)**

$$(1) \mathcal{B}(\mathbb{R}^2) = \mathcal{B}(\mathbb{R}) \otimes \mathcal{B}(\mathbb{R})$$

$$(2) \text{Hp: } X : (\Omega, \mathcal{A}) \rightarrow (E, \mathcal{E}), Y : (\Omega, \mathcal{A}) \rightarrow (F, \mathcal{F}) \text{ sono v. a.}$$

$$\text{Ts: } (X, Y) : (\Omega, \mathcal{A}) \rightarrow (E \times F, \mathcal{E} \otimes \mathcal{F}) \text{ è misurabile}$$

$$(3) \text{(misurabilità delle restrizioni) Hp: } X : (\Omega_1 \times \Omega_2, \mathcal{A}_1 \otimes \mathcal{A}_2) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$$

è una variabile aleatoria misurabile

$$\text{Ts: } \forall \omega_1 \in \Omega_1 \ X : (\Omega_2, \mathcal{A}_2) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R})), X(\omega_2) = X(\omega_1, \omega_2) \text{ è}$$

$$\text{misurabile, } \forall \omega_2 \in \Omega_2 \ X : (\Omega_1, \mathcal{A}_1) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R})), X(\omega_1) = X(\omega_1, \omega_2) \text{ è}$$

misurabile

(1) significa che  $\sigma((a_1, b_1) \times (a_2, b_2) : -\infty \leq a_i < b_i \leq +\infty \forall i = 1, 2)$  coincide con la  $\sigma$ -algebra prodotto. (2) risponde al primo problema, cioè definire un vettore aleatorio a partire da v. a.  $X : (\Omega, \mathcal{A}) \rightarrow (E, \mathcal{E}), Y : (\Omega, \mathcal{A}) \rightarrow (F, \mathcal{F})$ .  $(X, Y)$  vettore aleatorio significa che  $\forall C \in \mathcal{E} \otimes \mathcal{F} (X, Y)^{-1}(C) = \{\omega : (X, Y)(\omega) \in C\} \in \mathcal{A}$ : significa che un vettore che abbia per componenti v. a. è misurabile rispetto alla  $\sigma$ -algebra prodotto. Si noti che in (3) si inverte la situazione rispetto a (2).

Dati  $X : \Omega_1 \rightarrow E$  e  $Y : \Omega_2 \rightarrow F$  in due spazi di probabilità  $(\Omega_1, \mathcal{A}_1, \mathbf{P}_1), (\Omega_2, \mathcal{A}_2, \mathbf{P}_2)$ , si è visto che per costruire un modello di probabilità congiunto si prendono  $\Omega_1 \times \Omega_2$  e  $\mathcal{A} = \mathcal{A}_1 \otimes \mathcal{A}_2$ . Come definire la probabilità  $\mathbf{P} : \mathcal{A} \rightarrow [0, 1]$ ? Essa in questo caso è detta probabilità congiunta, mentre  $\mathbf{P}_1, \mathbf{P}_2$  sono probabilità marginali.

Dato  $A_1 \in \mathcal{A}_1, A_1 \times \Omega_2 \in \mathcal{A}$ : è ragionevole chiedere che  $\mathbf{P}$  sia coerente con le probabilità marginali, cioè che  $\mathbf{P}(A_1 \times \Omega_2) = \mathbf{P}_1(A_1)$ , e analogamente che  $\mathbf{P}(\Omega_1 \times A_2) = \mathbf{P}_2(A_2) \forall A_2 \in \mathcal{A}_2$ . Richiedo inoltre che  $\mathbf{P}((A_1 \times \Omega_2) \cap (\Omega_1 \times A_2)) = \mathbf{P}(A_1 \times \Omega_2) \mathbf{P}(\Omega_1 \times A_2)$ : questo equivale a chiedere che  $\mathbf{P}(A_1 \times A_2) = \mathbf{P}_1(A_1) \mathbf{P}_2(A_2)$ , cioè che gli esperimenti descritti dai due modelli marginali siano indipendenti. Quindi la  $\mathbf{P}$  prodotto ha senso solo per esperimenti indipendenti.

**Teo 13.1 (esistenza e unicità della probabilità prodotto)**

Hp:  $(\Omega_1, \mathcal{A}_1, \mathbf{P}_1), (\Omega_2, \mathcal{A}_2, \mathbf{P}_2)$  spazi di probabilità

Ts: (1)  $\exists !$  misura di probabilità  $\mathbf{P} : \mathcal{A}_1 \otimes \mathcal{A}_2 \rightarrow [0, 1]$  :

$$\mathbf{P}(A_1 \times A_2) = \mathbf{P}_1(A_1) \mathbf{P}_2(A_2) \forall A_1 \times A_2 \in \mathcal{A}_1 \times \mathcal{A}_2$$

Tale misura di probabilità si dice misura o probabilità prodotto e si indica con  $\mathbf{P}_1 \otimes \mathbf{P}_2$ . L'uguaglianza  $\mathbf{P}(A_1 \times A_2) = \mathbf{P}_1(A_1) \mathbf{P}_2(A_2)$  implica che valgano anche  $\mathbf{P}(A_1 \times \Omega_2) = \mathbf{P}_1(A_1), \mathbf{P}(\Omega_1 \times A_2) = \mathbf{P}_2(A_2)$ .

**Dim** Per il criterio di Caratheodory, essendo  $\mathcal{C} = \{A_1 \times A_2 : A_1 \in \mathcal{A}_1, A_2 \in \mathcal{A}_2\}$  un  $\pi$ -sistema su  $\mathcal{A}_1 \times \mathcal{A}_2$  e dato che  $\sigma(\mathcal{A}_1 \times \mathcal{A}_2) = \mathcal{A}_1 \otimes \mathcal{A}_2 = \mathcal{A}$  per definizione, le richieste sopra individuano univocamente  $\mathbf{P}$ . ■

**Def** Dati  $(\Omega_1, \mathcal{A}_1, \mathbf{P}_1), (\Omega_2, \mathcal{A}_2, \mathbf{P}_2)$  spazi di probabilità, lo spazio di probabilità  $(\Omega_1 \times \Omega_2, \mathcal{A}_1 \otimes \mathcal{A}_2, \mathbf{P}_1 \otimes \mathbf{P}_2)$  è detto spazio prodotto.



**Teo 13.1 (Fubini-Tonelli per v. a. nonnegative)**

Hp:  $(\Omega_1, \mathcal{A}_1, \mathbf{P}_1), (\Omega_2, \mathcal{A}_2, \mathbf{P}_2)$  spazi di probabilità,  $X : \Omega_1 \times \Omega_2 \rightarrow [0, +\infty]$

è una v.a.  $\mathcal{A}_1 \otimes \mathcal{A}_2$  misurabile

$$\begin{aligned} \text{Ts: (1)} \quad & \int_{\Omega_1} X(\omega_1, \omega_2) d\mathbf{P}(\omega_1) \text{ è } \mathcal{A}_2\text{-misurabile, } \int_{\Omega_2} X(\omega_1, \omega_2) d\mathbf{P}(\omega_2) \text{ è } \mathcal{A}_1\text{-misurabile} \\ (2) \quad & \int_{\Omega_1 \times \Omega_2} X(\omega_1, \omega_2) d\mathbf{P}_1 \otimes \mathbf{P}_2(\omega_1, \omega_2) = \int_{\Omega_2} \left( \int_{\Omega_1} X(\omega_1, \omega_2) d\mathbf{P}_1(\omega_1) \right) d\mathbf{P}_2(\omega_2) = \\ & \int_{\Omega_1} \left( \int_{\Omega_2} X(\omega_1, \omega_2) d\mathbf{P}_2(\omega_2) \right) d\mathbf{P}_1(\omega_1) \end{aligned}$$

Quindi in realtà questo non riguarda i vettori aleatori: è una v. a. definita su uno spazio prodotto.

$\int_{\Omega_1} X(\omega_1, \omega_2) d\mathbf{P}(\omega_1)$  è una funzione che ha come dominio  $\Omega_2$  e come codominio  $\mathbb{R}$ : la tesi afferma che è una variabile aleatoria.  $X \in L^1(\Omega_1 \times \Omega_2, \mathcal{A}_1 \otimes \mathcal{A}_2, \mathbf{P}_1 \otimes \mathbf{P}_2) \iff \int_{\Omega_2} \left( \int_{\Omega_1} X(\omega_1, \omega_2) d\mathbf{P}_1(\omega_1) \right) d\mathbf{P}_2(\omega_2) < +\infty$ .

Il teorema afferma che si può calcolare un integrale doppio con due integrali semplici.

**Teo 13.1 (Fubini-Tonelli per v. a. integrabili)**

Hp:  $(\Omega_1, \mathcal{A}_1, \mathbf{P}_1), (\Omega_2, \mathcal{A}_2, \mathbf{P}_2)$  spazi di probabilità,  $X : \Omega_1 \times \Omega_2 \rightarrow \mathbb{R}$ ,

$$X \in L^1(\Omega_1 \times \Omega_2, \mathcal{A}_1 \otimes \mathcal{A}_2, \mathbf{P}_1 \otimes \mathbf{P}_2)$$

Ts: (1)  $X(\omega_2) = \int_{\Omega_1} X(\omega_1, \omega_2) d\mathbf{P}_1(\omega_1)$  è  $L^1(\Omega_2, \mathcal{A}_2, \mathbf{P}_2)$  q. c. rispetto a  $\mathbf{P}_2$ ,

$X(\omega_1) = \int_{\Omega_2} X(\omega_1, \omega_2) d\mathbf{P}_2(\omega_2)$  è  $L^1(\Omega_1, \mathcal{A}_1, \mathbf{P}_1)$  q. c. rispetto a  $\mathbf{P}_1$

$$\begin{aligned} (2) \quad & \int_{\Omega_1} X(\omega_1, \omega_2) d\mathbf{P}(\omega_1) \text{ è } \mathcal{A}_2\text{-misurabile, } \int_{\Omega_2} X(\omega_1, \omega_2) d\mathbf{P}(\omega_2) \text{ è } \mathcal{A}_1\text{-misurabile} \\ (3) \quad & \int_{\Omega_1 \times \Omega_2} X(\omega_1, \omega_2) d\mathbf{P}_1 \otimes \mathbf{P}_2(\omega_1, \omega_2) = \int_{\Omega_2} \left( \int_{\Omega_1} X(\omega_1, \omega_2) d\mathbf{P}_1(\omega_1) \right) d\mathbf{P}_2(\omega_2) = \\ & \int_{\Omega_1} \left( \int_{\Omega_2} X(\omega_1, \omega_2) d\mathbf{P}_2(\omega_2) \right) d\mathbf{P}_1(\omega_1) \end{aligned}$$

(1) manca tesi di integrabilità di  $X, Y$

**Corollario 13.2 (la legge di v. a. indipendenti si fattorizza)**

Hp:  $(\Omega, \mathcal{A}, \mathbf{P})$  spazio di probabilità,  $(E, \mathcal{E}), (F, \mathcal{F})$  spazi misurabili,  $X : \Omega \rightarrow E$ ,

$Y : \Omega \rightarrow F$  v. a.,  $(X, Y) : \Omega \rightarrow E \times F$  ha legge  $P^{(X, Y)} : \mathcal{E} \otimes \mathcal{F} \rightarrow [0, 1]$

$$\text{Ts: } X \perp Y \iff P^{(X, Y)} = P^{(X)} \otimes P^{(Y)}$$

$P^{(X, Y)}$  si è definita quando si è definita la legge di una v. a.  $X : \Omega \rightarrow E$ . Si può parlare di legge del vettore aleatorio a partire da variabili aleatorie grazie alla definizione di spazio prodotto.

La tesi significa che la legge di un vettore aleatorio, quindi congiunta, avente come componenti variabili aleatorie indipendenti si fattorizza nel prodotto delle leggi marginali. Questo conferma che la definizione della probabilità prodotto ha senso quando si considerano eventi (o variabili aleatorie) indipendenti.

**Dim** Mostro l'implicazione da sinistra a destra.  $X \perp Y \iff \forall A \in \mathcal{E}, \forall B \in \mathcal{F} \mathbf{P}(X \in A, Y \in B) = \mathbf{P}(X \in A) \mathbf{P}(Y \in B)$ . Posso scrivere l'evento  $(X \in A, Y \in B) = ((X, Y) \in A \times B)$ . Allora  $\mathbf{P}((X, Y) \in A \times B) = \mathbf{P}(X \in A) \mathbf{P}(Y \in B)$  per ipotesi e  $P^{(X,Y)}(A \times B) = \mathbf{P}(X \in A) \mathbf{P}(Y \in B) = P^X(A) P^Y(B) \forall (A, B) \in \mathcal{E} \times \mathcal{F}$ , dove  $P^X(A)$  è la misura di  $A$  sotto la legge di  $X$ . Questo, per il teorema di Fubini-Tonelli, definisce univocamente  $P^{(X,Y)}$ , con la probabilità prodotto:  $P^{(X,Y)} = P^X \otimes P^Y$ .

Mostro l'implicazione da destra a sinistra. Sia  $P^{(X,Y)} = P^{(X)} \otimes P^{(Y)}$ : allora in particolare sui rettangoli  $P^{(X,Y)}(A \times B) = P^X(A) P^Y(B) \forall (A, B) \in \mathcal{E} \times \mathcal{F}$ , per cui  $P^{(X,Y)}(A \times B) = \mathbf{P}((X, Y) \in A \times B) = P^X(A) P^Y(B) = \mathbf{P}(X \in A) \mathbf{P}(Y \in B)$ , cioè  $X, Y$  sono indipendenti. ■

Quindi, se so che  $X, Y$  sono indipendenti e ho  $P^X, P^Y$ , lo spazio di probabilità  $(E \times F, \mathcal{E} \otimes \mathcal{F}, P^{(X)} \otimes P^{(Y)})$  è un buon modello.

**Corollario (la legge prodotto è coerente con le leggi marginali)**

Hp:  $X_1 : (\Omega_1, \mathcal{A}_1, \mathbf{P}_1) \rightarrow (E, \mathcal{E}), X_2 : (\Omega_2, \mathcal{A}_2, \mathbf{P}_2) \rightarrow (F, \mathcal{F})$  sono v.a.,

$$\Omega = \Omega_1 \times \Omega_2, \mathcal{A} = \mathcal{A}_1 \otimes \mathcal{A}_2, \mathbf{P} = \mathbf{P}_1 \otimes \mathbf{P}_2, (X, Y) : \Omega \rightarrow E \times F,$$

$$X(\omega_1, \omega_2) = X_1(\omega_1), Y(\omega_1, \omega_2) = X_2(\omega_2)$$

Ts: (1) date  $X : \Omega \rightarrow E, X_1 : \Omega_1 \rightarrow E$ , vale  $P^X = P^{X_1}$ ; date  $Y : \Omega \rightarrow F$ ,

$$X_2 : \Omega_2 \rightarrow F, \text{ vale } P^Y = P^{X_2}$$

$$(2) X \perp Y$$

La tesi 1 è  $\mathbf{P}(X \in B) = \mathbf{P}_1(X_1 \in B), \mathbf{P}(Y \in B) = \mathbf{P}_2(X_2 \in B)$ .

Nella tesi 2,  $X, Y$  possono essere confrontate perché sono definite sullo stesso  $\Omega$ . Il significato è che se mi aspetto che due v. a. siano indipendenti, posso costruire un modello di probabilità che me lo garantisce; in generale, se un vettore aleatorio è definito su uno spazio prodotto a partire da due variabili aleatorie definite su spazi diversi, allora queste sono indipendenti sullo spazio prodotto.

**Dim** (1)  $P^X(B_1) = \mathbf{P}(X \in B_1) \forall B_1 \in \mathcal{E}_1. (X \in B_1) = \{\omega \in \Omega : X(\omega_1, \omega_2) = X_1(\omega_1) \in B_1\} = (X_1 \in B_1) \times \Omega_2$ . Allora, per definizione di probabilità prodotto,  $\mathbf{P}(X \in B_1) = \mathbf{P}_1(X_1 \in B_1) \mathbf{P}_2(\Omega_2) = \mathbf{P}_1(X_1 \in B_1) = P^{X_1}(B_1)$ .

(2) Devo mostrare che  $\forall A \in \mathcal{E}, \forall B \in \mathcal{F} \mathbf{P}(X \in A, Y \in B) = \mathbf{P}(X \in A) \mathbf{P}(Y \in B)$ . Vale  $\mathbf{P}(X \in A, Y \in B) = \mathbf{P}((X, Y) \in A \times B) = \mathbf{P}((X \in A) \times (Y \in B)) = \mathbf{P}(X \in A) \mathbf{P}(Y \in B)$  per definizione di probabilità prodotto: quindi si ha la tesi. ■

1. Lancio 2 dadi.  $\Omega_1 = \{1, \dots, 6\}, \mathcal{A}_1 = 2^{\Omega_1}, \mathbf{P}_1$  uniforme su  $\Omega_1$  è il modello di probabilità per il primo esperimento, analogamente per il secondo.  $X_1$  descrive il risultato del primo lancio:  $X_1 : \Omega_1 \rightarrow \mathbb{R}, X_1(\omega_1) = \omega_1$ , e analogamente  $X_2(\omega_2) = \omega_2$ . E' naturale aspettarsi che i due lanci siano indipendenti: questo è vero per il corollario. Definisco  $\Omega = \{1, \dots, 6\}^2, \mathcal{A} = 2^{\Omega_1} \otimes 2^{\Omega_2}$  (che in questo caso coincide con  $2^{\Omega_1 \times \Omega_2}$ ),  $\mathbf{P} = \mathbf{P}_1 \otimes \mathbf{P}_2$  probabilità uniforme su  $\Omega$ ,  $X(\omega) = X_1(\omega_1), Y(\omega) = X_2(\omega_2)$ : allora  $X \perp Y$ .

### Famiglie di variabili aleatorie indipendenti

**Def 13.3** Dato  $(\Omega, \mathcal{A}, \mathbf{P})$  spazio di probabilità,  $(E_k, \mathcal{E}_k)$  spazio misurabile  $\forall k = 1, \dots, n$ ,  $X_k : \Omega \rightarrow E_k$  variabile aleatoria  $\forall k = 1, \dots, n$ , si dice che  $X_1, \dots, X_n$  sono una famiglia di variabili aleatorie indipendenti se  $\mathbf{P}\left(\bigcap_{j=1}^n (X_j \in B_j)\right) = \prod_{j=1}^n \mathbf{P}(X_j \in B_j) \forall B_j \in \mathcal{E}_j \forall j = 1, \dots, n$ .

Se  $n = 2$ , si ottiene la 12.1. Si può scrivere equivalentemente  $\mathbf{P}(X_1 \in B_1, X_2 \in B_2, \dots, X_n \in B_n) = \prod_{j=1}^n \mathbf{P}(X_j \in B_j) \forall (B_1, \dots, B_n) \in (\mathcal{E}_1 \times \mathcal{E}_2 \times \dots \times \mathcal{E}_n) \forall j = 1, \dots, n$ . A differenza della definizione data per famiglie di eventi indipendenti, non ci sono  $2^n - n - 1$  condizioni da confrontare: tuttavia, la definizione implica la richiesta di indipendenza su tutte le sottofamiglie; se e. g. si prende  $B_k = E_k$  per un certo  $\bar{k}$ , si ottiene  $\mathbf{P}(X_{\bar{k}} \in B_{\bar{k}}) = 1$  e si ritrova la definizione per  $n - 1$  variabili aleatorie. L'indipendenza permette di lavorare sui rettangoli, singolarmente su ogni spazio, cioè di separare le variabili, fattorizzare.

Dalla definizione segue che se  $X_1, \dots, X_n$  sono indipendenti e  $g$  è una funzione misurabile allora  $g(X_1), \dots, g(X_n)$  sono una famiglia di variabili aleatorie indipendenti: infatti  $\mathbf{P}\left(\bigcap_{j=1}^n (g(X_j) \in B_j)\right) = \mathbf{P}\left(\bigcap_{j=1}^n (X_j \in g^{-1}(B_j))\right) = \prod_{j=1}^n \mathbf{P}(X_j \in g^{-1}(B_j))$ .

1. Considero  $X_1, X_2, X_3$  variabili aleatorie. La condizione di indipendenza è che  $\mathbf{P}((X_1 \in B_1) \cap (X_2 \in B_2) \cap (X_3 \in B_3)) = \mathbf{P}(X_1 \in B_1) \mathbf{P}(X_2 \in B_2) \mathbf{P}(X_3 \in B_3) \forall (B_1, B_2, B_3) \in (\mathcal{E}_1 \times \mathcal{E}_2 \times \mathcal{E}_3)$ . Se si pone  $B_1 = E_1$ ,  $\mathbf{P}(X_1 \in B_1) = 1$  e si ottiene la richiesta di indipendenza tra due variabili aleatorie.

### Prop 13.4 (indipendenza di sottofamiglie disgiunte)

Hp:  $(\Omega, \mathcal{A}, \mathbf{P})$  spazio di probabilità,  $(E_k, \mathcal{E}_k)$  spazio misurabile e  $X_k : \Omega \rightarrow E_k$  v. a.

$\forall k = 1, \dots, n$ ,  $X_1, \dots, X_n$  v. a. indipendenti,  $J_1, J_2 \subseteq \{1, \dots, n\} : J_1 \cap J_2 = \emptyset$

Ts:  $(X_i)_{i \in J_1} \perp (X_j)_{j \in J_2}$

Se in particolare  $|J_1| = |J_2| = 2$ , la proposizione afferma che  $(X_h, X_k) \perp (X_i, X_j)$  se  $\{h, k\} \neq \{i, j\}$ . La tesi significa che i vettori  $(X_i), (X_j)$  sono indipendenti come variabili aleatorie: vedi esempio. Non vale l'implicazione inversa.

1. Considero  $X$  uniforme su  $\{-1, 1\}$  con  $\mathbf{P}(X = -1) = \frac{1}{2}, P(X = 1) = \frac{1}{2}$ , e analogamente  $Y$  uniforme su  $\{-1, 1\}$  con  $\mathbf{P}(Y = -1) = \frac{1}{2}, P(Y = 1) = \frac{1}{2}$ .  $X \perp Y$  (si è già verificato in un caso simile). Definisco  $Z = XY$ , tale che  $\mathbf{P}(Z = -1) = \frac{1}{2}, P(Z = 1) = \frac{1}{2}$ .  $Z \perp X$  perché  $\mathbf{P}(Z = 1, X = 1) = \frac{1}{4}$ , e  $ImZ = ImX = \{-1, 1\}$ . Allo stesso modo  $Z \perp Y$ . Tuttavia  $Z$  e  $(X, Y)$  non sono indipendenti perché il valore di  $(X, Y)$  individua univocamente  $Z$ . Quindi, pur essendo indipendenti a coppie, non sono una famiglia di v. a. indipendenti, usando l'implicazione inversa in 13.4.

**Def 14.1** Dato  $(\Omega, \mathcal{A}, \mathbf{P})$  spazio di probabilità,  $(E_k, \mathcal{E}_k)$  spazio misurabile  $\forall k \geq 1$ ,  $X_k : \Omega \rightarrow E_k$  variabile aleatoria  $\forall k \geq 1$ , si dice che la successione  $(X_n)_{n \geq 1}$  è una famiglia di variabili aleatorie indipendenti se qualsiasi sottofamiglia finita di  $(X_n)_{n \geq 1}$  è una famiglia di v.a. indipendenti.

Si mostra che  $(X_n)_{n \geq 1}$  è una famiglia di variabili aleatorie indipendenti  $\iff X_1, \dots, X_k$  è una famiglia di v.a. indipendenti  $\forall k \geq 2$ : non occorre controllare qualsiasi famiglia, ma solo le prime  $k$ .

Si può allora estendere il corollario 13.2:  $\mathbf{X}$  ha componenti indipendenti  $\iff P^{\mathbf{X}} = P^{X_1} \otimes \dots \otimes P^{X_n}$ .

**Def** Dato  $(\Omega, \mathcal{A}, \mathbf{P})$  spazio di probabilità,  $(E, \mathcal{E})$  spazio misurabile  $\forall k = 1, \dots, n$ ,  $X_k : \Omega \rightarrow E$  variabile aleatoria  $\forall k = 1, \dots, n$ , si dice che  $X_1, \dots, X_n$  sono indipendenti e identicamente distribuite se sono una famiglia di variabili aleatorie indipendenti e  $P^{X_1}(B) = \dots = P^{X_k}(B) \forall B \in \mathcal{E}$ .

**Def** Data  $(X_n)_{n \geq 1}$  famiglia di v.a. indipendenti definite su  $(\Omega, \mathcal{A}, \mathbf{P})$  spazio di probabilità, si indica con  $\mathcal{B}_n = \sigma(X_n)$ ,  $\mathcal{C}_n = \sigma\left(\bigcup_{p=n}^{+\infty} \mathcal{B}_p\right)$ ; si dice  $\sigma$ -algebra di coda della famiglia  $(X_n)_{n \geq 1}$ , e si indica con  $\mathcal{C}_\infty, \bigcap_{n=1}^{+\infty} \mathcal{C}_n$ .

Intuitivamente, una proprietà della successione che dipende dal suo comportamento "di coda", ad esempio la convergenza puntuale della successione, dev'essere un evento della sua  $\sigma$ -algebra di coda; essa contiene gli eventi il cui verificarsi o meno non è alterato dalla modifica di un numero finito di elementi della successione.

1.  $\{\omega : \sup_n X_n(\omega) < 1\} \notin \mathcal{C}_\infty$  perché l'estremo superiore della successione, fissato  $\omega$ , dipende da tutti i suoi termini, che può quindi essere modificato alterando un termine.
2.  $\{\omega : \lim_{n \rightarrow +\infty} X_n(\omega) \text{ esiste}\} \in \mathcal{C}_\infty$  perché riguarda il comportamento di coda della successione.
3.  $\{\omega : \limsup_n X_n(\omega) < 1\} \in \mathcal{C}_\infty$  perché il limsup è un concetto che si basa sui limiti delle sottosuccessioni di  $X_n$  e quindi sul comportamento di coda della successione. Questo vale per qualsiasi evento generato da  $\limsup_n X_n$ .

### Teo (legge zero-uno di Kolmogorov)

Hp:  $(X_n)_{n \geq 1}$  famiglia di v.a. indipendenti definite su  $(\Omega, \mathcal{A}, \mathbf{P})$

spazio di probabilità,  $\mathcal{C}_\infty$  loro  $\sigma$ -algebra di coda,  $C \in \mathcal{C}_\infty$

Ts:  $\mathbf{P}(C) = 0$  oppure  $\mathbf{P}(C) = 1$

Il teorema afferma che ogni evento della  $\sigma$ -algebra di coda ha probabilità zero o uno.

Ne segue che ogni v. a.  $\mathcal{C}_\infty$ -misurabile, e. g.  $\limsup_n X_n$  e  $\liminf_n X_n$ , è q. c. costante: infatti ogni evento generato da essa avrà probabilità zero o uno, in particolare gli eventi del tipo  $X_n = a$ .

## 10 Vettori aleatori

Sia  $\mathbf{X} = (X_1, \dots, X_n) : \Omega \rightarrow \mathbb{R}^n$  una n-upla di variabili aleatorie reali, detta vettore aleatorio, con  $(\Omega, \mathcal{A}, \mathbf{P})$  spazio di probabilità. E' noto che:

1. Come  $\sigma$ -algebra su  $\mathbb{R}^n$  si prende  $\mathcal{B}(\mathbb{R}^n) = \mathcal{B}(\mathbb{R}) \otimes \dots \otimes \mathcal{B}(\mathbb{R})$  ( $n$  fattori).
2.  $\mathbf{X}$  è misurabile rispetto alle 2 ( $n$ ?)  $\sigma$ -algre  $\iff X_1, \dots, X_n$  sono misurabili separatamente.
3. Se  $h : \mathbb{R}^n \rightarrow \mathbb{R}^k$  è boreliana,  $h(\mathbf{X}) : \Omega \rightarrow \mathbb{R}^k$  è un vettore aleatorio
4.  $\mathbf{X} = \mathbf{Y}$  q. c.  $\iff X_k = Y_k$  q. c.  $\forall k = 1, \dots, n$
5.  $P^{\mathbf{X}}(B) = \mathbf{P}(X \in B)$  (niente di nuovo: la definizione di legge si è data per  $E$  generico), che dipende solo da  $[X]$
6.  $\mathbf{X}$  ha componenti indipendenti  $\iff P^{\mathbf{X}} = P^{X_1} \otimes \dots \otimes P^{X_n}$
7. Se  $X_1, \dots, X_n \in L^p \iff \mathbf{X} \in L^p$  (come per le curve)
8. Date  $X_1, X_2 \in L^2(\mathbf{P})$ , allora  $X_1, X_2 \in L^1(\mathbf{P})$ . Se  $X_1, X_2 \in L^1(\mathbf{P})$  e  $X_1 \perp X_2$ ,  $X_1 X_2 \in L^1(\mathbf{P})$ .

**Def** Dato un vettore aleatorio  $\mathbf{X} : \Omega \rightarrow \mathbb{R}^n$  con legge  $P^{\mathbf{X}} : \mathcal{B}(\mathbb{R}^n) \rightarrow [0, 1]$ , si definisce funzione di ripartizione della sua legge  $F_{\mathbf{X}} : \mathbb{R}^n \rightarrow [0, 1]$ ,  $F_{\mathbf{X}}(\mathbf{x}) = P^{\mathbf{X}}((-\infty, x_1] \times \dots \times (-\infty, x_n]) = \mathbf{P}(X_1 \leq x_1, \dots, X_n \leq x_n)$ .

$P^{\mathbf{X}}((-\infty, x_1] \times \dots \times (-\infty, x_n])$  è la misura di un iperquadrante di sud-ovest sotto la legge di  $X$ . La definizione di  $F_{\mathbf{X}}$  non ha gli stessi vantaggi della definizione di una funzione di ripartizione da  $\mathbb{R}$  in  $[0, 1]$ : e. g., non si può parlare di monotonia perché in  $\mathbb{R}^n$  non è definibile una relazione d'ordine totale.

Dato  $\mathbf{X}$ ,  $X_1, \dots, X_n$  sono indipendenti  $\iff F_{\mathbf{X}}(\mathbf{x}) = F_{X_1}(x_1) \dots F_{X_n}(x_n)$ : questa proprietà permette una semplice verifica dell'indipendenza, anche quando alcune componenti del vettore sono discrete e altre continue. Significa che l'insieme degli iperquadranti di sud-ovest è un  $\pi$ -sistema.

Se  $F_{\mathbf{X}}$  è derivabile tranne e. g. un numero finito di punti,  $f_{\mathbf{X}}(\mathbf{x}) = \frac{\partial^2 F_{\mathbf{X}}(\mathbf{x})}{\partial x_1 \dots \partial x_n}$ .

## 10.1 Vettori aleatori discreti

**Def 14.2** Dato  $(\Omega, \mathcal{A}, \mathbf{P})$  spazio di probabilità, un vettore aleatorio  $\mathbf{X} : \Omega \rightarrow \mathbb{R}^n$  con legge  $P^{\mathbf{X}} : \mathcal{B}(\mathbb{R}^n) \rightarrow [0, 1]$  si dice vettore aleatorio discreto se  $P^{\mathbf{X}}$  è una misura di probabilità discreta su  $\mathcal{B}(\mathbb{R}^n)$ , cioè se  $\exists T \subseteq \mathbb{R}^n$  discreto e  $p_{\mathbf{X}} : T \rightarrow [0, 1]$ , detta densità discreta o densità congiunta di  $\mathbf{X}$ , tale che  $P^{\mathbf{X}}(B) = \sum_{\mathbf{x} \in T \cap B} p_{\mathbf{X}}(\mathbf{x}) \forall B \in \mathcal{B}(\mathbb{R}^n)$ . In tal caso si dice supporto di  $\mathbf{X}$ , e si indica con  $S_{\mathbf{X}}$ , il supporto di  $p_{\mathbf{X}}$ , cioè  $\{\mathbf{x} \in T : p_{\mathbf{X}}(\mathbf{x}) > 0\}$ .

Se  $\mathbf{X}$  descrive gli esiti del lancio di due dadi,  $T$  è  $\{1, \dots, 6\}^2$ , cioè un reticolo di 36 punti nel piano.

**Prop (definizioni equivalenti di vettore aleatorio discreto)**

Hp:  $(\Omega, \mathcal{A}, \mathbf{P})$  spazio di probabilità,  $\mathbf{X} : \Omega \rightarrow \mathbb{R}^n$  è vettore aleatorio con legge  $P^{\mathbf{X}}$

Ts: (1)  $\mathbf{X}$  è discreto  $\iff$  (2)  $\exists T' \subseteq \mathbb{R}^n$  discreto :  $\mathbf{P}(\mathbf{X} \in T') = 1 \iff$  (3)  $\exists \tilde{\mathbf{X}} : \mathbf{P}(\mathbf{X} = \tilde{\mathbf{X}}) = 1$

e  $Im(\tilde{\mathbf{X}}) = S_{\mathbf{X}} \iff$  (4)  $X_1, \dots, X_n$  sono variabili aleatorie discrete

(3) significa che non necessariamente  $\mathbf{X}$  ha come immagine  $S_{\mathbf{X}}$ , ma se è uguale quasi certamente con a un vettore che ha quella come immagine, allora è discreto!

**Dim** Dimostro che (4) implica (2). Se  $X_1, \dots, X_n$  sono variabili aleatorie discrete,  $P^{X_k}$  è una probabilità discreta con  $S_{X_k}$  discreto  $\forall k = 1, \dots, n$ . Allora definisco  $T' = S_{X_1} \times S_{X_2} \times \dots \times S_{X_n}$  e mostro che  $\mathbf{P}(\mathbf{X} \in T') = 1$ . Osservo che  $(\mathbf{X} \in S_{X_1} \times S_{X_2} \times \dots \times S_{X_n}) \iff (X_1 \in S_{X_1} \cap \dots \cap X_n \in S_{X_n}) \iff \bigcap_{k=1}^n (X_k \in S_{X_k})$  per la proprietà distributiva della controimmagine rispetto all'intersezione. Allora  $\mathbf{P}(\mathbf{X} \in T') = \mathbf{P}(\bigcap_{k=1}^n (X_k \in S_{X_k})) = 1 - \mathbf{P}(\bigcup_{k=1}^n (X_k \in S_{X_k})^c) \geq 1 - \sum_{k=1}^n \mathbf{P}((X_k \in S_{X_k})^c)$ . Ma  $\mathbf{P}(X_k \notin S_{X_k}) = 0 \forall k$  perché  $\mathbf{P}(X_k \notin S_{X_k}) = 1 - \mathbf{P}(X_k \in S_{X_k}) = 1 - \sum_{k=1}^n \mathbf{P}(X_k = x_k) = 1 - 1 = 0$ , dove  $x_k$  è il generico elemento di  $S_{X_k}$ , essendo  $\sum_{k=1}^n \mathbf{P}(X_k = x_k) = \sum_{k=1}^N p_{X_k}(x_k) = 1$ , quindi  $\mathbf{P}(\mathbf{X} \in T') \geq 1$ , cioè  $\mathbf{P}(\mathbf{X} \in T') = 1$ . ■

Se  $X_1, \dots, X_n$  sono variabili aleatorie discrete su  $T_1, \dots, T_n$  discreti (con densità  $p_{X_1}, \dots, p_{X_k}$ , dette densità marginali), per la proposizione  $\mathbf{P}(\mathbf{X} \in T = T_1 \times \dots \times T_n) = 1$ , ma è possibile che  $S_{\mathbf{X}} \subset T$ .

**Prop 14.3 (regola del valore atteso per vettori aleatori discreti, densità marginali da densità**

congiunta, fattorizzazione della densità)

Hp:  $(\Omega, \mathcal{A}, \mathbf{P})$  spazio di probabilità,  $\mathbf{X} : \Omega \rightarrow \mathbb{R}^n$  è vettore

aleatorio discreto con densità  $p_{\mathbf{X}}$  su  $T = T_1 \times \dots \times T_n$

Ts: (1) data  $h : \mathbb{R}^n \rightarrow \mathbb{R}$ , se  $h(\mathbf{x}) \geq 0$  o  $h \in L^1(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), P^{\mathbf{X}})$

$$\mathbf{E}(h(\mathbf{X})) = E^{\mathbf{X}}(h) = \sum_{t \in T} p_{\mathbf{X}}(t) h(t)$$

$$(2) \text{ la densità discreta di } X_k \text{ è } p_{X_k}(\hat{x}_k) = \sum_{\substack{x_1 \in T_1 \\ \vdots \\ x_{k-1} \in T_{k-1} \\ x_{k+1} \in T_{k+1} \\ \vdots \\ x_n \in T_n}} p_{\mathbf{X}}(x_1, \dots, x_{k-1}, \hat{x}_k, \dots, x_n)$$

$$\text{con } (x_1, \dots, x_{k-1}, \hat{x}_k, \dots, x_n) \in T$$

$$(3) X_1, \dots, X_n \text{ sono indipendenti} \iff p_{\mathbf{X}}(x_1, \dots, x_n) = p_{X_1}(x_1) \dots p_{X_n}(x_n) \quad \forall \mathbf{x}$$

(1) implica che  $h \in L^1(P^{\mathbf{X}}) \iff \sum_{t \in T} |h(t)| p_{\mathbf{X}}(t) < +\infty$ . (2) è ben posta perché  $\mathbf{X}$  è discreto se e solo se tutte le sue componenti sono discrete; significa che si possono ottenere le densità marginali dalla densità congiunta, cioè avendo la legge del vettore si può sapere tutto sulle componenti. (3) significa che per v. a. indipendenti la densità congiunta si fattorizza nel prodotto delle densità marginali.

**Dim** (1) vale per il teo 10.1 nel caso particolare  $(E, \mathcal{E}) = (\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ .

(2) Nel caso particolare di  $n = 2$  per semplicità di notazione. Se  $x_1 \in T_1$ ,  $p_{X_1}(x_1) = \mathbf{P}(X_1 = x_1) = \mathbf{P}(X_1 = x_1, X_2 \in \mathbb{R}) = P^{\mathbf{X}}(\{x_1\} \times \mathbb{R})$  perché  $(X_1 = \hat{x}_1 \cap X_2 \in \mathbb{R}) = X^{-1}(\hat{x}_1) \cap \Omega = X^{-1}(\hat{x}_1)$ . Dunque  $p_{X_1}(x_1) = P^{\mathbf{X}}(\{x_1\} \times \mathbb{R}) = \sum_{x_2 \in T_2} p_{\mathbf{X}}(\mathbf{x})$  per definizione di probabilità discreta.

(3) Se  $X_1, \dots, X_n$  sono indipendenti,  $p_{\mathbf{X}}(x_1, \dots, x_n) = \mathbf{P}(X_1 = x_1, \dots, X_n = x_n) = \mathbf{P}(X_1 = x_1) \dots \mathbf{P}(X_n = x_n) = p_{X_1}(x_1) \dots p_{X_n}(x_n)$  per definizione di indipendenza.

Se  $p_{\mathbf{X}}(x_1, \dots, x_n) = p_{X_1}(x_1) \dots p_{X_n}(x_n) \quad \forall \mathbf{x}, \forall B_1, \dots, B_n \in \mathcal{B}(\mathbb{R}^n)$  vale  $\mathbf{P}(X_1 \in B_1, \dots, X_n \in B_n) = \mathbf{P}(\mathbf{X} \in B_1 \times \dots \times B_n) = \sum_{\mathbf{x} \in B \cap S} p_{\mathbf{X}}(\mathbf{x})$ : scrivendo la sommatoria componente per componente si ottiene  $\sum_{\mathbf{x} \in B \cap S} p_{\mathbf{X}}(\mathbf{x}) \dots$  ■

1.  $\Omega = \{1, \dots, 6\}^3$ ,  $\mathcal{A} = 2^\Omega$ ,  $\mathbf{P}$  uniforme. Lancio 3 dadi:  $X_1(\omega) = \omega_1, X_2 = \omega_2, X_3 = \omega_3$  descrivono il risultato dei lanci. Come trovo la densità congiunta delle tre, cioè la densità del vettore  $(X_1, X_2, X_3)$ ? Se considero  $\omega = (1, 2, 1)$ ,  $p_{\mathbf{X}}(1, 2, 1) = \mathbf{P}(X_1 = 1, X_2 = 2, X_3 = 1) = \mathbf{P}(X_1 = 1) \mathbf{P}(X_2 = 2) \mathbf{P}(X_3 = 1)$ , perché le tre v. a. sono indipendenti, quindi  $p_{\mathbf{X}}(1, 2, 1) = p_{X_1}(1) p_{X_2}(2) p_{X_3}(1)$ .

$Y = X_1 + X_2$  è una v. a. discreta perché  $Im(X_1 + X_2) = \{2, \dots, 12\}$  è discreta. Costruisco il vettore  $(Y, X_1)$ , che è un vettore aleatorio discreto perché ogni componente è discreta. Ne trovo la densità su

$T = \{2, \dots, 12\} \times \{1, \dots, 6\}$ : calcolo ad esempio  $p_{(Y, X_1)}(2, 1) = \mathbf{P}(Y = 2, X_1 = 1) = \mathbf{P}(X_1 = 1, X_2 = 1) = \frac{1}{36}$  perché  $X_1$  e  $X_2$  sono indipendenti (cf. v. a. che descrivono l'esito di due lanci di moneta). Dato  $(t_1, t_2) \in T$ ,  $p_{(Y, X_1)}(t_1, t_2) = 0$  se  $t_1 \leq t_2$  o se  $t_1 \geq 8$  e  $t_2 \leq t_1 - 7$ , quindi la matrice delle densità è triangolare superiore.

$Y/X_1$	1	2	3	4	5	6	$p_Y$
2	$\frac{1}{36}$	0	0	0	0	0	$\frac{1}{36}$
3	$\frac{1}{36}$	$\frac{1}{36}$	0	0	0	0	$\frac{1}{18}$
4	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	0	0	0	$\frac{1}{12}$
5	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	0	0	$\frac{1}{9}$
6	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	0	$\frac{5}{36}$
7	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{6}$
...	...	...	...	...	...	...	...
$p_{X_1}$	$\frac{1}{6}$	$\frac{5}{36}$	$\frac{1}{9}$	$\frac{1}{12}$	$\frac{1}{18}$	$\frac{1}{36}$	1

Mostro che  $X_1$  e  $Y$  non sono indipendenti: basta trovare  $t : p_{(Y, X_1)}(t) = 0$ , infatti e. g.  $p_{(Y, X_1)}(2, 2) = 0 \neq p_Y(2)p_{X_1}(2) = \frac{1}{36} \frac{1}{6}$ . In generale, nei punti in cui la densità congiunta è 0 cade l'indipendenza tra le componenti del vettore, perché sono i punti in cui la densità marginale di  $Y$  è forzata da quella di  $X_1$ . Dato quindi che in questo caso il supporto di  $(Y, X_1)$  non è un rettangolo, non posso separare le variabili:  $S_{(Y, X_1)}$  non è fattorizzabile.

Invece  $Y \perp X_3$  perché  $Y = h(X_1, X_2) \perp g(X_3) \forall h, g$  misurabili, per 12.2, essendo  $(X_1, X_2)$  indipendente da  $X_3$ .

## 10.2 Vettori assolutamente continui

**Def 14.4** Si dice misura di Lebesgue su  $\mathcal{B}(\mathbb{R}^n)$  una funzione  $m_n : \mathcal{B}(\mathbb{R}^n) \rightarrow [0, +\infty]$  con le seguenti proprietà:

**M1**  $m_n(A_1 \times \dots \times A_n) = m(A_1) \dots m(A_n) \forall A_i \in \mathcal{B}(\mathbb{R}), \forall i = 1, \dots, n$

**M2**  $\sigma$ -additività:  $\forall (B_n)_{n \geq 1} : B_n \in \mathcal{B}(\mathbb{R}^n) \forall n$  e  $B_k \cap B_h \forall k \neq h$ , vale  $m_n\left(\bigcup_{n=1}^{+\infty} B_n\right) = \sum_{n=1}^{+\infty} m_n(B_n)$

M1 è ben posta perché  $\mathcal{B}(\mathbb{R}^n) = \mathcal{B}(\mathbb{R}) \otimes \dots \otimes \mathcal{B}(\mathbb{R})$ : dev'essere  $m_n(A_1 \times \dots \times A_n) = m(A_1) \dots m(A_n) \forall A_i \in \mathcal{B}(\mathbb{R})$ , come per **P**.  $A_1 \times \dots \times A_n$  è un generico elemento del  $\pi$ -sistema dei rettangoli in  $\mathbb{R}^n$ .

**Prop**

$\exists !$  misura di Lebesgue su  $\mathbb{R}^n$



**Prop (proprietà di  $m_n$ )**

Hp:  $m_n : \mathcal{B}(\mathbb{R}^n) \rightarrow [0, +\infty]$  è la misura di Lebesgue su  $\mathbb{R}^n$

$$\text{Ts: (1) } m_2([a, b] \times [c, d]) = (b - a)(d - c)$$

$$(2) m_3([a, b] \times [c, d] \times [e, f]) = (b - a)(d - c)(f - e)$$

$$(3) m_n(\mathbb{R}^n) = +\infty$$

$$(4) m_n(\{\mathbf{x}\}) = 0 \quad \forall \mathbf{x} \in \mathbb{R}^n$$

$$(5) m_n(\{\mathbf{x}\} \times \mathbb{R}) = 0 \quad \forall \mathbf{x} \in \mathbb{R}^{n-1}$$

$$(6) m_2(G_f) = 0 \quad \forall f \text{ boreliana}$$

$$(7) m_n(A) = 0 \quad \forall A \text{ sottospazio di } \mathbb{R}^n : \dim A < n$$

In (6)  $G_f = \{(x, y) \in \mathbb{R}^2 : y = f(x)\}$ .  $m_2(\{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq r^2\}) = \pi r^2$ : per  $n = 2$  e  $n = 3$  la misura di Lebesgue coincide con le ordinarie nozioni di area e volume. Per (7) in realtà vale  $m_n(A) = 0 \quad \forall A$ : esiste un sottospazio  $V$  di  $\mathbb{R}^n : \dim V < n$  e  $A \subseteq V$ , dove  $V$  è un sottospazio eventualmente affine ( $V = \mathbf{v}_0 + H$ ,  $\dim V = \dim H$ ).

**Dim\*** (1) (2) Sono conseguenze banali di M1 e della definizione di misura di Lebesgue in  $\mathbb{R}$ .

(3) E' ancora una conseguenza di M1, dato che  $m(\mathbb{R}) = +\infty$ .

(5) significa in particolare che  $m_2(\{x\} \times \mathbb{R}) = 0 \quad \forall x \in \mathbb{R}$ , cioè la misura di una retta verticale in  $\mathbb{R}^2$  è nulla. Si dimostra considerando un intorno tubolare della retta di ampiezza che tende a 0. ■

Si può definire, data  $h : \mathbb{R}^n \rightarrow \mathbb{R}, h \geq 0$  misurabile, l'integrale rispetto alla misura multidimensionale  $\int_{\mathbb{R}^n} h dm_n$  come fatto in precedenza. Scriveremo, con abuso,  $\int_{\mathbb{R}^n} h dm_n$  come  $\int_{\mathbb{R}^n} h(x_1, \dots, x_n) dx_1 \dots dx_n$ . Si può definire analogamente  $L^1(m_n) = \{[h] : h \text{ è misurabile e } \int_{\mathbb{R}^n} |h| dm_n < +\infty\}$ . Si definisce  $\int_A h dm_n$  come  $\int_{\mathbb{R}^n} h I_A dm_n$ ; vale quindi  $\int_A dm_n = \int_{\mathbb{R}^n} I_A dm_n = m_n(A)$ . Valgono le stesse proprietà dell'integrale di Lebesgue in  $\mathbb{R}$ , con due eccezioni:  $h$  limitata e misurabile non implica  $h$  integrabile; non è vero che  $L^q(m_n) \subseteq L^p(m_n) \quad \forall p \leq q$ , sempre per il fatto che la misura non è finita.

**Def 14.5** Dato  $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \mathbf{P})$  spazio di probabilità,  $\mathbf{P}$  si dice misura di probabilità assolutamente continua se  $\exists f : \mathbb{R}^n \rightarrow \mathbb{R}$  boreliana tale che  $\mathbf{P}(A) = \int_A f(x_1, \dots, x_n) dm_n \quad \forall A \in \mathcal{B}(\mathbb{R}^n)$ .

Si scrive con abuso  $\int_A f(x_1, \dots, x_n) dx_1 \dots dx_n$ . Se  $n = 1$ , si ottiene la definizione già data, per 11.5.

**Def 14.6** Dato  $(\Omega, \mathcal{A}, \mathbf{P})$  spazio di probabilità e  $\mathbf{X} : \Omega \rightarrow \mathbb{R}^n$  vettore aleatorio con legge  $P^{\mathbf{X}}$ ,  $X$  si dice assolutamente continuo se  $P^{\mathbf{X}}$  è una misura di probabilità assolutamente continua sullo spazio misurabile  $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ . In tal caso si indica con  $f_{\mathbf{X}}$  una sua densità.

**Teo 14.7 (caratterizzazione delle densità)**

Hp:  $(\Omega, \mathcal{A}, \mathbf{P})$  è uno spazio di probabilità,  $f : \mathbb{R}^n \rightarrow \mathbb{R}$

Ts: (1)  $f$  è una densità per  $\mathbf{P} \iff f$  è misurabile,  $f(\mathbf{x}) \geq 0 \forall \mathbf{x}$ ,  $\int_{\mathbb{R}^n} f(x_1, \dots, x_n) dm_n = 1$

(2) se vale 1,  $\mathbf{P}(A) = \int_A f(x_1, \dots, x_n) dm_n \forall A \in \mathcal{B}(\mathbb{R}^n)$

(3) se  $f_1 = f_2$  q. o. e una delle due è una densità per  $\mathbf{P}$ , anche l'altra è

una densità per  $\mathbf{P}$

(3) significa che  $\mathbf{P}$  è caratterizzata da  $[f]$ .

1. Considero  $(\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2))$  e  $C = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq r^2\}$ : voglio assegnare allo spazio una misura di probabilità uniforme su  $C$ . Questo significa intuitivamente che c'è una densità di probabilità costante su  $C$ , cioè  $f(x, y) = cI_C(x, y)$  (ne segue che si definisce  $\mathbf{P}$  assolutamente continua). Per determinare  $c \geq 0$  si usa (1):  $\int_{\mathbb{R}^2} cI_C(x, y) dx dy = cm_2(C) = c\pi r^2 = 1 \iff c = \frac{1}{r^2\pi}$ .

Dato  $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \mathbf{P})$  con  $\mathbf{P}$  assolutamente continua con densità  $f$ , dato  $A \in \mathcal{B}(\mathbb{R}^n) : m_n(A) = 0$ , allora  $\mathbf{P}(A) = 0$ .

**Teo 15.1 (regola del valore atteso per vettori aleatori continui, densità marginali da densità congiunta, fattorizzazione della densità)**

Hp:  $(\Omega, \mathcal{A}, \mathbf{P})$  spazio di probabilità,  $\mathbf{X} : \Omega \rightarrow \mathbb{R}^n$  v. a. con legge  $P^{\mathbf{X}}$

assolutamente continuo con densità  $f_{\mathbf{X}}$

Ts: (1)  $\forall h : \mathbb{R}^n \rightarrow \mathbb{R}, h \in L^1(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), P^{\mathbf{X}}) \iff hf_{\mathbf{X}} \in L^1(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), m_n);$

$\forall h \in L^1(P^{\mathbf{X}})$  o  $h \geq 0$  misurabile vale  $\mathbf{E}(h(\mathbf{X})) = \int_{\mathbb{R}^n} h(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x}) dm_n$

(2)  $\forall k = 1, \dots, n$   $X_k : (\Omega, \mathcal{A}, \mathbf{P}) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$  è assolutamente

continua con densità  $f_{X_k}(\hat{x}_k) = \int_{\mathbb{R}^{n-1}} f_{\mathbf{X}}(x_1, \dots, \hat{x}_k, \dots, x_n) dx_1 \dots dx_{k-1} dx_{k+1} \dots dx_n$

(3)  $X_1, \dots, X_n$  sono indipendenti e assolutamente continue con densità  $f_{X_1},$

$\dots, f_{X_n} \iff \mathbf{X}$  è assolutamente continuo e  $f_{\mathbf{X}}(x_1, \dots, x_n) = f_{X_1}(x_1) \dots f_{X_n}(x_n)$  q. o.

E' l'analogo di 14.3 per vettori aleatori continui.  $\int_{\mathbb{R}^n} h(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x}) dm_n = \int_{\mathbb{R}^n} h(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x}) dx_1 \dots dx_n$ .  $hf_{\mathbf{X}} \in L^1(m_n)$  significa  $\int_{\mathbb{R}^n} |h(\mathbf{x})| f_{\mathbf{X}}(\mathbf{x}) dm_n < +\infty$ . Ricordo che  $\mathbf{E}(h(\mathbf{X})) = \int_{\mathbb{R}^n} h(x_1, \dots, x_n) dP^{\mathbf{X}}(x_1, \dots, x_n)$  per la

regola del valore atteso. (2) significa che la densità della componente  $k$  si ottiene integrando la densità congiunta avendo fissato la variabile  $k$ , come per le densità discrete.

**Dim** (3) Se  $X_1, \dots, X_n$  sono indipendenti e assolutamente continue con densità  $f_{X_1}, \dots, f_{X_n}$ , allora la legge di  $\mathbf{X}$  è  $P^{\mathbf{X}}(B) = \mathbf{P}(X_1 \in B_1, \dots, X_n \in B_n)$ , cioè, per indipendenza e assoluta continuità delle componenti,  $\mathbf{P}(X_1 \in B_1) \dots \mathbf{P}(X_n \in B_n) = \int_{B_1} f_{X_1}(x_1) dx_1 \dots \int_{B_n} f_{X_n}(x_n) dx_n = \int_{B_1 \times \dots \times B_n} f_{X_1}(x_1) \dots f_{X_n}(x_n) dx_1 \dots dx_n$ , dove l'ultima uguaglianza vale per il teorema di Fubini-Tonelli, essendo l'integrale fatto rispetto a  $m_n$ , che è la misura prodotto  $m \otimes \dots \otimes m$ . ■

Relativamente a (3), può capitare che  $f_{\mathbf{X}}$  sia assegnata già fattorizzata in  $n$  fattori, cioè  $f_{\mathbf{X}}(x_1, \dots, x_n) = h_1(x_1) \dots h_n(x_n)$ : si sta affermando implicitamente che le componenti sono indipendenti; in tal caso i fattori sono le densità marginali, a meno di una costante da determinare imponendo che  $\int_{\mathbb{R}} f_{X_k}(x_k) dx_k = 1 \forall k$ .

Ovviamente  $f_{\mathbf{X}}$  è assegnata già fattorizzata se questo è possibile: se e. g.  $f_{(X,Y)}(x,y) = \frac{1}{\pi} I_C(x,y)$ , con  $C = \{(x,y) \in \mathbb{R}^2 : x^2 + y^2 \leq 1\}$ ,  $\frac{1}{\pi} I_C(x,y) = \frac{1}{\pi} I_{[-1,1]}(x) I_{[-\sqrt{1-x^2}, \sqrt{1-x^2}]}(y)$ : ma la seconda  $I$  dipende anche da  $x$ , quindi  $f$  non è fattorizzabile. Lo sarebbe passando in coordinate polari, ma con una tale trasformazione si cambierebbe la legge del vettore.

1. Se  $n = 2$  e ho  $\mathbf{X} = (X_1, X_2)$  assolutamente continuo, la seconda densità marginale è  $f_{X_2}(\hat{x}_2) = \int_{\mathbb{R}} f_{\mathbf{X}}(x_1, \hat{x}_2) dx_1$ .

Se  $\mathbf{X}$  è assolutamente continuo, ogni componente  $X_k$  è assolutamente continua per 15.1 (2); in generale non è vero il viceversa, a differenza del caso discreto, ma lo diventa se le componenti sono indipendenti.

1. Considero  $(\Omega, \mathcal{A}, \mathbf{P})$  e  $X : \Omega \rightarrow \mathbb{R}$  con legge esponenziale  $\varepsilon(\lambda)$ : è assolutamente continua con densità  $f_X(t) = \lambda e^{-\lambda t} I_{(0,+\infty)}(t)$ . Si è già visto che  $X^2$  è assolutamente continua con densità  $f_{X^2}(t) = \frac{\lambda}{2\sqrt{t}} e^{-\lambda\sqrt{t}} I_{(0,+\infty)}(t)$ . Mi chiedo se il vettore  $\mathbf{X} = (X, X^2)$  è assolutamente continuo. Osservo che  $\forall \omega \in \Omega (X(\omega), X^2(\omega)) \in G_f$ , con  $f(x) = x^2$ , cioè  $\mathbf{P}(\mathbf{X} \in G_f) = 1$  (il boreliano  $G_f$  ha misura 1 sotto la legge di  $\mathbf{X}$ ), ma si è visto sopra che se  $P^{\mathbf{X}}$  è assolutamente continua  $m_2(G_f) = 0$ , quindi  $P^{\mathbf{X}}(G_f) = 0$ , che è assurdo. Dunque  $P^{\mathbf{X}}$  non è assolutamente continua (vedi def 2.2.13)

### Teo 15.2 (l'indipendenza a blocchi preserva l'assoluta continuità)

Hp:  $(\Omega, \mathcal{A}, \mathbf{P})$  spazio di probabilità,  $\mathbf{X}_1 : \Omega \rightarrow \mathbb{R}^{n_1}$ ,  $\mathbf{X}_2 : \Omega \rightarrow \mathbb{R}^{n_2}$  sono v. a.

assolutamente continui con densità  $f_{\mathbf{X}_1}, f_{\mathbf{X}_2}$  e indipendenti;  $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2) : \Omega \rightarrow \mathbb{R}^{n_1+n_2}$

Ts:  $\mathbf{X}$  è assolutamente continuo con densità

$$f_{\mathbf{X}}(x_1, \dots, x_{n_1}, \dots, x_{n_1+n_2}) = f_{\mathbf{X}_1}(x_1, \dots, x_{n_1}) f_{\mathbf{X}_2}(x_{n_1+1}, \dots, x_{n_1+n_2})$$

Il teorema è una generalizzazione di 15.1 (3): nel caso  $n_1 = n_2 = 1$  afferma che un vettore con componenti assolutamente continue e indipendenti è assolutamente continuo, con densità che si fattorizza. L'indipendenza a blocchi permette di preservare l'assoluta continuità.

E' noto che se  $\mathbf{X} : \Omega \rightarrow \mathbb{R}^n$  è un v. a. e  $\mathbf{h} : \mathbb{R}^n \rightarrow \mathbb{R}^k$  è misurabile, allora  $\mathbf{h}(\mathbf{X}) : \Omega \rightarrow \mathbb{R}^k$  è un vettore aleatorio discreto se  $\mathbf{X}$  è discreto.

Se  $\mathbf{X}$  è assolutamente continuo,  $\mathbf{h}(\mathbf{X})$  è discreto se  $\mathbf{h}$  ha immagine discreta. Se  $\mathbf{h}$  è un diffeomorfismo ( $C^1$  con inversa  $C^1$ ) anche  $\mathbf{h}(\mathbf{X})$  è assolutamente continuo. Se  $\mathbf{h}$  non rientra in nessuna di queste categorie,  $\mathbf{h}(\mathbf{X})$  è generalmente difficile da trattare.

1. Se  $(X, Y)$  è un ve. a. assolutamente continuo con densità  $f_{(X,Y)}(x, y)$  e  $Z = X + Y$ , allora  $Z$  è una v.

a. assolutamente continua con densità  $f_Z(z) = \int_{-\infty}^{+\infty} f_{(X,Y)}(z-x, x) dy$ . Infatti  $F_Z(t) = \mathbf{P}(X+Y \leq t) = P^{(X,Y)}(A)$ , con  $A = \{(x, y) : x \in \mathbb{R}, y \in (-\infty, t-x)\}$ . Allora  $P^{(X,Y)}(A) = \int_{-\infty}^{+\infty} \left( \int_{-\infty}^{t-x} f_{(X,Y)}(x, y) dy \right) dx$ , e per il teorema fondamentale del calcolo integrale  $f_Z(t) = \frac{d}{dt} \left( \int_{-\infty}^{+\infty} \left( \int_{-\infty}^{t-x} f_{(X,Y)}(x, y) dy \right) dx \right) = \int_{-\infty}^{+\infty} \frac{d}{dt} \left( \int_{-\infty}^{t-x} f_{(X,Y)}(x, y) dy \right) dx$  perché  $\lim_{y \rightarrow -\infty} f_{(X,Y)}(x, y) = 0$ . Se  $X \perp Y$ ,  $f_Z(z) = \int_{-\infty}^{+\infty} f_X(x) f_Y(t-x) dx$ , detto prodotto di convoluzione delle due densità.

Alternativamente, si usa il vettore  $\begin{pmatrix} U \\ V \end{pmatrix} = \begin{pmatrix} X+Y \\ Y \end{pmatrix} = \mathbf{h}(X, Y) : \mathbf{g}(U, V) = \begin{pmatrix} U-V \\ V \end{pmatrix}$  ha  $J_{\mathbf{g}}(u, v) = \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix}$ , per cui  $f_{(U,V)}(u, v) = f_{(X,Y)}(u-v, v) \cdot 1$ , e la densità marginale di  $U$  è  $f_Z(u) = \int_{-\infty}^{+\infty} f_{(X,Y)}(u-v, v) dv$ .

## 10.3 Covarianza

Dato  $(\Omega, \mathcal{A}, \mathbf{P})$ , considero  $(X, Y) : \Omega \rightarrow \mathbb{R}^2$  v. a., con legge  $P^{(X,Y)} : \mathcal{B}(\mathbb{R}^2) \rightarrow [0, 1]$ . Cerco un indice che misuri la relazione tra le componenti del vettore. Ho già  $\mathbf{E}(X), \mathbf{E}(Y), \text{var}(X), \text{var}(Y)$ , per calcolare i quali non è necessario conoscere  $P^{(X,Y)}$ , e che non dicono nulla su  $X, Y$  congiuntamente.

**Def 15.3** Dato  $(\Omega, \mathcal{A}, \mathbf{P})$  spazio di probabilità e  $(X, Y) : \Omega \rightarrow \mathbb{R}^2$  v. a. tale che  $X, Y \in L^2(\Omega, \mathcal{A}, \mathbf{P})$ , si definisce covarianza di  $X$  e  $Y$ , e si indica con  $\text{cov}(X, Y)$ , il numero reale  $\mathbf{E}((X - \mathbf{E}(X))(Y - \mathbf{E}(Y)))$ . Se  $\text{cov}(X, Y) > 0$ ,  $X$  e  $Y$  si dicono positivamente correlate; se  $\text{cov}(X, Y) = 0$ ,  $X$  e  $Y$  si dicono scorrelate; se  $\text{cov}(X, Y) < 0$ ,  $X$  e  $Y$  si dicono negativamente correlate.

La definizione è ben posta: perché se  $X, Y \in L^2$ , allora  $X, Y \in L^1$ , e perché  $X - \mathbf{E}(X), Y - \mathbf{E}(Y) \in L^2$  (i numeri reali appartengono a  $L^p \forall p$  e  $L^2$  è uno spazio vettoriale) e il prodotto di v. a.  $L^2$  è  $L^1$ .

Una covarianza positiva esprime la tendenza a crescere di una v. a. quando l'altra cresce, mentre se è negativa una cresce quando l'altra decresce. L'unità di misura della varianza è l'unità di misura delle v. a. al quadrato.

Si può vedere  $cov(X, Y)$  come  $\mathbf{E}(h(X, Y))$  con  $h(X, Y) = (X - \mu_X)(Y - \mu_Y)$ , per cui  $cov(X, Y) = \int_{\mathbb{R}^2} (x - \mu_X)(y - \mu_Y) dP^{(X, Y)}$  per la regola del valore atteso.

Se  $(X, Y)$  è discreto con  $T \subseteq \mathbb{R}^2$  discreto e densità  $p_{(X, Y)}$ , si ha  $cov(X, Y) = \sum_{(x, y) \in T} (x - \mu_X)(y - \mu_Y) p_{(X, Y)}(x, y)$ ; se  $(X, Y)$  è assolutamente continuo con densità  $f_{(X, Y)}$ , si ha  $cov(X, Y) = \int_{\mathbb{R}^2} (x - \mu_X)(y - \mu_Y) f_{(X, Y)}(x, y) dx dy$ .

#### Teo 15.4 (proprietà della covarianza)

Hp:  $(\Omega, \mathcal{A}, \mathbf{P})$  spazio di probabilità,  $(X, Y) : \Omega \rightarrow \mathbb{R}^2$  v.a.,  $X, Y, Z \in L^2(\Omega, \mathcal{A}, \mathbf{P})$

$$\text{Ts: (1) } cov(X, X) = var(X)$$

$$(2) \text{ (commutatività) } cov(X, Y) = cov(Y, X)$$

$$(3) cov(X, Y) = \mathbf{E}(XY) - \mathbf{E}(X)\mathbf{E}(Y)$$

$$(4) \text{ (bilinearità) } cov(aX + bY + c, Z) = acov(X, Z) + bcov(Y, Z)$$

$$(5) \text{ se } X \perp Y, cov(X, Y) = 0$$

$$(6) \text{ (Cauchy-Schwarz) } |cov(X, Y)| \leq \sqrt{var(X)var(Y)}$$

$$(7) \text{ se } X_k \in L^2 \forall k = 1, \dots, n, var\left(\sum_{k=1}^n X_k\right) = \sum_{k=1}^n var(X_k) + 2 \sum_{h \neq k} cov(X_k, X_h)$$

$\mathbf{E}(XY) = \mathbf{E}(h(X, Y))$  con  $h(x, y) = xy$ : vedi regola valore atteso.

Le proprietà (1), (2) e (4), unitamente al fatto che  $cov(X, X) \geq 0$ , implicano che  $cov : L^2(\Omega, \mathcal{A}, \mathbf{P}) \times L^2(\Omega, \mathcal{A}, \mathbf{P}) \rightarrow \mathbb{R}$  definisce un'operazione con proprietà simili a un prodotto scalare su  $L^2(\Omega, \mathcal{A}, \mathbf{P})$ ; non vale tuttavia  $cov(X, X) = 0 \iff X = 0$ , ma  $X = c$  q. c.. E' un prodotto scalare sullo spazio  $L^2(\Omega, \mathcal{A}, \mathbf{P})$  quozientato rispetto alla relazione di equivalenza "differire quasi certamente per una costante".

La norma indotta da tale prodotto scalare è  $\|X\| = \sqrt{var(X)}$ . (3) è un'uguaglianza analoga a quella che vale per la varianza, che diventa un caso particolare di (3).

Due v. a. indipendenti hanno covarianza nulla, ma non vale il contrario. Per calcolare  $\mathbf{E}(XY)$  occorre  $P^{(X, Y)}$ . In (4) si avrebbe  $cov(c, Z) = 0$  perché la covarianza di una costante e qualsiasi v. a. è nulla. (7), se  $n = 2$ , è  $var(X + Y) = var(X) + var(Y) + 2cov(X, Y)$ . Se le  $X_k$  sono v. a. indipendenti, vale  $var(\sum_{k=1}^n X_k) = \sum_{k=1}^n var(X_k)$ , perché ogni covarianza è nulla per la (5), essendo le  $X_k$  v. a. indipendenti anche due a due.

**Dim** (1)  $cov(X, X) = \mathbf{E}((X - \mathbf{E}(X))(X - \mathbf{E}(X))) = var(X)$  per definizione.

(2)  $cov(X, Y) = \mathbf{E}((X - \mathbf{E}(X))(Y - \mathbf{E}(Y))) = \mathbf{E}((Y - \mathbf{E}(Y))(X - \mathbf{E}(X))) = cov(Y, X)$  per commutatività del prodotto in  $\mathbb{R}$ .

(3)  $cov(X, Y) = \mathbf{E}((X - \mathbf{E}(X))(Y - \mathbf{E}(Y))) = \mathbf{E}(XY - X\mathbf{E}(Y) - Y\mathbf{E}(X) + \mathbf{E}(X)\mathbf{E}(Y))$ , che è, per linearità,  $\mathbf{E}(XY) - \mathbf{E}(X)\mathbf{E}(Y) - \mathbf{E}(Y)\mathbf{E}(X) + \mathbf{E}(X)\mathbf{E}(Y) = \mathbf{E}(XY) - \mathbf{E}(X)\mathbf{E}(Y)$ .

(4) Per definizione  $cov(aX + bY + c, Z) = \mathbf{E}((aX + bY + c - \mathbf{E}(aX + bY + c))(Z - \mathbf{E}(Z)))$ . L'argomento è  $(aX + bY + c)Z - \mathbf{E}(Z)(aX + bY + c) - Z\mathbf{E}(aX + bY + c) + \mathbf{E}(aX + bY + c)\mathbf{E}(Z)$ : i termini in cui compare  $c$  sono  $cZ - c\mathbf{E}(Z) - Zc + c\mathbf{E}(Z) = 0$ , e si ottiene quindi  $(aX + bY)Z - \mathbf{E}(Z)(aX + bY) - Z\mathbf{E}(aX + bY) + \mathbf{E}(aX + bY)\mathbf{E}(Z)$ , che coincide con  $a(X - \mathbf{E}(X))(Z - \mathbf{E}(Z)) + b(Y - \mathbf{E}(Y))(Z - \mathbf{E}(Z))$ . Calcolando il valore atteso di ciò si ottiene per linearità  $\mathbf{E}(a(X - \mathbf{E}(X))(Z - \mathbf{E}(Z))) + \mathbf{E}(b(Y - \mathbf{E}(Y))(Z - \mathbf{E}(Z))) = acov(X, Z) + bcov(Y, Z)$ .

(5) Per 12.3, per v. a. indipendenti  $\mathbf{E}(XY) = \mathbf{E}(X)\mathbf{E}(Y)$ , quindi usando (3) si ha  $cov(X, Y) = 0$ .

(6) È una conseguenza della struttura di spazio euclideo di  $L^2(\Omega, \mathcal{A}, \mathbf{P})$ : ponendo  $cov(X, Y) = \langle X, Y \rangle$ , si ha per la disuguaglianza di Cauchy-Schwarz  $|\langle X, Y \rangle| \leq \|X\| \|Y\| = \sqrt{var(X) var(Y)}$ .

(7) Per bilinearità del prodotto scalare si ottiene  $var(\sum_{k=1}^n X_k) = \langle X_1 + \dots + X_n, X_1 + \dots + X_n \rangle = \|X_1\|^2 + \dots + \|X_n\|^2 + 2\langle X_1, X_2 \rangle + \dots + 2\langle X_1, X_n \rangle + \dots + 2\langle X_{n-1}, X_n \rangle = var(X_1) + \dots + var(X_n) + 2\sum_{h \neq k} cov(X_h, X_k)$ , dove l'ultimo addendo aggiunge tutti i doppi prodotti. ■

**Def 15.5** Dato  $(\Omega, \mathcal{A}, \mathbf{P})$  spazio di probabilità,  $(X, Y) : \Omega \rightarrow \mathbb{R}^2$  v.a. tali che  $X, Y \in L^2(\Omega, \mathcal{A}, \mathbf{P})$ ,  $var(X) = \sigma_X^2 > 0$  e  $var(Y) = \sigma_Y^2 > 0$ , si definisce coefficiente di correlazione lineare di  $X$  e  $Y$ , e si indica con  $\rho(X, Y)$ ,  $\frac{cov(X, Y)}{\sqrt{\sigma_X^2 \sigma_Y^2}}$ .

La definizione è ben posta perché  $\sigma_X^2, \sigma_Y^2 > 0$ . Per (6)  $|\rho(X, Y)| \leq 1$ : quindi la correlazione ha lo stesso significato della covarianza, ma è più facilmente comprensibile perché normalizzata.

**Prop 15.6 (correlazione unitaria)**

Hp:  $(\Omega, \mathcal{A}, \mathbf{P})$  spazio di probabilità,  $(X, Y) : \Omega \rightarrow \mathbb{R}^2$  v.a.,

$$X, Y \in L^2(\Omega, \mathcal{A}, \mathbf{P}), \sigma_X^2 > 0, \sigma_Y^2 > 0$$

Ts:  $|\rho(X, Y)| = 1 \iff \exists a \neq 0, b \in \mathbb{R} : Y = aX + b$  q. c. e

$$a = \frac{cov(X, Y)}{\sigma_X^2}, b = \mathbf{E}(Y) - \frac{cov(X, Y)\mathbf{E}(X)}{\sigma_X^2}$$

La proposizione afferma che due v. a. hanno coefficiente di correlazione unitario se e solo se una è una trasformazione lineare affine dell'altra. Nella tesi si può avere equivalentemente  $a = \rho(X, Y) \frac{\sigma_Y}{\sigma_X}, b = \mu_Y - \mu_X \rho(X, Y) \frac{\sigma_X}{\sigma_Y}$ .

**Dim\*** Dimostro l'implicazione da destra a sinistra. Se  $Y = aX + b$ ,  $var(Y) = a^2 var(X)$  e  $cov(X, Y) = avar(X)$ , per cui  $\frac{cov(X, Y)}{\sqrt{\sigma_X^2 \sigma_Y^2}} = \frac{avar(X)}{\sqrt{var(X) a^2 var(X)}} = 1$ . ■

**Def 16.1** Dato  $(\Omega, \mathcal{A}, \mathbf{P})$  spazio di probabilità,  $\mathbf{X} : \Omega \rightarrow \mathbb{R}^n$  vettore aleatorio con  $X_k \in L^1((\Omega, \mathcal{A}, \mathbf{P})) \forall k = 1, \dots, n$ , si definisce vettore delle medie di  $\mathbf{X}$ , e si indica con  $\mathbf{E}(\mathbf{X})$  o  $\mu_{\mathbf{X}}$ , il vettore  $\begin{pmatrix} \mathbf{E}(X_1) \\ \dots \\ \mathbf{E}(X_n) \end{pmatrix} \in \mathbb{R}^n$ .

**Def 16.2** Dato  $(\Omega, \mathcal{A}, \mathbf{P})$  spazio di probabilità,  $\mathbf{X} : \Omega \rightarrow \mathbb{R}^n$  vettore aleatorio con  $X_k \in L^2((\Omega, \mathcal{A}, \mathbf{P})) \forall k = 1, \dots, n$ , si definisce matrice delle covarianze di  $\mathbf{X}$ , e si indica con  $C_{\mathbf{X}}$ , la matrice  $\mathbf{E}((\mathbf{X} - \mu_{\mathbf{X}})(\mathbf{X} - \mu_{\mathbf{X}})^T) \in M_{\mathbb{R}}(n, n)$ .

$$(\mathbf{X} - \mu_{\mathbf{X}})(\mathbf{X} - \mu_{\mathbf{X}})^T = \begin{bmatrix} (X_1 - \mathbf{E}(X_1))^2 & \dots & (X_1 - \mathbf{E}(X_1))(X_n - \mathbf{E}(X_n)) \\ \dots & \dots & \dots \\ (X_n - \mathbf{E}(X_n))(X_1 - \mathbf{E}(X_1)) & \dots & (X_n - \mathbf{E}(X_n))^2 \end{bmatrix} : \text{se ne calcola il}$$

valore atteso componente per componente. Dalla definizione è evidente che  $(C_{\mathbf{X}})_{ij} = E((X_i - \mathbf{E}(X_i))(X_j - \mathbf{E}(X_j))) =$

$$cov(X_i, X_j), \text{ cioè } C_{\mathbf{X}} = \begin{bmatrix} var(X_1) & \dots & cov(X_1, X_n) \\ \dots & \dots & \dots \\ cov(X_n, X_1) & \dots & var(X_n) \end{bmatrix}.$$

**Teo 16.3 (proprietà della matrice di covarianza)**

Hp:  $(\Omega, \mathcal{A}, \mathbf{P})$  spazio di probabilità,  $\mathbf{X} : \Omega \rightarrow \mathbb{R}^n$  vettore aleatorio,

$X_k \in L^2((\Omega, \mathcal{A}, \mathbf{P})) \forall k = 1, \dots, n$ ,  $C_{\mathbf{X}}$  matrice delle covarianze di  $\mathbf{X}$

Ts: (1)  $C_{\mathbf{X}}$  è simmetrica e semidefinita positiva

(2)  $\mathbf{X} \in colC_{\mathbf{X}} + \mathbf{E}(\mathbf{X})$  q. c.

(3) se  $A \in M_{\mathbb{R}}(k, n)$ ,  $\mathbf{b} \in \mathbb{R}^k$ ,  $\mathbf{Y} = A\mathbf{X} + \mathbf{b}$ , allora

$$\mathbf{E}(\mathbf{Y}) = A\mathbf{E}(\mathbf{X}) + \mathbf{b} \text{ e } C_{\mathbf{Y}} = AC_{\mathbf{X}}A^T$$

$C_{\mathbf{X}}$  semidefinita positiva significa che  $\mathbf{a}^T C_{\mathbf{X}} \mathbf{a} \geq 0 \forall \mathbf{a} \in \mathbb{R}^n$ .

(2) significa che  $\mathbf{E}(\mathbf{X})$  rappresenta un valore centrale per  $\mathbf{X}$ , attorno a cui  $\mathbf{X}$  oscilla con variabilità governata dalla matrice delle covarianze.  $colC_{\mathbf{X}} + \mathbf{E}(\mathbf{X})$  è uno spazio lineare affine. Se  $\det C_{\mathbf{X}} \neq 0$ , (2) diventa  $\mathbf{X} \in \mathbb{R}^n$  q. c., che è un'informazione inutile: e. g., se  $n = 1$ , significa che  $X \in \{\mathbf{E}(X) + avar(X) : a \in \mathbb{R}\}$  q. c., cioè  $X$  appartiene q. c. all'asse reale se  $var(X) \neq 0$ . Se invece  $r(C_{\mathbf{X}}) = \dim(colC_{\mathbf{X}}) < n$ ,  $\mathbf{X}$  non assume tutti i valori in  $\mathbb{R}^n$ , e soprattutto  $\mathbf{P}(\mathbf{X} \in colC_{\mathbf{X}} + \mathbf{E}(\mathbf{X})) = 1$ , dove  $colC_{\mathbf{X}} + \mathbf{E}(\mathbf{X})$  è un sottospazio affine di  $\mathbb{R}^n$  di dimensione inferiore a  $n$ . Ma è noto che in tal caso  $m_n(colC_{\mathbf{X}} + \mathbf{E}(\mathbf{X})) = 0$ , per cui la legge di  $\mathbf{X}$  non è una probabilità assolutamente

continua, altrimenti dovrebbe preservare la misura nulla. Per capire se  $\mathbf{X}$  non è assolutamente continuo è quindi sufficiente calcolare  $\det C_{\mathbf{X}}$ : se  $\mathbf{X}$  è assolutamente continuo  $C_{\mathbf{X}}$  è definita positiva.

**Dim** (1) Poiché  $(C_{\mathbf{X}})_{ij} = \text{cov}(X_i, X_j)$ ,  $(C_{\mathbf{X}})_{ij} = (C_{\mathbf{X}})_{ji}$  per commutatività della varianza. Inoltre, per definizione di forma quadratica,  $\mathbf{a}^T C_{\mathbf{X}} \mathbf{a} = \sum_{k=1}^n \sum_{l=1}^n a_k (C_{\mathbf{X}})_{kl} a_l = \sum_{k=1}^n \sum_{l=1}^n a_k \text{cov}(X_k, X_l) a_l$ , cioè, per bilinearità della covarianza,  $\sum_{k=1}^n a_k \text{cov}(X_k, \sum_{l=1}^n a_l X_l) = \text{cov}(\sum_{k=1}^n a_k X_k, \sum_{l=1}^n a_l X_l) = \text{cov}(a_1 X_1 + \dots + a_n X_n, a_1 X_1 + \dots + a_n X_n) = \text{var}(\sum_{k=1}^n a_k X_k) \geq 0$  per positività della varianza. Questo è un modo per calcolare una varianza utile.

(2) Mostro che  $\mathbf{X} - \mathbf{E}(\mathbf{X}) \in (\ker C_{\mathbf{X}})^\perp$  q. c., cioè che  $\forall \mathbf{v} : C_{\mathbf{X}} \mathbf{v} = \mathbf{0}$  vale  $\langle \mathbf{X} - \mu, \mathbf{v} \rangle = 0$  q. c.. E' noto che  $C_{\mathbf{X}} = \mathbf{E}((\mathbf{X} - \mu)(\mathbf{X} - \mu)^T)$ : quindi se  $C_{\mathbf{X}} \mathbf{v} = \mathbf{0}$  allora  $\mathbf{E}((\mathbf{X} - \mu)(\mathbf{X} - \mu)^T) \mathbf{v} = \mathbf{0}$  e  $\mathbf{E}(\mathbf{v}^T (\mathbf{X} - \mu)(\mathbf{X} - \mu)^T \mathbf{v}) = 0$ . Ma  $\mathbf{v}^T (\mathbf{X} - \mu)(\mathbf{X} - \mu)^T \mathbf{v} = \left\| (\mathbf{X} - \mu)^T \mathbf{v} \right\|^2 \geq 0$ , e se il valore atteso di una v. a. nonnegativa è zero allora la v. a. è nulla q. c., cioè  $(\mathbf{X} - \mu)^T \mathbf{v} = 0$  q. c..

(3) Per definizione  $C_{\mathbf{Y}} = \mathbf{E}((\mathbf{Y} - \mu_{\mathbf{Y}})(\mathbf{Y} - \mu_{\mathbf{Y}})^T) = \mathbf{E}((\mathbf{A}\mathbf{X} + \mathbf{b} - \mathbf{A}\mathbf{E}(\mathbf{X}) - \mathbf{b})(\mathbf{A}\mathbf{X} + \mathbf{b} - \mathbf{A}\mathbf{E}(\mathbf{X}) - \mathbf{b})^T) = \mathbf{E}((\mathbf{A}(\mathbf{X} - \mathbf{E}(\mathbf{X}))) (\mathbf{A}(\mathbf{X} - \mathbf{E}(\mathbf{X})))^T) = \mathbf{A} \mathbf{E}((\mathbf{X} - \mathbf{E}(\mathbf{X})) (\mathbf{X} - \mathbf{E}(\mathbf{X}))^T) \mathbf{A}^T$ , per linearità del valore atteso e trasposizione di un prodotto. ■

Date  $X_1, \dots, X_n : \Omega \rightarrow \mathbb{R}$  v. a. indipendenti e identicamente distribuite con  $X_k \in L^2(\Omega, \mathcal{A}, \mathbf{P})$ , posso vedere tali v. a. come componenti di un vettore  $\mathbf{X} : \Omega \rightarrow \mathbb{R}^n$ .  $\mathbf{E}(\mathbf{X})$  è un vettore con tutte le componenti uguali a  $\mu$ , perché  $\mathbf{E}(X_i)$  dipende solo dalla legge di  $X_i$  per la legge del valore atteso.  $C_{\mathbf{X}} = \begin{bmatrix} \sigma^2 & \dots & 0 \\ \dots & \dots & \dots \\ 0 & \dots & \sigma^2 \end{bmatrix}$  è una matrice diagonale, multipla dell'identità, perché per indipendenza ogni covarianza tra v. a. diverse è nulla; la varianza è la stessa per tutte le v. a. perché dipende solo dalla legge.

Se definisco media campionaria  $\bar{X}_n = \frac{X_1 + \dots + X_n}{n}$ , vale per linearità che  $\mathbf{E}(\bar{X}_n) = \frac{1}{n} n \mu = \mu$ . Per 15.4 (7) vale inoltre  $\text{var}(\sum_{k=1}^n X_k) = \sum_{k=1}^n \text{var}(X_k) = n \sigma^2$ , per cui  $\text{var}(\bar{X}_n) = \frac{1}{n^2} n \sigma^2 = \frac{\sigma^2}{n}$ .

Se invece definisco varianza campionaria  $S_n^2 = \frac{1}{n-1} \sum_{k=1}^n (X_k - \bar{X}_n)^2$ , vale  $S_n^2 = \frac{1}{n-1} \sum_{k=1}^n (X_k^2 + \bar{X}_n^2 - 2X_k \bar{X}_n)$ . Poiché  $\text{var}(X_k) = \mathbf{E}(X_k^2) - \mathbf{E}^2(X_k)$ , vale  $\mathbf{E}(X_k^2) = \sigma^2 + \mu^2$ , e  $\mathbf{E}(X_k \bar{X}_n) = \text{cov}(X_k, \bar{X}_n) + \mathbf{E}(X_k) \mathbf{E}(\bar{X}_n) = \frac{1}{n} \text{var}(X_k) + \mathbf{E}(X_k) \mathbf{E}(\bar{X}_n)$ : per linearità  $\mathbf{E}(S_n^2) = \frac{1}{n-1} \left( n(\sigma^2 + \mu^2) + n \left( \frac{\sigma^2}{n} + \mu^2 \right) - 2n \frac{\sigma^2}{n} - 2n \mu^2 \right) = \frac{1}{n-1} (n \sigma^2 + \sigma^2 - 2 \sigma^2) = \sigma^2$ .



## 11 Funzione caratteristica

Nel seguito  $\mathbf{x} = \begin{pmatrix} x_1 \\ \dots \\ x_n \end{pmatrix}$  indica il generico vettore colonna di  $\mathbb{R}^n$ ,  $\mathbf{x}^T \in M_{\mathbb{R}}(1, n)$ , il prodotto scalare standard di  $\mathbb{R}^n$  è  $\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{i=1}^n x_i y_i = \langle \mathbf{x} | \mathbf{y} \rangle$ .

Introdurremo un nuovo modo per caratterizzare una probabilità su  $\mathbb{R}^n$  - oltre alle già viste funzione di ripartizione e densità - che semplificherà il calcolo dei valori attesi e faciliterà lo studio dell'indipendenza e delle combinazioni lineari di variabili aleatorie.

**Def 16.4** Dato  $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \mathbf{P})$  spazio di probabilità, si dice funzione caratteristica di  $\mathbf{P}$ , e si indica con  $\phi$  o  $\hat{\mathbf{P}}$ , la funzione  $\phi : \mathbb{R}^n \rightarrow \mathbb{C}$ ,  $\phi(\mathbf{u}) = \int_{\mathbb{R}^n} e^{i\langle \mathbf{u}, \mathbf{x} \rangle} d\mathbf{P}(\mathbf{x})$ .

(greg dice antitrasformata) Si noti che il prodotto scalare non ha senso se  $\mathbf{P}$  non è una probabilità sui boreliani.

La funzione integranda è  $h : \mathbb{R}^n \rightarrow \mathbb{C}$ ,  $h(\mathbf{x}) = e^{i\langle \mathbf{u}, \mathbf{x} \rangle} : \exists h_1, h_2 : \mathbb{R} \rightarrow \mathbb{R} : h(\mathbf{x}) = h_1(\mathbf{x}) + i h_2(\mathbf{x})$ , dunque  $h$  è univocamente individuata dal vettore delle sue componenti. Per definizione  $\int_{\mathbb{R}^n} e^{i\langle \mathbf{u}, \mathbf{x} \rangle} d\mathbf{P}(\mathbf{x}) := \int_{\mathbb{R}^n} h_1(\mathbf{x}) d\mathbf{P}(\mathbf{x}) + i \int_{\mathbb{R}^n} h_2(\mathbf{x}) d\mathbf{P}(\mathbf{x})$ , dunque tutte le proprietà e i teoremi dell'integrale rispetto a una misura di probabilità, visti per  $h$  a valori reali, valgono anche per funzioni complesse, poiché valgono separatamente per parte reale e immaginaria. In particolare  $h \in L^1(\mathbf{P}) \iff h_1, h_2 \in L^1(\mathbf{P})$ .

Poiché  $e^{i\theta} = \cos \theta + i \sin \theta$ ,  $|e^{i\theta}| = 1$ , e anche  $|e^{i\langle \mathbf{u}, \mathbf{x} \rangle}| = 1$ . Quindi  $|\int_{\mathbb{R}^n} h(\mathbf{x}) d\mathbf{P}(\mathbf{x})| \leq \int_{\mathbb{R}^n} |h(\mathbf{x})| d\mathbf{P}(\mathbf{x})$ . Se  $h(\mathbf{x}) = e^{i\langle \mathbf{u}, \mathbf{x} \rangle}$ ,  $h = h_{\mathbf{u}}$  è limitata e quindi integrabile rispetto a  $\mathbf{P}$ , dunque la definizione di  $\phi$  è ben posta qualsiasi sia  $\mathbf{P}$  (come si può vedere anche separatamente da parte reale e parte immaginaria, che sono funzioni misurabili limitate e quindi integrabili rispetto a  $\mathbf{P}$ ); inoltre  $|\int_{\mathbb{R}^n} e^{i\langle \mathbf{u}, \mathbf{x} \rangle} d\mathbf{P}(\mathbf{x})| \leq \int_{\mathbb{R}^n} |e^{i\langle \mathbf{u}, \mathbf{x} \rangle}| d\mathbf{P}(\mathbf{x}) = \int_{\mathbb{R}^n} 1 d\mathbf{P}(\mathbf{x}) = \mathbf{P}(\mathbb{R}^n) = 1$ , cioè  $\phi$  è limitata in modulo.

**Def 16.5** Dato  $(\Omega, \mathcal{A}, \mathbf{P})$  spazio di probabilità e  $\mathbf{X} : \Omega \rightarrow \mathbb{R}^n$  vettore aleatorio con legge  $P^{\mathbf{X}} : \mathcal{B}(\mathbb{R}^n) \rightarrow [0, 1]$ , si dice funzione caratteristica di  $P^{\mathbf{X}}$ , e si indica con  $\phi_{\mathbf{X}}$  o  $\hat{P}^{\mathbf{X}}$ , la funzione  $\phi_{\mathbf{X}} : \mathbb{R}^n \rightarrow \mathbb{C}$ ,  $\phi_{\mathbf{X}}(\mathbf{u}) = \int_{\mathbb{R}^n} e^{i\langle \mathbf{u}, \mathbf{x} \rangle} dP^{\mathbf{X}}(\mathbf{x})$ .

Per la regola del valore atteso  $\int_{\mathbb{R}^n} e^{i\langle \mathbf{u}, \mathbf{x} \rangle} dP^{\mathbf{X}}(\mathbf{x}) = \int_{\Omega} e^{i\langle \mathbf{u}, \mathbf{X}(\omega) \rangle} d\mathbf{P}(\omega) = \mathbf{E}(e^{i\langle \mathbf{u}, \mathbf{X}(\omega) \rangle})$ .  $\phi_{\mathbf{X}}$  esiste qualsiasi sia  $\mathbf{X}$ , come la funzione di ripartizione. Se in particolare  $\mathbf{X}$  è assolutamente continuo,  $\phi_{\mathbf{X}}(\mathbf{u}) = \int_{\mathbb{R}^n} f_{\mathbf{X}}(\mathbf{x}) e^{i\langle \mathbf{u}, \mathbf{x} \rangle} d\mathbf{x}$ .

1. Considero  $X \sim \delta_{\mu}$ .  $p_X(\mu) = 1$ ,  $p_X(k) = 0 \forall k \neq \mu$ .  $\phi_X(u) = \int_{\mathbb{R}} e^{iux} dP^X(x) = \mathbf{E}(e^{iuX}) = e^{iu\mu} \mathbf{P}(X = \mu) = e^{iu\mu}$  perché  $e^{iuX}$  è una v. a. semplice.
2. Considero  $X \sim \text{poiss}(\lambda)$ .  $\phi_X(u) = \int_{\mathbb{R}} e^{iux} dP^X(x) = \mathbf{E}(e^{iuX}) = \sum_{k=0}^{+\infty} e^{iuk} e^{-\lambda} \frac{\lambda^k}{k!} = e^{-\lambda} \sum_{k=0}^{+\infty} \frac{(e^{iu}\lambda)^k}{k!} = e^{-\lambda} e^{\lambda e^{iu}}$  per la regola del valore atteso per v. a. discrete e perché la serie è assolutamente convergente, dato che  $|e^{iu}| = 1$ .

3. Considero  $X \sim \varepsilon(\lambda)$ .  $\phi_X(u) = \int_{\mathbb{R}} e^{iux} dP^X(x) = \mathbf{E}(e^{iuX}) = \int_0^{+\infty} e^{iux} \lambda e^{-\lambda x} dx = \lambda \int_0^{+\infty} e^{x(iu-\lambda)} dx = \frac{\lambda}{\lambda - iu}$   
per la regola del valore atteso per v. a. continue e perché  $\lim_{x \rightarrow +\infty} \frac{1}{iu-\lambda} e^{x(iu-\lambda)} = 0$  perché  $|e^{x(iu-\lambda)}| = |e^{iux} e^{-\lambda x}| \leq e^{-\lambda x}$  e  $\lim_{x \rightarrow +\infty} e^{-\lambda x} = 0$  (spiraleggia).

**Teo 17.1** ( $\phi$  caratterizza  $\mathbf{P}$ ; proprietà analitiche di  $\phi$ )

Hp:  $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \mathbf{P})$  spazio di probabilità,  $\hat{\mathbf{P}}/\phi$  è la funzione caratteristica di  $\mathbf{P}$

Ts: (1) se  $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \mathbf{Q})$  è uno spazio di probabilità,  $\mathbf{P} = \mathbf{Q} \iff \hat{\mathbf{Q}}(\mathbf{u}) = \hat{\mathbf{P}}(\mathbf{u}) \quad \forall \mathbf{u}$

$$(2) \phi(\mathbf{0}) = 1; |\phi(\mathbf{u})| \leq 1 \quad \forall \mathbf{u}$$

$$(3) \phi \text{ è continua in } \mathbb{R}^n$$

In (1) l'implicazione da sinistra e destra è ovvia per la definizione di  $\phi$ . (1) significa che il funzionale che a ogni probabilità associa la funzione caratteristica è biunivoco, quindi per controllare se due misure di probabilità coincidono è sufficiente confrontare le loro funzioni caratteristiche, che è molto più semplice. (3) implica che  $\phi$  in generale è una funzione più regolare sia di  $F$  che, quando esiste, di  $f$ . Si può dimostrare che  $\phi$  è anche uniformemente continua.

**Dim\*** (2)  $\phi(\mathbf{0}) = \int_{\mathbb{R}^n} 1 d\mathbf{P}(\mathbf{x}) = \mathbf{P}(\mathbb{R}^n) = 1$ . Per quanto già visto  $|\phi(\mathbf{u})| = \left| \int_{\mathbb{R}^n} e^{i\langle \mathbf{u}, \mathbf{x} \rangle} d\mathbf{P}(\mathbf{x}) \right| \leq \int_{\mathbb{R}^n} |e^{i\langle \mathbf{u}, \mathbf{x} \rangle}| d\mathbf{P}(\mathbf{x}) = \int_{\mathbb{R}^n} 1 d\mathbf{P}(\mathbf{x}) = \mathbf{P}(\mathbb{R}^n) = 1$ .

(3) Si vuole dimostrare che  $\forall \{\mathbf{u}_k\} : \lim_{k \rightarrow +\infty} \mathbf{u}_k = \mathbf{u}$  vale  $\lim_{k \rightarrow +\infty} \phi(\mathbf{u}_k) = \phi(\mathbf{u})$ .  $\lim_{k \rightarrow +\infty} e^{i\langle \mathbf{u}_k, \mathbf{x} \rangle} = e^{i\langle \mathbf{u}, \mathbf{x} \rangle}$  per continuità dell'esponenziale;  $\phi(\mathbf{u}_k) = \int_{\mathbb{R}^n} e^{i\langle \mathbf{u}_k, \mathbf{x} \rangle} d\mathbf{P}(\mathbf{x})$  e la funzione  $|e^{i\langle \mathbf{u}_k, \mathbf{x} \rangle}| = 1$  è limitata da 1, che è integrabile, quindi per il teorema di convergenza dominata (nel caso particolare  $\Omega = \mathbb{R}$ ; gli elementi della successione sono tutti misurabili perché funzioni continue), dato che  $e^{i\langle \mathbf{u}_k, \mathbf{x} \rangle} \rightarrow e^{i\langle \mathbf{u}, \mathbf{x} \rangle}$ ,  $\int_{\mathbb{R}^n} e^{i\langle \mathbf{u}_k, \mathbf{x} \rangle} d\mathbf{P}(\mathbf{x})$  converge a  $\int_{\mathbb{R}^n} e^{i\langle \mathbf{u}, \mathbf{x} \rangle} d\mathbf{P}(\mathbf{x})$ . ■

**Teo 17.2** (regolarità di  $\phi$ , formula dei momenti, fattorizzazione della funzione caratteristica)

Hp:  $(\Omega, \mathcal{A}, \mathbf{P})$  spazio di probabilità,  $\mathbf{X} : \Omega \rightarrow \mathbb{R}^n$  vettore aleatorio

con legge  $P^{\mathbf{X}} : \mathcal{B}(\mathbb{R}^n) \rightarrow [0, 1]$  e funzione caratteristica  $\phi_{\mathbf{X}}$

Ts: (1) se  $\exists m \in \mathbb{N}^* : X_k \in L^m(\mathbb{R}) \quad \forall k = 1, \dots, n$ , allora  $\phi_{\mathbf{X}} \in C^m(\mathbb{R}^n, \mathbb{C})$

e date  $k_1, \dots, k_m$  direzioni,  $\frac{\partial^m}{\partial u_{k_1} \dots \partial u_{k_m}} \phi_{\mathbf{X}}(\mathbf{u}) = i^m \mathbf{E}(X_{k_1} \dots X_{k_m} e^{i\langle \mathbf{u}, \mathbf{x} \rangle})$

(2) se  $A \in M_{\mathbb{R}}(k, n)$ ,  $\mathbf{b} \in \mathbb{R}^k$ ,  $\mathbf{Y} = A\mathbf{X} + \mathbf{b}$ , allora  $\phi_{\mathbf{Y}}(\mathbf{v}) = e^{i\langle \mathbf{b}, \mathbf{v} \rangle} \phi_{\mathbf{X}}(A^T \mathbf{v})$

(3)  $X_1, \dots, X_n$  sono v. a. indipendenti  $\iff \phi_{\mathbf{X}}(\mathbf{u}) = \phi_{X_1}(u_1) \dots \phi_{X_n}(u_n) \quad \forall \mathbf{u} \in \mathbb{R}^n$

(1.1) afferma che l'integrabilità delle componenti di  $\mathbf{X}$  trasferisce regolarità a  $\phi_{\mathbf{X}}$ ; non vale il viceversa. In (1.2) le direzioni possono anche non essere tutte distinte. (1.2) afferma la possibilità di scambiare derivata e integrale: in  $\mathbf{E}(X_{k_1} \dots X_{k_m} e^{i\langle \mathbf{u}, \mathbf{X} \rangle})$  si è derivato  $e^{i\langle \mathbf{u}, \mathbf{X} \rangle}$  rispetto alle direzioni  $u_{k_1}, \dots, u_{k_m}$  prima di calcolare l'integrale: l'uguaglianza è infatti  $\frac{\partial^m}{\partial u_{k_1} \dots \partial u_{k_m}} \int_{\mathbb{R}^n} e^{i\langle \mathbf{u}, \mathbf{x} \rangle} dP^{\mathbf{X}}(\mathbf{x}) = i^m \int_{\mathbb{R}^n} x_{k_1} \dots x_{k_m} e^{i\langle \mathbf{u}, \mathbf{x} \rangle} dP^{\mathbf{X}}(\mathbf{x})$ . (2) permette di calcolare  $\phi_{\mathbf{Y}}$  con  $\phi_{\mathbf{X}}$  se  $\mathbf{Y}$  è una trasformazione lineare affine di  $\mathbf{X}$ . Se in particolare  $Y = -X$ ,  $\phi_Y(v) = \phi_X(-v) = \bar{\phi}_X(v)$ . (3) afferma che  $n$  v. a. sono indipendenti se e solo se la  $\phi$  del vettore che esse compongono si fattorizza. Non è sufficiente che  $\phi_{\mathbf{X}}(u, u, \dots, u) = \phi_{X_1}(u) \dots \phi_{X_n}(u) \forall u \in \mathbb{R}$ .

$X_k \in L^m(\mathbf{P}) \iff \mathbf{E}(|X_k|^m) < +\infty$ . Quindi  $|X_{k_1} \dots X_{k_m} e^{i\langle \mathbf{u}, \mathbf{X} \rangle}| = |X_{k_1}| \dots |X_{k_m}|$ , che è il prodotto di  $m$  v. a. ciascuna appartenente a  $L^m$ : come il prodotto di 2 v. a. in  $L^2$  appartiene a  $L^1$ ,  $|X_{k_1}| \dots |X_{k_m}|$  appartiene a  $L^1$  e dunque  $X_{k_1} \dots X_{k_m} \in L^1$ : perciò il lato destro della formula è ben definito (!).

**Dim** (2)  $\phi_{\mathbf{Y}}(\mathbf{v}) = \mathbf{E}(e^{i\langle \mathbf{v}, \mathbf{Y} \rangle}) = \mathbf{E}(e^{i\langle \mathbf{v}, A\mathbf{X} + \mathbf{b} \rangle}) = \mathbf{E}(e^{i\langle \mathbf{v}, A\mathbf{X} \rangle} e^{i\langle \mathbf{v}, \mathbf{b} \rangle}) = e^{i\langle \mathbf{v}, \mathbf{b} \rangle} \mathbf{E}(e^{i\langle \mathbf{v}, A\mathbf{X} \rangle})$ . Poiché  $\langle \mathbf{v}, A\mathbf{X} \rangle = \mathbf{v}^T A\mathbf{X} = (A^T \mathbf{v})^T \mathbf{X}$ , si ha  $\phi_{\mathbf{Y}}(\mathbf{v}) = e^{i\langle \mathbf{v}, \mathbf{b} \rangle} \mathbf{E}(e^{i\langle A^T \mathbf{v}, \mathbf{X} \rangle}) = e^{i\langle \mathbf{v}, \mathbf{b} \rangle} \phi_{\mathbf{X}}(A^T \mathbf{v})$ .

(3) Se  $X_1, \dots, X_n$  sono indipendenti,  $\phi_{\mathbf{X}}(\mathbf{u}) = \mathbf{E}(e^{i\langle \mathbf{u}, \mathbf{X} \rangle}) = \mathbf{E}(e^{iu_1 X_1} \dots e^{iu_n X_n}) = \mathbf{E}(e^{iu_1 X_1}) \dots \mathbf{E}(e^{iu_n X_n}) = \phi_{X_1}(u_1) \dots \phi_{X_n}(u_n)$ . Il valore atteso si fattorizza perché  $e^{iu_1 X_1}, \dots, e^{iu_n X_n}$  sono una famiglia di v. a. indipendenti e  $e^{iux}$  è misurabile e limitata.

Mostro l'implicazione da destra a sinistra: voglio mostrare che le  $X_i$  sono indipendenti, cioè che la legge di  $\mathbf{X}$  è la legge prodotto. Allora considero la legge  $P^{\mathbf{X}} = P^{X_1} \otimes \dots \otimes P^{X_n}$  e mostro che ha funzione caratteristica  $\phi_{\mathbf{X}}(\mathbf{u}) = \phi_{X_1}(u_1) \dots \phi_{X_n}(u_n)$ , per cui quella è la legge del vettore  $\mathbf{X}$ , dato che  $\phi_{\mathbf{X}}$  caratterizza  $P^{\mathbf{X}}$ . La funzione caratteristica della probabilità prodotto è  $\int_{\mathbb{R}^n} e^{i\langle \mathbf{u}, \mathbf{x} \rangle} dP^{X_1} \otimes \dots \otimes P^{X_n}(\mathbf{x}) = \int_{\mathbb{R}^n} e^{iu_1 x_1} \dots e^{iu_n x_n} dP^{X_1} \otimes \dots \otimes P^{X_n}(\mathbf{x}) = \int_{\mathbb{R}^n} e^{iu_1 x_1} dP^{X_1}(x_1) \dots \int e^{iu_n x_n} dP^{X_n}(x_n)$ , dove l'ultima uguaglianza vale per il teorema di Fubini Tonelli (super Fubini 9.12), dato che la funzione integranda è a variabili separabili. Quindi  $\int_{\mathbb{R}^n} e^{i\langle \mathbf{u}, \mathbf{x} \rangle} dP^{X_1} \otimes \dots \otimes P^{X_n}(\mathbf{x}) = \phi_{X_1}(x_1) \dots \phi_{X_n}(x_n)$ . Allora la legge prodotto dev'essere la legge di  $\mathbf{X}$  e le sue componenti sono una famiglia di v. a. indipendenti. ■

1. La formula 1.2 è detta dei momenti perché permette di calcolare più facilmente i momenti di una variabile aleatoria, prendendo  $\mathbf{u} = \mathbf{0}$ : si ottiene  $\frac{\partial^m}{\partial u_{k_1} \dots \partial u_{k_m}} \phi_{\mathbf{X}}(\mathbf{0}) = i^m \mathbf{E}(X_{k_1} \dots X_{k_m})$ . Sia  $n = 1$ ,  $X \in L^2(\mathbf{P})$ : in questo caso  $m = 2$ , quindi  $\phi_X \in C^2(\mathbb{R}, \mathbb{C})$ . Ma poiché  $L^2 \subseteq L^1$ , posso anche scegliere  $m = 1$ : applicando la formula,  $\frac{d}{du} \phi_X(0) = i \mathbf{E}(X)$ , quindi  $\mathbf{E}(X) = -i \phi'_X(0)$ . Analogamente, con  $m = 2$   $\frac{d^2}{du^2} \phi_X(0) = i^2 \mathbf{E}(X^2)$ , cioè  $\mathbf{E}(X^2) = -\phi''_X(0)$ . In generale, se si vuole conoscere il momento  $m$ -esimo, si prende  $k_1 = \dots = k_m = 1$  e si ottiene  $\phi_X^{(m)}(0) = i^m \mathbf{E}(X^m)$ . Vale dunque  $\text{var}(X) = -\phi''_X(0) + (\phi'_X(0))^2$ .

Sia  $n = 2$ ,  $X_1, X_2 \in L^2(\mathbf{P})$ : essendo  $m = 2$ ,  $\phi_X \in C^2(\mathbb{R}^2, \mathbb{C})$ . Vale quindi  $\frac{\partial^2 \phi_{\mathbf{X}}(0,0)}{\partial u_1 \partial u_2} = i^2 \mathbf{E}(X_1 X_2)$ , che

implica  $\mathbf{E}(X_1 X_2) = -\frac{\partial^2 \phi_{\mathbf{X}}(0,0)}{\partial u_1 \partial u_2}$ . Con  $m = 1$ , prendendo  $k_1 = k_2 = 1$  si ottiene ancora  $\mathbf{E}(X_1) = -i\phi'_{X_1}(0) = -i\frac{\partial}{\partial u_1}\phi_{\mathbf{X}}(0,0)$ ,  $\mathbf{E}(X_2) = -i\frac{\partial}{\partial u_2}\phi_{\mathbf{X}}(0,0)$ , dunque  $\text{cov}(X_1, X_2) = -\frac{\partial^2 \phi_{\mathbf{X}}(\mathbf{0})}{\partial u_1 \partial u_2} + \frac{\partial}{\partial u_1}\phi_{\mathbf{X}}(\mathbf{0})\frac{\partial}{\partial u_2}\phi_{\mathbf{X}}(\mathbf{0})$ . L'integrabilità delle componenti permette quindi di risalire a tutti gli indici desiderati semplicemente calcolando derivate.

Si noti che vale  $\phi'_{X_1}(0) = \frac{\partial}{\partial u_1}\phi_{\mathbf{X}}(0,0)$  perché con  $m = 1$  si ha  $\frac{\partial}{\partial u_1}\phi_{\mathbf{X}}(\mathbf{0}) = i\mathbf{E}(X_1)$ , cioè  $\mathbf{E}(X_1) = -i\frac{\partial}{\partial u_1}\phi_{\mathbf{X}}(\mathbf{0})$ .

2. Considero  $X \sim \text{geom}(p)$ :  $p_X(k) = (1-p)^{k-1}p \forall k \geq 1$ . Posso calcolare  $\mathbf{E}(X)$  usando la formula dei momenti:  $\mathbf{E}(X) = -i\phi'_X(0)$ . Poiché  $\phi_X(u) = \mathbf{E}(e^{iuX}) = \sum_{k=1}^{+\infty} e^{iuk} (1-p)^{k-1} p = pe^{iu} \sum_{k=1}^{+\infty} (e^{iu}(1-p))^{k-1} = \frac{pe^{iu}}{1-e^{iu}(1-p)}$ ,  $\phi'_X(u) = \frac{ipe^{iu}}{(1-e^{iu}(1-p))^2}$ , quindi  $\mathbf{E}(X) = \frac{p}{p^2} = \frac{1}{p}$ .

3. Considero  $X \sim \mathcal{N}(0,1)$ .  $\phi_X(u) = \int_{\mathbb{R}} e^{iux} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx = \int_{\mathbb{R}} \cos ux \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx + i \int_{\mathbb{R}} \sin ux \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx = \int_{\mathbb{R}} \cos ux \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx$  (il secondo addendo è nullo perché la funzione integranda è dispari). Si è visto prima che  $\phi'_X(u) = i\mathbf{E}(Xe^{iuX}) = i \int_{\mathbb{R}} x \cos ux \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx - \int_{\mathbb{R}} x \sin ux \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx = -\frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} x \sin uxe^{-\frac{1}{2}x^2} dx$ . Integrando per parti si ha  $\int x \sin(ux) e^{-\frac{1}{2}x^2} dx = -e^{-\frac{1}{2}x^2} \sin(ux) + \int ue^{-\frac{1}{2}x^2} \cos(ux) dx$ , per cui  $\phi'_X(u) = -\frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} ue^{-\frac{1}{2}x^2} \cos uxdx = -u\phi_X(u)$ . Si è quindi ottenuto un problema di Cauchy  $\begin{cases} \phi'_X(u) = -u\phi_X(u) \\ \phi_X(0) = 1 \end{cases}$ : esiste un'unica soluzione  $\phi_X(u) = e^{-\frac{1}{2}u^2}$ , che è limitata come dev'essere e ha parte immaginaria nulla.

Considero  $X \sim \mathcal{N}(\mu, \sigma^2)$ .  $\exists Z \sim \mathcal{N}(0,1) : X = \sigma Z + \mu$ : allora per 17.3 (2)  $\phi_X(u) = e^{iu\mu} \phi_X(\sigma u) = e^{iu\mu} e^{-\frac{1}{2}\sigma^2 u^2}$ : si ha un termine lineare e uno quadratico all'esponente.

4. Dato  $\mathbf{X}$ , se  $\phi_{\mathbf{X}}(\mathbf{u})$  è nota, come si ottiene  $\phi_{X_k}(v)$ ? Vale  $X_k = \langle \mathbf{e}_k, \mathbf{X} \rangle = \mathbf{e}_k^T \mathbf{X}$ , quindi  $\phi_{X_k}(v) = \phi_{\mathbf{X}}(\mathbf{e}_k v) = \phi_{\mathbf{X}}(0, \dots, 0, v, 0, \dots, 0)$ .

### Corollario 17.3

Hp:  $(\Omega, \mathcal{A}, \mathbf{P})$  spazio di probabilità,  $\mathbf{X}_1 : \Omega \rightarrow \mathbb{R}^{n_1}, \mathbf{X}_2 : \Omega \rightarrow \mathbb{R}^{n_2}$

vettori aleatori con funzioni caratteristiche  $\phi_{\mathbf{X}_1}, \phi_{\mathbf{X}_2}$

Ts:  $\mathbf{X}_1 \perp \mathbf{X}_2 \iff \phi_{(\mathbf{X}_1, \mathbf{X}_2)}(\mathbf{u}) = \phi_{\mathbf{X}_1}(u_1, \dots, u_{n_1}) \phi_{\mathbf{X}_2}(u_{n_1+1}, \dots, u_{n_1+n_2})$

**Dim** Si applica 17.2 (3) a  $\mathbf{X}_1, \mathbf{X}_2$ . ■

**Corollario 17.4 ( $\phi$  della somma)**

Hp:  $(\Omega, \mathcal{A}, \mathbf{P})$  spazio di probabilità,  $\mathbf{X} : \Omega \rightarrow \mathbb{R}^n$  vettore

aleatorio con funzione caratteristica  $\phi_{\mathbf{X}}$ ,  $S_n = X_1 + \dots + X_n$

Ts:  $\phi_{S_n}(v) = \phi_{\mathbf{X}}(v, \dots, v)$

(ii) se  $X_1, \dots, X_k$  sono indipendenti,  $\phi_{S_n}(v) = \phi_{X_1}(v) \dots \phi_{X_n}(v)$

(iii) se  $X_1, \dots, X_k$  sono iid,  $\phi_{S_n}(v) = (\phi_{X_i}(v))^n$

Non vale il viceversa in (ii), perché il fatto che  $\phi_{\mathbf{X}}(v, \dots, v)$  si fattorizzi  $\forall v$  non è sufficiente a concludere che si fattorizzi  $\forall \mathbf{v} = (v_1, \dots, v_n)$ .

**Dim** Vale  $S_n = \left\langle \begin{pmatrix} 1 \\ \dots \\ 1 \end{pmatrix}, \mathbf{X} \right\rangle$ , quindi per 17.2 (2)  $\phi_{S_n}(v) = \phi_{\mathbf{X}}(v, \dots, v)$ . Se inoltre le componenti sono indipendenti, per 17.2 (3) la funzione caratteristica si fattorizza:  $\phi_{S_n}(v) = \phi_{X_1}(v) \dots \phi_{X_n}(v)$ . Se inoltre le componenti

hanno tutte la stessa legge, hanno la stessa funzione caratteristica e  $\phi_{S_n}(v) = (\phi_{X_i}(v))^n$ . ■

1. Siano  $X_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$ ,  $X_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$ ,  $X_1 \perp X_2$ . Mostro che  $X_1 + X_2 \sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$ . Infatti per indipendenza vale  $\phi_{S_2}(v) = e^{iu\mu_1} e^{-\frac{1}{2}\sigma_1^2 u^2} e^{iu\mu_2} e^{-\frac{1}{2}\sigma_2^2 u^2} = e^{iu(\mu_1 + \mu_2)} e^{-\frac{1}{2}(\sigma_1^2 + \sigma_2^2)u^2}$ , che è la funzione caratteristica di  $Y \sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$ . Poiché  $\phi_Y$  individua univocamente la legge della variabile aleatoria,  $S_2$  deve avere la legge di  $Y$ .
2. Siano  $X_1 \sim \text{poiss}(\lambda_1)$ ,  $X_2 \sim \text{poiss}(\lambda_2)$ ,  $X_1 \perp X_2$ ,  $\lambda_1, \lambda_2 > 0$ . Mostro che  $X_1 + X_2 \sim \text{poiss}(\lambda_1 + \lambda_2)$ . Infatti per indipendenza vale  $\phi_{S_2}(v) = e^{-\lambda_1} e^{\lambda_1 e^{iu}} e^{-\lambda_2} e^{\lambda_2 e^{iu}} = e^{-(\lambda_1 + \lambda_2)} e^{(\lambda_1 + \lambda_2) e^{iu}}$ , che è la funzione caratteristica di  $Y \sim \text{poiss}(\lambda_1 + \lambda_2)$ . Poiché  $\phi_Y$  individua univocamente la legge della variabile aleatoria,  $S_2$  deve avere la legge di  $Y$ .
3. Sia  $Y \sim \text{bin}(n, p)$ .  $\phi_Y(u) = \mathbf{E}(e^{iuX}) = \sum_{k=0}^n e^{iuk} \binom{n}{k} p^k (1-p)^{n-k} = \sum_{k=0}^n \binom{n}{k} (e^{iu}p)^k (1-p)^{n-k}$ , che è il binomio di Newton:  $(e^{iu}p + 1 - p)^n$ . Se in particolare  $n = 1$ , si ha  $X \sim \text{bin}(1, p) = b(p)$  e  $\phi_X(u) = e^{iu}p + 1 - p$ .

Se  $X_1 \sim b(p)$ ,  $X_2 \sim b(p)$ ,  $X_1 \perp X_2$ , allora  $X_1 + X_2 \sim \text{bin}(2, p)$ : infatti per quanto visto sopra  $\phi_{S_2}(u) = (e^{iu}p + 1 - p)^2$ , che è la funzione caratteristica di  $Y \sim \text{bin}(2, p)$ , quindi  $S_2$  deve avere la legge di  $Y$ . In generale se  $X_1, \dots, X_n \sim b(p)$  e  $X_1, \dots, X_n$  sono una famiglia di v. a. indipendenti, allora  $S_n \sim \text{bin}(n, p)$ . Da questo si deduce che  $\mathbf{E}(Y) = np$ ,  $\text{var}(Y) = np(1 - p)$ .

Se  $Y_1 \sim \text{bin}(n_1, p)$ ,  $Y_2 \sim \text{bin}(n_2, p)$ ,  $Y_1 \perp Y_2$ , allora  $Y_1 + Y_2 \sim \text{bin}(n_1 + n_2, p)$ . Infatti  $\phi_{S_2}(u) = (e^{iu}p + 1 - p)^{n_1} (e^{iu}p + 1 - p)^{n_2} = (e^{iu}p + 1 - p)^{n_1 + n_2}$ , che è la funzione caratteristica di  $Y \sim \text{bin}(n_1 + n_2, p)$ .

4. Siano  $X_1 \sim \Gamma(\alpha, \lambda)$ ,  $X_2 \sim \Gamma(\beta, \lambda)$ ,  $X_1 \perp X_2$ ,  $\alpha, \beta, \lambda > 0$ . Mostro che  $X_1 + X_2 \sim \Gamma(\alpha + \beta, \lambda)$ . Infatti  $\phi_{(X_1, X_2)}(\mathbf{u}) = \phi_{X_1}(u_1) \phi_{X_2}(u_2) = \left(\frac{\lambda}{\lambda - iu}\right)^\alpha \left(\frac{\lambda}{\lambda - iu}\right)^\beta = \left(\frac{\lambda}{\lambda - iu}\right)^{\alpha + \beta}$ , che è la funzione caratteristica di una  $\Gamma(\alpha + \beta, \lambda)$ .

### Teo 17.5 (legge della somma di due variabili aleatorie)

Hp:  $\mathbf{X} : \Omega \rightarrow \mathbb{R}^2$  è un vettore aleatorio con legge  $P^{\mathbf{X}} : \mathcal{B}(\mathbb{R}^2) \rightarrow [0, 1]$ ,

$Y : \Omega \rightarrow \mathbb{R}$ ,  $Y = X_1 + X_2$  ha legge  $P^Y : \mathcal{B}(\mathbb{R}) \rightarrow [0, 1]$

Ts: (1)  $P^Y(B) = \int_{\mathbb{R}^2} I_B(x_1 + x_2) dP^{\mathbf{X}}(x_1, x_2) \quad \forall B \in \mathcal{B}(\mathbb{R})$

(2) se  $\mathbf{X}$  è assolutamente continuo con densità  $f_{\mathbf{X}}$ ,  $Y$  è assolutamente

continua e  $f_Y(y) = \int_{\mathbb{R}} f_{\mathbf{X}}(y - t, t) dt$ ; se inoltre  $X_1 \perp X_2$ ,  $f_Y(y) = \int_{\mathbb{R}} f_{X_1}(y - t) f_{X_2}(t) dt$

(3) se  $\mathbf{X}$  è discreto con densità discreta  $p_{\mathbf{X}}$ ,  $Y$  è discreta e con densità discreta

$$p_Y(y) = \sum_{x_2 \in S_2} p_{\mathbf{X}}(y - x_2, x_2); \text{ se inoltre } X_1 \perp X_2, p_Y(y) = \sum_{x_2 \in S_2} p_{X_1}(y - x_2) p_{X_2}(x_2)$$

(2) è equivalentemente  $\int_{\mathbb{R}} f_{X_2}(y - t) f_{X_1}(t) dt$  (basta fare un cambio di variabile).  $\int_{\mathbb{R}} f_{X_1}(y - t) f_{X_2}(t) dt$  è detto prodotto di convoluzione di  $f_{X_1}$  e  $f_{X_2}$ ;  $\sum_{x_2 \in S_2} p_{\mathbf{X}}(y - x_2, x_2)$  è il prodotto di convoluzione discreto.  $f_{X_1}(y - t) f_{X_2}(t)$  rappresenta il prodotto della probabilità che  $X_1$  sia circa  $y - t$  e  $X_2$  circa  $t$ , quindi la probabilità che la loro somma sia circa  $y$ ; la somma viene quindi fatta su tutti i valori di  $t$ .

$$\text{Dim* (1) } P^Y(B) = \mathbf{P}(Y \in B) = \mathbf{E}(I_{(Y \in B)}) = \mathbf{E}(I_B(Y)) = \mathbf{E}(I_B(X_1 + X_2)) = \int_{\mathbb{R}^2} I_B(x_1 + x_2) dP^{\mathbf{X}}(x_1, x_2)$$

per la regola del valore atteso.

(2)  $F_Y(t) = \mathbf{P}(X_1 + X_2 \leq t) = P^{(X_1, X_2)}(A)$ , con  $A = \{(x, y) : x \in (-\infty, t - y], y \in \mathbb{R}\}$ . Allora  $P^{(X_1, X_2)}(A) = \int_{-\infty}^{+\infty} \left( \int_{-\infty}^{t-y} f_{(X_1, X_2)}(x, y) dx \right) dy$ , e per il teorema fondamentale del calcolo integrale  $f_Y(t) = \frac{d}{dt} \left( \int_{-\infty}^{+\infty} \left( \int_{-\infty}^{t-y} f_{(X_1, X_2)}(x, y) dx \right) dy \right) = \int_{-\infty}^{+\infty} \frac{d}{dt} \left( \int_{-\infty}^{t-y} f_{(X_1, X_2)}(x, y) dx \right) dy = \int_{-\infty}^{+\infty} f_{(X_1, X_2)}(t - y, y) dy$  perché  $\lim_{y \rightarrow -\infty} f_{(X, Y)}(x, y) = 0$ . Se inoltre  $X \perp Y$ , la densità congiunta si fattorizza e  $f_Y(t) = \int_{-\infty}^{+\infty} f_{X_1}(t - y) f_{X_2}(y) dy$ . Alternativamente, si usa il vettore  $\begin{pmatrix} U \\ V \end{pmatrix} = \begin{pmatrix} X + Y \\ Y \end{pmatrix} = \mathbf{h}(X, Y)$ :  $\mathbf{g}(U, V) = \begin{pmatrix} U - V \\ V \end{pmatrix}$  ha  $J_{\mathbf{g}}(u, v) = \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix}$ , per cui  $f_{(U, V)}(u, v) = f_{(X, Y)}(u - v, v) \cdot 1$ , e la densità marginale di  $U$  è  $f_Z(u) = \int_{-\infty}^{+\infty} f_{(X, Y)}(u - v, v) dv$ . ■

## 12 Vettori gaussiani

Si è notato che  $X \sim \mathcal{N}(\mu, \sigma^2)$  ha funzione caratteristica  $\phi_X(u) = e^{iu\mu} e^{-\frac{1}{2}\sigma^2 u^2}$ . Dato che  $\phi_X$  caratterizza la legge di  $X$ , si potrebbe dare una definizione equivalente di variabile aleatoria gaussiana a partire da  $\phi_X$ . In verità usare la funzione caratteristica permette di includere tra le v. a. gaussiane nuove variabili aleatorie, perché non occorre chiedere  $\sigma^2 \neq 0$  come fatto per definire la densità della normale: nel caso  $\sigma^2 = 0$ , per 9.10, si ottiene infatti  $\phi_X(u) = e^{iu\mu}$ , che è la funzione caratteristica di una v. a.  $X = \mu$  q. c., cioè  $X \sim \delta_\mu$ : quindi si sono incluse anche le v. a. quasi certamente costanti, la cui densità può essere vista come limite della densità della normale al ridursi di  $\sigma^2$ .

**Def 18.1** Dato  $\mathbf{X} : \Omega \rightarrow \mathbb{R}^n$  vettore aleatorio e i parametri  $\mu = \begin{pmatrix} \mu_1 \\ \dots \\ \mu_n \end{pmatrix}$ ,  $C = (c_{ij})_{i,j=1,\dots,n} \in S^+(n, n)$ ,  $\mathbf{X}$  si

dice vettore gaussiano di parametri  $\mu, C$  se ha funzione caratteristica  $\phi : \mathbb{R}^n \rightarrow \mathbb{C}$ ,  $\phi_{\mathbf{X}}(\mathbf{u}) = e^{i\langle \mathbf{u}, \mu \rangle - \frac{1}{2}\langle \mathbf{u}, C \mathbf{u} \rangle}$ .

La definizione generalizza il caso trattato sopra.  $S^+$  indica l'insieme delle matrici simmetriche semidefinite positive.

Affinché la definizione sia ben posta occorre che  $e^{i\langle \mathbf{u}, \mu \rangle - \frac{1}{2}\langle \mathbf{u}, C \mathbf{u} \rangle}$  sia effettivamente una funzione caratteristica per qualsiasi  $\mu, C$ , altrimenti non esisterebbero vettori gaussiani con determinati parametri.

**Prop 18.2 (la definizione è ben posta)**

$$\text{Hp: } \mu = \begin{pmatrix} \mu_1 \\ \dots \\ \mu_n \end{pmatrix} \in \mathbb{R}^n, C \in S^+(n, n)$$

$$\text{Ts: } \exists \mathbf{X} : \Omega \rightarrow \mathbb{R}^n \text{ vettore aleatorio tale che } \hat{P}^{\mathbf{X}}(\mathbf{u}) = e^{i\langle \mathbf{u}, \mu \rangle - \frac{1}{2}\langle \mathbf{u}, C \mathbf{u} \rangle}$$

La dimostrazione è costruttiva:  $\mathbf{X}$  è costruito esplicitamente.

**Dim** Sia  $n = 1$ . Se  $\sigma^2 \neq 0$ ,  $X \sim \mathcal{N}(\mu, \sigma^2)$  ha legge con la funzione caratteristica richiesta. Se  $\sigma^2 = 0$ ,  $X \sim \delta_\mu$  ha  $\hat{P}^{\mathbf{X}}(\mathbf{u}) = e^{iu\mu}$ .

Sia  $n > 1$ ,  $\mu = \mathbf{0}$ ,  $C = Id$ . Considero  $Z_1, \dots, Z_n \sim \mathcal{N}(0, 1)$  indipendenti: il vettore  $\mathbf{Z} = \begin{pmatrix} Z_1 \\ \dots \\ Z_n \end{pmatrix}$  ha funzione caratteristica  $\phi_{\mathbf{Z}}(\mathbf{u}) = \phi_{Z_1}(u_1) \dots \phi_{Z_n}(u_n) = e^{-\frac{1}{2}u_1^2} \dots e^{-\frac{1}{2}u_n^2} = e^{-\frac{1}{2}\|\mathbf{u}\|^2} = e^{i\langle \mathbf{u}, \mathbf{0} \rangle - \frac{1}{2}\langle \mathbf{u}, Id \mathbf{u} \rangle}$ , quindi la tesi per  $\mu, C$  scelti è dimostrata.

Se  $C \in S^+(n, n)$ , è ben definita  $\sqrt{C}$ , che ha la proprietà  $\sqrt{C}(\sqrt{C})^T = C$  (p. 258 FT). Definisco allora  $\mathbf{X} = \sqrt{C}\mathbf{Z} + \mu$ :  $\phi_{\mathbf{X}}(\mathbf{v}) = e^{i\langle \mathbf{v}, \mu \rangle} \phi_{\mathbf{Z}}\left(\left(\sqrt{C}\right)^T \mathbf{v}\right) = e^{i\langle \mathbf{v}, \mu \rangle} e^{-\frac{1}{2}\langle (\sqrt{C})^T \mathbf{v}, (\sqrt{C})^T \mathbf{v} \rangle} = e^{i\langle \mathbf{v}, \mu \rangle} e^{-\frac{1}{2}\mathbf{v}^T \sqrt{C}(\sqrt{C})^T \mathbf{v}} = e^{i\langle \mathbf{v}, \mu \rangle - \frac{1}{2}\langle \mathbf{v}, C\mathbf{v} \rangle}$ . ■

$\mathbf{Z}$ , per com'è stato definito sopra, ha  $\mathbf{E}(\mathbf{Z}) = \mathbf{0}$  e  $C_{\mathbf{Z}} = Id$  perché ogni componente ha  $\mu_i = 0, \sigma_i^2 = 1$ , e le componenti sono tutte indipendenti. Si scrive allora  $\mathbf{Z} \sim \mathcal{N}(\mathbf{0}, Id)$ . Dalla dimostrazione si ricava che ogni vettore gaussiano ha la stessa legge della trasformazione lineare affine di un vettore gaussiano standard (è la generalizzazione della standardizzazione vista per v. a. normali), cioè se  $\mathbf{X}$  è gaussiano con parametri  $\mu, C$  ( $\mathbf{X} \sim \mathcal{N}(\mu, C)$ ),  $\mathbf{X} = \sqrt{C}\mathbf{Z} + \mu$  con  $\mathbf{Z} \sim \mathcal{N}(\mathbf{0}, Id)$ .

### Prop 18.3 (proprietà dei vettori gaussiani)

Hp:  $\mu \in \mathbb{R}^n, C \in S^+(n, n), \mathbf{X} \sim \mathcal{N}(\mu, C)$

Ts: (1)  $\mathbf{E}(\mathbf{X}) = \mu, C_{\mathbf{X}} = C$

(2)  $\forall i = 1, \dots, n X_i \sim \mathcal{N}(\mu_i, c_{ii})$

(3) se  $A \in M_{\mathbb{R}}(k, n), \mathbf{b} \in \mathbb{R}^k, \mathbf{Y} = A\mathbf{X} + \mathbf{b}$ , allora  $\mathbf{Y} \sim \mathcal{N}(A\mu + \mathbf{b}, AC_{\mathbf{X}}A^T)$

(4)  $X_1, \dots, X_n$  sono indipendenti  $\iff cov(X_i, X_j) = 0 \forall i, j = 1, \dots, n$

(2) afferma che ogni componente di un vettore gaussiano è una v. a. normale. (3) afferma che con trasformazioni lineari affini di gaussiane si ottengono ancora gaussiane: quindi in particolare ogni sottovettore di  $\mathbf{X}$ , potendo essere ottenuto come  $A\mathbf{X}$ , è ancora gaussiano. (4) significa che, a differenza di quanto accade in generale, le componenti di un vettore gaussiano sono indipendenti se e solo se  $C_{\mathbf{X}}$  è diagonale.

**Dim** (1) E' noto che  $\mathbf{X}$  ha la stessa distribuzione di  $\sqrt{C}\mathbf{Z} + \mu$  con  $\mathbf{Z} \sim \mathcal{N}(\mathbf{0}, Id)$  (non posso scrivere l'uguaglianza perché  $\mathbf{X}$  potrebbe essere discreta ed essere definita su uno spazio di probabilità molto piccolo su cui non è definito  $\mathbf{Z}$ , a differenza di prima in cui erano definite sullo stesso spazio di probabilità 50:24). Allora  $\mathbf{E}(\mathbf{X}) = \sqrt{C}\mathbf{E}(\mathbf{Z}) + \mu = \mu$  e  $C_{\mathbf{X}} = \sqrt{C}C_{\mathbf{Z}}(\sqrt{C})^T = \sqrt{C}Id(\sqrt{C})^T = C$ . Si può dimostrare alternativamente con la formula dei momenti.

(2) Per quanto visto sopra  $X_i = \mathbf{e}_i^T \mathbf{X}$ , quindi  $\phi_{X_i}(v_i) = \phi_{\mathbf{X}}(\mathbf{e}_i^T v_i) = e^{iv_i \mu_i - \frac{1}{2}C_{ii}v_i^2}$ . Per unicità di  $\phi$  allora  $X_i \sim \mathcal{N}(\mu_i, C_{ii})$ .

(3) E' noto che  $\phi_{\mathbf{Y}}(\mathbf{v}) = e^{i\langle \mathbf{v}, \mathbf{b} \rangle} \phi_{\mathbf{X}}(A^T \mathbf{v})$ , quindi, essendo  $\phi_{\mathbf{X}}(\mathbf{u}) = e^{i\langle \mathbf{u}, \mu \rangle - \frac{1}{2}\langle \mathbf{u}, C\mathbf{u} \rangle}$ , si ottiene  $\phi_{\mathbf{Y}}(\mathbf{v}) = e^{i\langle \mathbf{v}, \mathbf{b} \rangle} e^{i\langle A^T \mathbf{v}, \mu \rangle - \frac{1}{2}\langle A^T \mathbf{v}, CA^T \mathbf{v} \rangle}$ . Vale  $e^{i\langle \mathbf{v}, \mathbf{b} \rangle} e^{i\langle A^T \mathbf{v}, \mu \rangle} = e^{i\langle \mathbf{v}, A\mu \rangle} e^{i\langle \mathbf{v}, \mathbf{b} \rangle} = e^{i\langle \mathbf{v}, A\mu + \mathbf{b} \rangle}$ , mentre  $\langle A^T \mathbf{v}, CA^T \mathbf{v} \rangle = \langle \mathbf{v}, ACA^T \mathbf{v} \rangle$ , quindi  $\phi_{\mathbf{Y}}(\mathbf{v}) = e^{i\langle \mathbf{v}, A\mu + \mathbf{b} \rangle - \frac{1}{2}\langle \mathbf{v}, ACA^T \mathbf{v} \rangle}$ , quindi  $\mathbf{Y} \sim \mathcal{N}(A\mu + \mathbf{b}, ACA^T)$ .

(4) L'implicazione da sinistra a destra deriva da 15.4, perché l'indipendenza della famiglia implica anche l'indipendenza a coppie e quindi covarianza nulla.



Dimostro l'implicazione da destra a sinistra:  $C = \begin{bmatrix} \sigma_1^2 & \dots & 0 \\ \dots & \dots & \dots \\ 0 & \dots & \sigma_n^2 \end{bmatrix}$ , per cui  $\phi_{\mathbf{X}}(\mathbf{u}) = e^{i\langle \mathbf{u}, \mu \rangle - \frac{1}{2} \langle \mathbf{u}, C \mathbf{u} \rangle} = e^{iu_1\mu_1 + \dots + iu_n\mu_n} e^{-\frac{1}{2}(\sigma_1^2 u_1^2 + \dots + \sigma_n^2 u_n^2)} = e^{iu_1\mu_1 - \frac{1}{2}\sigma_1^2 u_1^2} \dots e^{iu_n\mu_n - \frac{1}{2}\sigma_n^2 u_n^2}$ . Poiché ogni fattore è la funzione caratteristica di una componente, significa che  $\phi_{\mathbf{X}}$  si fattorizza, e quindi  $X_1, \dots, X_n$  sono indipendenti per 17.2 (3). ■

1. Dato  $\mathbf{X} \sim \mathcal{N}(\mu, C)$ , se  $\exists k \neq h : \text{cov}(X_k, X_h) = 0$ , allora  $\begin{pmatrix} X_k \\ X_h \end{pmatrix}$  è gaussiano (perché trasformazione lineare di  $\mathbf{X}$ ) con  $X_k \perp X_h$  (perché la nuova matrice di covarianza sarà diagonale). Se e. g.

$$n = 3, \text{ sia } C = \begin{bmatrix} \sigma_1^2 & * & * \\ * & \sigma_2^2 & 0 \\ * & 0 & \sigma_3^2 \end{bmatrix}. \quad \begin{pmatrix} X_2 \\ X_3 \end{pmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix}, \text{ quindi } C_{(X_2, X_3)} = A C A^T =$$

$$\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \sigma_1^2 & * & * \\ * & \sigma_2^2 & 0 \\ * & 0 & \sigma_3^2 \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} * & * \\ \sigma_2^2 & 0 \\ 0 & \sigma_3^2 \end{bmatrix} = \begin{bmatrix} \sigma_2^2 & 0 \\ 0 & \sigma_3^2 \end{bmatrix}. \text{ Si applica allora 18.3}$$

(4) a  $(X_2, X_3)$  e si conclude che  $X_2 \perp X_3$ .

Ha senso chiedersi se vale l'implicazione opposta in 18.3 (2): in generale no.

#### Prop 18.4

(1) Hp:  $X_1, \dots, X_n : \Omega \rightarrow \mathbb{R}^n$  variabili aleatorie

Ts:  $X_1, \dots, X_n$  sono indipendenti e  $X_k \sim \mathcal{N}(\mu_k, \sigma_k^2) \forall k = 1, \dots, n \iff$

$$\mathbf{X} = \begin{pmatrix} X_1 \\ \dots \\ X_n \end{pmatrix} \sim \mathcal{N}(\mu_X, C_X) \text{ con } \mu_X = \begin{pmatrix} \mu_1 \\ \dots \\ \mu_n \end{pmatrix} \text{ e } C_{\mathbf{X}} = \begin{bmatrix} \sigma_1^2 & \dots & 0 \\ \dots & \dots & \dots \\ 0 & \dots & \sigma_n^2 \end{bmatrix}$$

(2) Hp:  $\mathbf{X}_1 : \Omega \rightarrow \mathbb{R}^{n_1}, \mathbf{X}_2 : \Omega \rightarrow \mathbb{R}^{n_2}$  vettori aleatorii

Ts:  $\mathbf{X}_1 \sim \mathcal{N}(\mu_1, C_1), \mathbf{X}_2 \sim \mathcal{N}(\mu_2, C_2), \mathbf{X}_1 \perp \mathbf{X}_2 \iff$

$$\mathbf{X} = \begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{bmatrix} C_1 & 0_M \\ 0_M & C_2 \end{bmatrix}\right)$$

L'implicazione da destra a sinistra in (1) vale per quanto visto sopra; quella da sinistra a destra si basa sulla fattorizzazione della funzione caratteristica. La tesi 2, in cui figura come matrice di covarianza una matrice diagonale a blocchi, è la 1 nel caso di  $n_1 = 1, n_2 = 1$ .

1. Siano  $X_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$ ,  $X_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$ ,  $X_1 \perp X_2$ . Allora  $\alpha X_1 + \beta X_2 + c \sim \mathcal{N}(\alpha\mu_1 + \beta\mu_2 + c, \alpha^2\sigma_1^2 + \beta^2\sigma_2^2)$  perché per 18.4 (1)  $\mathbf{X} = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$  è gaussiano con  $\mu_{\mathbf{X}} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}$  e  $C_{\mathbf{X}} = \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix}$ , quindi  $Y = [\alpha|\beta] \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} + c$  per 18.3 (3) è normale con parametri  $\alpha\mu_1 + \beta\mu_2 + c$  e  $\alpha^2\sigma_1^2 + \beta^2\sigma_2^2$ .

### Teo 18.5 (caratterizzazione dei vettori gaussiani)

Hp:  $(\Omega, \mathcal{A}, \mathbf{P})$  spazio di probabilità,  $\mathbf{X} : \Omega \rightarrow \mathbb{R}^n$  vettore aleatorio

Ts:  $\exists \mu \in \mathbb{R}^n, C \in S^+(n, n) : \mathbf{X} \sim \mathcal{N}(\mu, C) \iff \forall \mathbf{a} \in \mathbb{R}^n \langle \mathbf{a}, \mathbf{X} \rangle$  è gaussiana

L'implicazione da sinistra a destra è ovvia:  $\langle \mathbf{a}, \mathbf{X} \rangle$  è gaussiana con parametri  $\mathbf{a}^T \mu$ ,  $\mathbf{a}^T C \mathbf{a}$ . L'implicazione da destra a sinistra ha un'ipotesi molto più forte della richiesta di  $X_1, \dots, X_n$  gaussiane: deve esserlo ogni loro combinazione lineare, non solo le combinazioni lineari del tipo  $\langle \mathbf{e}_k, \mathbf{X} \rangle$ .

**Dim** L'implicazione da sinistra a destra è ovvia per 18.3, che si usa con  $A = \mathbf{a}^T$ , quindi  $\langle \mathbf{a}, \mathbf{X} \rangle$  è gaussiana.

Se  $\forall \mathbf{a} \in \mathbb{R}^n Y = \langle \mathbf{a}, \mathbf{X} \rangle$  è gaussiana e si pone  $\mu := \mathbf{E}(\mathbf{X})$ ,  $C := C_{\mathbf{X}}$ , si ha  $\phi_Y(v) = \phi_{\mathbf{X}}(\mathbf{a}^T v) = e^{i\langle v, \mu_Y \rangle - \frac{1}{2} \sigma_Y^2 v^2} = e^{i\langle v, \mathbf{a}^T \mu_{\mathbf{X}} \rangle - \frac{1}{2} \mathbf{a}^T C \mathbf{a} v^2}$ , quindi se  $v = 1$  si ha  $\phi_{\mathbf{X}}(\mathbf{a}) = e^{i\langle \mathbf{a}, \mu_{\mathbf{X}} \rangle - \frac{1}{2} \mathbf{a}^T C \mathbf{a}}$ . ■

Ora ci chiediamo se i vettori gaussiani sono discreti o assolutamente continui. E' noto che per  $n = 1$  se  $\sigma^2 = 0$   $X$  gaussiana è discreta con supporto  $S = \mu$ , mentre se  $\sigma^2 \neq 0$   $X$  gaussiana è normale secondo la vecchia definizione, con densità  $f_X(t) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(t-\mu)^2}$ .

Si è visto in passato che  $\mathbf{X} \in \text{col} C_X + \mu$  q. c., e che se  $\mathbf{X}$  è assolutamente continuo allora  $\det C_{\mathbf{X}} > 0$ . Il seguente teorema afferma che per i vettori gaussiani è possibile invertire l'ultima implicazione.

[Se  $\mathbf{X} \sim \mathcal{N}(\mu, C)$ ,  $h(\mathbf{X})$  se  $h$  non è lineare affine non è gaussiana??. Quindi ogni vettore gaussiano non può avere come immagine  $G_f$  a meno che lineare affine, cioè sottospazio lineare affine. I vettori gaussiani hanno come immagine sottospazi vettoriali affini. Per questo vale il se e solo se: non posso avere i casi  $\begin{pmatrix} X \\ f(X) \end{pmatrix}$  in cui  $\det > 0$  ma vettore non assolutamente continuo.

Se  $\mathbf{X} \sim \mathcal{N}(\mu, C)$ , vale  $\mathbf{X} \sim \mu + \sqrt{C}\mathbf{Z}$ : ogni componente di  $\mathbf{Z}$  è una normale standard che assume con probabilità non nulla ogni valore reale; poiché le componenti sono indipendenti  $\mathbf{Z}$  assume con probabilità non nulla ogni valore di  $\mathbb{R}^n$ . Se  $\sqrt{C}$  è invertibile anche  $\mu + \sqrt{C}\mathbf{Z}$  assume con probabilità non nulla ogni valore di  $\mathbb{R}^n$ . Se invece  $\sqrt{C}$  non è invertibile  $\mu + \sqrt{C}\mathbf{Z}$  ha come supporto un sottospazio vettoriale di dimensione pari a  $r(\sqrt{C})$ . Questo implica che ogni vettore gaussiano non può avere supporto qualsiasi, ma solo un sottospazio vettoriale affine di  $\mathbb{R}^n$ ; se  $n = 2$ , non può essere un grafico di funzione, a meno che non si tratti di una funzione lineare affine, né può

avere come supporto un cerchio o un quadrato. E' questo fatto, connesso alla gaussianità, che rende possibile l'equivalenza tra assoluta continuità del vettore e positività del determinante della matrice di covarianza: i casi in cui il determinante della matrice di covarianza di un vettore è positivo, ma il vettore non è assolutamente continuo, non possono verificarsi se un vettore è gaussiano, perché i vettori per cui accade ciò sono del tipo  $\begin{pmatrix} X \\ f(X) \end{pmatrix}$ .]

1. Nonostante non esistano vettori gaussiani del tipo  $\begin{pmatrix} X \\ f(X) \end{pmatrix}$  con  $f$  non lineare affine, è comunque possibile che, data  $X \sim \mathcal{N}(0, 1)$ ,  $f(X)$  sia gaussiana con  $f$  non lineare affine. Denominata  $F$  la funzione di ripartizione di una  $\chi^2(1)$ , si consideri  $f(X) = \Phi^{-1}(F(X^2))$ :  $\mathbf{P}(f(X) \leq t) = \mathbf{P}(\Phi^{-1}(F(X^2)) \leq t) = \mathbf{P}(F(X^2) \leq \Phi(t))$ . Ma è noto che  $F(X^2) \sim \mathcal{U}((0, 1))$ , quindi  $\mathbf{P}(f(X) \leq t) = \Phi(t)$ , cioè  $f(X) \sim \mathcal{N}(0, 1)$ .

### Teo (normalizzazione di un vettore gaussiano)

$$\text{Hp: } \mu \in \mathbb{R}^n, C \in S^+(n, n), \mathbf{X} \sim \mathcal{N}(\mu, C)$$

$$\text{Ts: } \exists Y_1, \dots, Y_n \text{ v.a. indipendenti tali che}$$

$$Y_i \sim \mathcal{N}(0, \lambda_i) \quad \forall i, \lambda_i \geq 0 \quad \forall i \text{ e } \exists U \text{ ortogonale : } \mathbf{X} = \mu + U\mathbf{Y}$$

$$\text{La tesi implica che } \mathbf{Y} \sim \mathcal{N}\left(0, \begin{bmatrix} \lambda_1 & \dots & 0 \\ \dots & \dots & \dots \\ 0 & \dots & \lambda_n \end{bmatrix}\right); \text{ questa normalizzazione è diversa da } \mu + \sqrt{C}\mathbf{Z}, \text{ perché}$$

quest'ultimo è uguale a  $\mathbf{X}$  solo in legge.

Il teorema esprime il fatto che ogni vettore gaussiano nasce dalla combinazione lineare di v. a. gaussiane indipendenti.

**Dim** Poiché  $C$  è simmetrica, per il teorema spettrale  $\exists U$  ortogonale e  $D$  diagonale tale che  $C = UDU^T$ . Definendo  $\mathbf{Y} = U^T(\mathbf{X} - \mu)$ , si ha che  $\mathbf{Y} \sim \mathcal{N}(0, U^T C U)$ , dato che è una trasformazione lineare affine di  $\mathbf{X}$  gaussiano. Vale inoltre  $U^T C U = D$  diagonale: dunque effettivamente esistono  $Y_1, \dots, Y_n$  indipendenti con  $Y_i \sim \mathcal{N}(0, \lambda_i) \quad \forall i, \lambda_i \geq 0 \quad \forall i$ . ■

### Teo 18.5 (assoluta continuità dei vettori gaussiani)

$$\text{Hp: } \mu \in \mathbb{R}^n, C \in S^+(n, n), \mathbf{X} \sim \mathcal{N}(\mu, C)$$

$$\text{Ts: } \mathbf{X} \text{ è assolutamente continuo } \iff \det C_{\mathbf{X}} > 0, \text{ e in tal}$$

$$\text{caso } f_{\mathbf{X}}(x_1, \dots, x_n) = \frac{1}{(2\pi)^{n/2} \sqrt{\det C_X}} e^{-\frac{1}{2} \langle \mathbf{x} - \mu, C^{-1}(\mathbf{x} - \mu) \rangle}$$

Si noti che la formula per la densità è ben posta perché  $C_X$  è invertibile. Nel caso  $n = 1$  il teorema si riduce a dire che una v. a. con  $\phi_X(u) = e^{iu\mu - \frac{1}{2}\sigma^2 u^2}$  è assolutamente continua se e solo se  $\sigma^2 \neq 0$ , e in tal caso  $f_X(x) = \frac{1}{(2\pi)^{1/2}\sqrt{\sigma^2}} e^{-\frac{1}{2}(x-\mu)\frac{1}{\sigma^2}(x-\mu)}$ : cioè se  $X$  non è una delta di Dirac ma una normale secondo la vecchia definizione.

Il teorema implica che  $\mathbf{Z} \sim \mathcal{N}(\mathbf{0}, Id)$  è assolutamente continua, dato che  $\det Id = 1$ .

**Dim** E' noto che  $\mathbf{X} = \sqrt{C}\mathbf{Z} + \mu$ : poiché esiste  $C^{-1}$ , posso scegliere  $\sqrt{C}$  invertibile: in tal caso la trasformazione scritta è un diffeomorfismo. Allora anche  $\mathbf{X}$  ammette densità continua.

Sia  $\det C_{\mathbf{X}} > 0$ . Allora  $\exists \mathbf{Y} \sim \mathcal{N}\left(0, \begin{bmatrix} \lambda_1 & \dots & 0 \\ \dots & \dots & \dots \\ 0 & \dots & \lambda_n \end{bmatrix}\right)$ ,  $U$  ortogonale e  $D$  diagonale tale che  $C = UDU^T$  e

$\mathbf{X} = \mu + U\mathbf{Y}$ . Poiché  $Y_i \sim \mathcal{N}(0, \lambda_i)$  e le  $Y_i$  sono indipendenti,  $\mathbf{Y}$  ammette la densità  $f_{\mathbf{Y}}(\mathbf{y}) = \prod_{j=1}^n \frac{1}{\sqrt{2\pi\lambda_j}} e^{-\frac{1}{2\lambda_j} y_j^2} = \frac{1}{2\pi^{n/2}\sqrt{\det C_{\mathbf{X}}}} e^{-\frac{1}{2\det C_{\mathbf{X}}} \mathbf{y}^T \mathbf{y}}$ : allora  $\mathbf{X}$  ha la densità (...)

■

Ricapitolando, se  $\det C > 0$ ,  $\exists C^{-1}$ ,  $\dim(\text{col}C) = n$  e  $\mathbf{X}$  è assolutamente continuo. Se invece  $\det C = 0$ ,  $\dim(\text{col}C) < n$ : quindi la trasformazione  $\mathbf{X} = \sqrt{C}\mathbf{Z} + \mu$  non è di rango massimo ( $\det \sqrt{C} = 0$ ).

Sia  $\mathbf{X} \sim \mathcal{N}(\mu, C)$  con  $\det C = 0$  (e affinché questo accada è sufficiente un solo autovalore nullo): allora  $\ker C$  non contiene il solo vettore nullo, cioè  $\exists \hat{\mathbf{a}} \neq \mathbf{0} : C\hat{\mathbf{a}} = \mathbf{0}$ . Poiché  $\mathbf{X}$  è gaussiano,  $\hat{\mathbf{a}}^T \mathbf{X}$  è gaussiana con parametri  $\hat{\mathbf{a}}^T \mu$ ,  $\hat{\mathbf{a}} C \hat{\mathbf{a}}^T$ , cioè  $\hat{\mathbf{a}}^T \mathbf{X} \sim \mathcal{N}(\hat{\mathbf{a}}^T \mu, 0_M)$ . Questo implica che, poiché  $\hat{\mathbf{a}}^T \mathbf{X} \in \text{col}0_M + \hat{\mathbf{a}}^T \mu$ ,  $\hat{\mathbf{a}}^T \mathbf{X} = \hat{\mathbf{a}}^T \mu$  q. c., cioè  $\hat{\mathbf{a}}^T \mathbf{X}$  è una v. a. degenera. Questo è equivalente a dire che  $\langle \hat{\mathbf{a}}, \mathbf{X} - \mu \rangle = 0$ , cioè  $\mathbf{P}(\mathbf{X} \in H = \{\mathbf{x} \in \mathbb{R}^n : \langle \mathbf{x} - \mu, \hat{\mathbf{a}} \rangle = 0\}) = 1$ :  $\mathbf{X}$  appartiene quasi certamente all'iperpiano definito dall'equazione  $\langle \mathbf{x} - \mu, \hat{\mathbf{a}} \rangle = 0$ , che è un sottospazio vettoriale affine di  $\mathbb{R}^n$ .

1. Sia  $\mathbf{X} = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}\right)$ .  $X_1, X_2 \sim \mathcal{N}(1, 1)$ ;  $\det C = 0$ , quindi  $\dim \ker C = 1$ . Una base normale di  $\ker C$  è  $\left\{ \begin{pmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{pmatrix} \right\}$ . Allora  $\mathbf{P}(\mathbf{X} \in H = \{\mathbf{x} \in \mathbb{R}^2 : \frac{x_1-1}{\sqrt{2}} - \frac{x_2-1}{\sqrt{2}} = 0\}) = 1$ , cioè  $\mathbf{X}$  appartiene q. c. al piano definito da  $x_1 = x_2$ , che ha come base  $t \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ : questo significa che  $\mathbf{P}(X_1 = X_2) = 1$ .

Se e. g. voglio calcolare  $\mathbf{P}(\|\mathbf{X}\|^2 \leq 1)$ , ho  $\mathbf{P}(X_1^2 + X_2^2 \leq 1) = \mathbf{P}(X_1^2 + X_2^2 \leq 1, X_1 = X_2)$  (se  $\mathbf{P}(B) = 1$ ,  $\mathbf{P}(A \cap B) = \mathbf{P}(A)$ ), quindi si ottiene  $\mathbf{P}(2X_1^2 \leq 1, X_1 = X_2) = \mathbf{P}(2X_1^2 \leq 1) = \mathbf{P}\left(X_1 \in \left[-\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right]\right)$ , che si può ricavare attraverso  $f_{X_1}$ . Quindi conoscere l'iperpiano cui appartiene q. c.  $\mathbf{X}$  permette di calcolare le

probabilità che lo coinvolgono riducendosi al più grande sottospazio in cui esiste una densità, che seleziono usando le componenti che vedo trovando la più grande sottomatrice di  $C$  con determinante non nullo.

## 12.1 Funzioni non lineari di vettori gaussiani e statistiche campionarie

**Def** Dato  $\mathbf{Z} \sim \mathcal{N}(\mathbf{0}, Id)$  vettore gaussiano standard a valori in  $\mathbb{R}^n$ , si dice legge chi-quadro a  $n$  gradi di libertà, e si indica con  $\chi^2(n)$ , la legge della variabile aleatoria  $V_n = \|\mathbf{Z}\|^2 = Z_1^2 + \dots + Z_n^2$ .

Poiché le  $Z_i$  sono indipendenti e quindi anche le  $Z_i^2$ , si ricava che  $\mathbf{E}(V_n) = n$  e  $var(V_n) = nvar(Z_1^2) = n(\mathbf{E}(Z_1^4) - \mathbf{E}^2(Z_1^2)) = 2n$ . Una chi-quadro a  $n$  gradi di libertà è quindi uguale in legge alla somma di  $n$  gaussiane standard ed è non negativa.

### Prop (densità della chi-quadro)

$$\text{Hp: } V_n \sim \chi^2(n)$$

Ts:  $V_n$  è assolutamente continua con

$$\text{densità } f_{V_n}(t) = \frac{\left(\frac{1}{2}\right)^{\frac{n}{2}}}{\Gamma\left(\frac{n}{2}\right)} t^{\frac{n}{2}-1} e^{-\frac{t}{2}} I_{(0,+\infty)}(t)$$

Quindi  $\Gamma\left(\frac{n}{2}, \frac{1}{2}\right) = \chi^2(n)$ :  $var(V_n) = \frac{\alpha}{\chi^2} = 2n$ , essendo la varianza di una gamma nota.

1. Se  $X_1, \dots, X_n \sim iid\mathcal{N}(\mu, \sigma^2)$ ,  $\sum_{k=1}^n \frac{(X_k - \mu)^2}{\sigma^2}$  è la somma di  $n$  gaussiane standard indipendenti al quadrato e quindi ha legge  $\chi^2(n)$ .

**Def** Data  $Z \sim \mathcal{N}(0, 1)$  e  $V_n \sim \chi^2(n)$ ,  $V_n \perp Z$ , si dice legge  $t$  di Student a  $n$  gradi di libertà, e si indica con  $t(n)$ , la legge della variabile aleatoria  $T = \frac{Z}{\sqrt{\frac{V_n}{n}}}$ .

### Prop (densità della $t$ di Student)

$$\text{Hp: } T \sim t(n)$$

Ts:  $T$  è assolutamente continua con

$$\text{densità } f_T(t) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\sqrt{n\pi}} \left(1 + \frac{t^2}{n}\right)^{-\frac{n+1}{2}}$$

$$\lim_{n \rightarrow +\infty} \left( f_T(t) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\sqrt{n\pi}} \left(1 + \frac{t^2}{n}\right)^{-\frac{n+1}{2}} = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\sqrt{n\pi}} \left[ \left(1 + \frac{t^2}{n}\right)^n \right]^{-\frac{1}{2}} \left(1 + \frac{t^2}{n}\right)^{-\frac{1}{2}} \right) = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} \quad \forall t \in \mathbb{R} \text{ (dim),}$$

cioè la densità di una  $t$  di Student, al crescere del numero di gradi di libertà, si avvicina alla densità di una normale standard (quindi  $T_n \rightarrow Z$ ). Se  $n = 1$ ,  $f_{T_n}(t) = \frac{1}{\pi(1+t^2)}$ , che è la densità di una v. a. di Cauchy: quindi  $\exists \mathbf{E}(T_1)$ .

Se invece  $n > 1$ ,  $\exists \mathbf{E}(T_n) = 0$  perché  $f_{T_n}$  è simmetrica. Se  $n = 1, 2$   $var(T_n) \exists$ ; se  $n > 2$   $var(T_n) = \frac{n}{n-2}$ . In generale esistono solo i momenti per  $k < n$ :  $\mathbf{E}(T_n^k) < +\infty \iff k < n$ .

Se definisco media campionaria  $\bar{X}_n = \frac{X_1 + \dots + X_n}{n}$ , vale per linearità che  $\mathbf{E}(\bar{X}_n) = \frac{1}{n}n\mu = \mu$ . Per 15.4 (7) vale inoltre  $\text{var}(\sum_{k=1}^n X_k) = \sum_{k=1}^n \text{var}(X_k) = n\sigma^2$ , per cui  $\text{var}(\bar{X}_n) = \frac{1}{n^2}n\sigma^2 = \frac{\sigma^2}{n}$ .

Sia  $\begin{pmatrix} X_1 \\ \dots \\ X_n \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mu \\ \dots \\ \mu \end{pmatrix}, \begin{bmatrix} \sigma^2 & \dots & 0 \\ \dots & \dots & \dots \\ 0 & \dots & \sigma^2 \end{bmatrix}\right)$  un campione casuale, cioè un insieme di v. a. iid con distribuzione  $\mathcal{N}(\mu, \sigma^2)$ . Siano, come già visto, la media campionaria  $\bar{X}_n = \frac{X_1 + \dots + X_n}{n}$ , la varianza campionaria  $S_n^2 = \frac{1}{n-1} \sum_{k=1}^n (X_k - \bar{X}_n)^2$ .

, vale  $S_n^2 = \frac{1}{n-1} \sum_{k=1}^n (X_k^2 + \bar{X}_n^2 - 2X_k\bar{X}_n)$ . Poiché  $\text{var}(X_k) = \mathbf{E}(X_k^2) - \mathbf{E}^2(X_k)$ , vale  $\mathbf{E}(X_k^2) = \sigma^2 + \mu^2$ , e  $\mathbf{E}(X_k\bar{X}_n) = \text{cov}(X_k, \bar{X}_n) + \mathbf{E}(X_k)\mathbf{E}(\bar{X}_n) = \frac{1}{n}\text{var}(X_k) + \mathbf{E}(X_k)\mathbf{E}(\bar{X}_n)$ : per linearità  $\mathbf{E}(S_n^2) = \frac{1}{n-1} \left( n(\sigma^2 + \mu^2) + n\left(\frac{\sigma^2}{n} + \mu^2\right) \right) = \frac{1}{n-1} (n\sigma^2 + \sigma^2 - 2\sigma^2) = \sigma^2$ .

### Teo 19.1 (relazioni tra statistiche campionarie)

Hp:  $X_1, \dots, X_n$  è un campione casuale estratto da  $X \sim \mathcal{N}(\mu, \sigma^2)$

$$\text{Ts: (1) } \bar{X}_n \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$$

$$(2) \frac{n-1}{\sigma^2} S_n^2 \sim \chi^2(n-1)$$

$$(3) \bar{X}_n \perp S_n^2$$

$$(4) \frac{\bar{X}_n - \mu}{\sqrt{\frac{S_n^2}{n}}} \sim t(n-1)$$

Da (2) si ricava che  $\text{var}(S_n^2) = \frac{2\sigma^4}{n-1}$  e  $\phi_{S_n^2}(u) = \left(1 - \frac{2i\sigma^2}{n-1}u\right)^{-\frac{n-1}{2}}$ .

**Dim** (1)  $\mathbf{E}\left(\frac{X_1 + \dots + X_n}{n}\right) = \frac{1}{n} \sum_{i=1}^n \mathbf{E}(X_i) = \frac{1}{n}n\mu = \mu$ .  $\text{var}\left(\frac{X_1 + \dots + X_n}{n}\right) = \frac{1}{n^2} \sum_{i=1}^n \text{var}(X_i) = \frac{1}{n^2}n\sigma^2 = \frac{\sigma^2}{n}$ .

(2) (3) Suppongo  $\mu = 0, \sigma^2 = 1$ . Allora  $\begin{pmatrix} X_1 \\ \dots \\ X_n \end{pmatrix} \sim \mathcal{N}(\mathbf{0}, Id)$  perché le componenti sono indipendenti.

Considero  $U \in M_{\mathbb{R}}(n, n)$  ortogonale, avente la prima riga  $\left[\frac{1}{\sqrt{n}} | \dots | \frac{1}{\sqrt{n}}\right]$  e tutte le altre righe ortonormali alla prima. Costruisco  $\mathbf{Z} = U\mathbf{X}$ : sarà  $\mathbf{Z} \sim \mathcal{N}(U \cdot \mathbf{0}, UIdU^T) = \mathcal{N}(\mathbf{0}, Id)$ . Inoltre  $Z_1 = \frac{1}{\sqrt{n}}(X_1 + \dots + X_n) = \sqrt{n}\bar{X}_n$ .  $\sum_{k=2}^n Z_k^2 = \|\mathbf{Z}\|^2 - Z_1^2 = f(Z_2, \dots, Z_n)$ , che è indipendente da  $Z_1$  perché  $\mathbf{Z}$  ha matrice  $C$  diagonale e quindi componenti indipendenti. Si è già visto che  $(n-1)S_n^2 = \sum_{k=1}^n (X_k - \bar{X}_n)^2 = \|\mathbf{X}\|^2 - n\bar{X}_n^2$ . Ma  $\|\mathbf{Z}\|^2 = \langle U\mathbf{X}, U\mathbf{X} \rangle = \|\mathbf{X}\|^2$  perché  $U$  è ortogonale, quindi  $(n-1)S_n^2 = \|\mathbf{Z}\|^2 - n\bar{X}_n^2 = \|\mathbf{Z}\|^2 - Z_1^2 = \sum_{k=2}^n Z_k^2$  perché  $\frac{Z_1}{\sqrt{n}} = \bar{X}_n$ .  $\sum_{k=2}^n Z_k^2 \sim \chi^2(n-1)$  per definizione di  $\chi^2$ , ed è dimostrata (2) perché sono nel caso di  $\sigma^2 = 1$ .

Ne segue che  $\sum_{k=1}^n (X_k - \bar{X}_n)^2 = \sum_{k=2}^n Z_k^2$  è indipendente da  $Z_1 = \sqrt{n}\bar{X}_n$ , per cui  $S_n^2 = \frac{1}{n-1} \sum_{k=1}^n (X_k - \bar{X}_n)^2$  è indipendente da  $\bar{X}_n = \frac{Z_1}{\sqrt{n}}$  (applicare funzioni misurabili a due v. a. indipendenti preserva l'indipendenza). Con

la standardizzazione si mostra che  $\frac{n-1}{\sigma^2} S_n^2 \sim \chi^2(n-1)$  e  $\bar{X}_n \perp S_n^2$ .

(4) Voglio scrivere  $\frac{\bar{X}_n - \mu}{\sqrt{\frac{S_n^2}{n}}}$  come  $\frac{Z}{\sqrt{\frac{V_{n-1}}{n-1}}}$ . Vale  $Z = \frac{\bar{X}_n - \mu}{\sqrt{\frac{\sigma^2}{n}}} \sim N(0, 1)$  e  $V_{n-1} = \frac{(n-1)S_{n-1}^2}{\sigma^2} \sim \chi^2(n-1)$ , ma  $\frac{\bar{X}_n - \mu}{\sqrt{\frac{S_n^2}{n}}}$  è proprio uguale a  $\frac{Z}{\sqrt{\frac{V_{n-1}}{n-1}}}$ , quindi  $\frac{\bar{X}_n - \mu}{\sqrt{\frac{S_n^2}{n}}} \sim t(n-1)$ . ■

## 13 Leggi condizionali

Sia  $\begin{pmatrix} X \\ Y \end{pmatrix} : \Omega \rightarrow \mathbb{R}^2$  un vettore aleatorio discreto con densità congiunta  $p_{(X,Y)}$  e densità marginali  $p_X, p_Y$ , con supporti delle componenti  $S_X, S_Y$  rispettivamente. Supponiamo di venire a sapere che  $X = x_i$ . In assenza di questa informazione, la legge di  $Y$  si calcola come  $\mathbf{P}(Y \in B)$ , partendo dalla terna  $(\Omega, \mathcal{A}, \mathbf{P})$ ; avendola, per calcolare probabilità legate a  $Y$  ha senso usare  $\mathbf{P}(Y \in B | X = x_i)$ , il che definisce una nuova legge. In generale, dato  $x \in S_X$ , la funzione avente come dominio  $\mathcal{B}(\mathbb{R})$  che a ogni  $B \in \mathcal{B}(\mathbb{R})$  associa  $\mathbf{P}(Y \in B | X = x)$  è una misura di probabilità (si è già visto che la probabilità condizionata è una misura di probabilità nel suo primo argomento).  $\mathbf{P}(Y \in B | X = x)$  si indica con  $P^{Y|X}(B|x)$ , con  $x \in S_X$  e  $B \in \mathcal{B}(\mathbb{R})$ .

La v. a.  $Y | X = x$  non è ben definita in sé: è la notazione usata per indicare le v. a. con la legge  $\mathbf{P}(Y \in B | X = x)$ .

$Y$  è discreta anche rispetto a questa misura di probabilità, perché  $\mathbf{P}(Y \in S_Y | X = x) = \frac{\mathbf{P}(Y \in S_Y, X = x)}{\mathbf{P}(X = x)} = \frac{\mathbf{P}(X = x)}{\mathbf{P}(X = x)} = 1$ , dato che  $\mathbf{P}(A \cap B) = \mathbf{P}(B)$  se  $\mathbf{P}(A) = 1$ .

Calcolo la densità discreta di  $Y$  rispetto alla nuova misura: dato  $y \in S_Y$ ,  $p_{Y|X}(y|x) = \mathbf{P}(Y = y | X = x) = \frac{p_{(X,Y)}(x,y)}{p_X(x)}$ . E' facile mostrare che è una densità discreta.

Il valore atteso di  $Y$  rispetto alla nuova misura di probabilità  $P^{Y|X}(B|x)$ , che si indica con  $\mathbf{E}(Y | X = x)$ , è  $\int_{\Omega} Y(\omega) P(d\omega | X = x) = \int_{\mathbb{R}} y P^{Y|X}(dy|x) = \sum_{y \in S_Y} y p_{Y|X}(y|x)$ . In generale, data  $h : \mathbb{R} \rightarrow \mathbb{R}$  boreliana e limitata oppure nonnegativa,  $\mathbf{E}(h(Y) | X = x) = \sum_{y \in S_Y} h(y) p_{Y|X}(y|x)$ .

Vale inoltre  $p_{(X,Y)}(x,y) = p_{X|Y}(y|x) p_X(x) \forall x \in S_X, \forall y \in S_Y$ .

1. Lancio un dado equilibrato: si descrive il risultato  $k$  con una v. a.  $D \sim \mathcal{U}(\{1, \dots, 6\})$ , con densità  $p_D(d) = \frac{1}{6} \forall d \in \{1, \dots, 6\}$ . Lancio quindi una moneta equa per  $k$  volte; il numero di teste è la v. a.  $T$ . So che  $T$ , una volta lanciato il dado, è binomiale di parametri  $d$  e  $\frac{1}{2}$ : per  $d \in \{1, \dots, 6\}$ ,  $n \in \{1, \dots, d\}$  vale  $p_{T|D}(n|d) = P(T = n | D = d) = \binom{d}{n} \left(\frac{1}{2}\right)^n \left(\frac{1}{2}\right)^{d-n} = \binom{d}{n} \left(\frac{1}{2}\right)^d$ . La densità congiunta di  $D$  e  $T$  è  $p_{(D,T)}(d,n) = p_{T|D}(n|d) p_D(d) = \binom{d}{n} \left(\frac{1}{2}\right)^d \frac{1}{6}$  per quanto detto sulla densità discreta della legge condizionata.  $D$  e  $T$  non sono indipendenti perché il supporto di  $(D,T)$  non è quadrato: ho degli zeri. La legge marginale di  $T$  è  $p_T(n) = \sum_{d=n}^6 \binom{d}{n} \left(\frac{1}{2}\right)^d \frac{1}{6}$ .

Finora si sono considerate  $X, Y$  discrete; tuttavia, quanto detto non cambia se  $Y$  non è discreta, perché  $\mathbf{P}(Y \in B|X = x)$  si definisce come visto,  $\forall x \in S_X$ . In generale è invece necessario che  $\mathbf{P}(X = x) \neq 0$  per poter definire  $\mathbf{P}(Y \in B|X = x)$  come fatto finora: questa ipotesi dev'essere rimossa per trattare il caso di  $X$  assolutamente continua. L'obiettivo è quindi, dato che  $\mathbf{P}(X = x) = 0$ , rendere sensata la scrittura  $\mathbf{P}(Y \in B, X = x) = \mathbf{P}(X = x) \mathbf{P}(Y \in B|X = x)$  definendo arbitrariamente  $\mathbf{P}(Y \in B|X = x)$ .

Traiamo ancora ispirazione dal caso in cui  $X$  è discreta. In questo caso, usando la disintegrazione grazie alla proprietà distributiva dell'intersezione rispetto all'unione  $(X^{-1}(A) \cap Y^{-1}(B)) = [\bigcup_{x \in A \cap S_X} X^{-1}(\{x\})] \cap Y^{-1}(B) = \bigcup_{x \in A \cap S_X} (X^{-1}(\{x\}) \cap Y^{-1}(B))$ , vale  $\mathbf{P}(X \in A, Y \in B) = \sum_{x \in S_X \cap A} \mathbf{P}(X = x, Y \in B) = \sum_{x \in S_X} I_A(x) \mathbf{P}(Y \in B|X = x) \mathbf{P}(X = x)$  si cerca una formula analoga per il caso continuo, in modo da ottenere un'uguaglianza del tipo  $\mathbf{P}(X \in A, Y \in B) = \int_{\mathbb{R}} I_A(x) dP^X(x)$ , con  $dP^X(x)$  funzione che estenda quanto visto. Infatti  $\sum_{x \in S_X} I_A(x) \mathbf{P}(Y \in B|X = x) \mathbf{P}(X = x)$  può essere vista come integrale di Lebesgue rispetto a una legge discreta:  $\int_{\mathbb{R}} I_A(x) \mathbf{P}(Y \in B|X = x) dP^X(x)$ .

**Def 20.1** Dato  $(\Omega, \mathcal{A}, \mathbf{P})$  spazio di probabilità,  $(E, \mathcal{E}), (F, \mathcal{F})$  spazi misurabili,  $X : \Omega \rightarrow E, Y : \Omega \rightarrow F$  variabili aleatorie, si dice legge condizionale di  $Y$  data  $X$  una funzione  $Q : E \times \mathcal{F} \rightarrow [0, 1]$  tale che

**C1**  $\forall B \in \mathcal{F} Q : E \rightarrow [0, 1], Q = Q(\cdot, B) = Q_B(\cdot)$  è  $\mathcal{E}$ -misurabile

**C2**  $\forall x \in E Q : \mathcal{F} \rightarrow [0, 1], Q = Q(x, \cdot) = Q_x(\cdot)$  è una misura di probabilità su  $(F, \mathcal{F})$

**C3**  $\forall A \in \mathcal{E}, \forall B \in \mathcal{F}$  vale  $\mathbf{P}(X \in A, Y \in B) = \int_E I_A(x) Q(x, B) dP^X(x)$

C1 serve perché per integrare rispetto a  $P^X$  serve che la funzione integranda sia misurabile rispetto a  $\mathcal{E}$ . Si noti che la definizione è estremamente generale:  $X$  e  $Y$  sono funzioni misurabili da  $\Omega$  a spazi misurabili qualsiasi (e. g., se  $X$  e  $Y$  sono vettori aleatori).  $Q$  è definita su  $E \times \mathcal{F}$  perché serve a calcolare una probabilità relativa a  $Y$  (quindi di  $B \in \mathcal{F}$ ) avendo un valore di  $X$  (quindi  $x \in E$ ). C1 significa che  $\forall C \in \mathcal{B}([0, 1]) Q_B^{-1}(C) \in \mathcal{E}$ : serve e. g. a poter calcolare la probabilità che  $X$  appartenga all'insieme degli  $x$  tali che  $Q_B(x) \in C$ .

C3 è un'estensione della formula delle probabilità totali: è noto che, se  $E_1, \dots, E_n$  sono una partizione di  $A$ ,  $\mathbf{P}(A \cap B) = \sum \mathbf{P}(B|E_n) \mathbf{P}(E_n)$ , mentre ora  $P^{(X,Y)}(A \times B) = \int_E I_A(x) Q(x, B) dP^X(x)$ . In C3, le  $x$  "partitionano"  $A$ . L'integrale è ben definito perché, fissato  $B$ , per C1  $Q(x, B)$  è misurabile rispetto a  $\mathcal{E}$ , limitata e nonnegativa, quindi integrabile secondo la legge di  $X$ . Vale inoltre l'uguaglianza

$$P^{(X,Y)}(A \times B) = \int_E I_A(x) Q(x, B) dP^X(x) = \int_E I_A(x) \left( \int_F I_B(y) Q(x, dy) \right) dP^X(x)$$

per C2, dato che in generale  $P(A) = \int_{\Omega} I_A(\omega) dP(\omega)$ .

Ho quindi trovato l'estensione desiderata:  $Q(x, B)$  fa le veci di  $\mathbf{P}(Y \in B|X = x)$  vista nel caso discreto, per cui  $Q(x, dy)$  si merita di essere indicata anche con  $P^{Y|X}(dy|x)$ .



Se  $Q$  è una legge condizionale, grazie a C3 (che definisce  $P^{(X,Y)}(C)$  solo per  $C \in \mathcal{E} \times \mathcal{F}$ ) si ha che  $\forall C \in \mathcal{E} \otimes \mathcal{F}$   $P^{(X,Y)}(C) = \int_E \left( \int_F I_C(x,y) Q(x,dy) \right) dP^X(x)$ , perché  $\mathcal{E} \times \mathcal{F}$  è un  $\pi$ -sistema: non vale più  $I_C(x,y) = I_A(x) I_B(y)$  perché  $C$  non è necessariamente un rettangolo!

In generale  $(P^{(X,Y)}(C) = \mathbf{E}^{(X,Y)}(I_C(X,Y)))$ ,  $\forall h : \mathbb{R}^2 \rightarrow \mathbb{R} : |h| \leq M$  o  $h \geq 0$  vale  $\mathbf{E}(h(X,Y)) = \int_{E \times F} h(x,y) dP^{(X,Y)}(x,y) = \int_E \left( \int_F h(x,y) Q(x,dy) \right) dP^X(x)$ : questo è quanto afferma il teorema di Fubini-Tonelli generalizzato, che non richiede più di lavorare sulla legge prodotto e può essere quindi applicato anche se  $X$  e  $Y$  non sono indipendenti.

**Prop 20.2 (unicità della legge condizionale se  $F = \mathbb{R}^n$ )**

Hp:  $(\Omega, \mathcal{A}, \mathbf{P})$  spazio di probabilità,  $(E, \mathcal{E}), (\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$

spazi misurabili,  $X : \Omega \rightarrow E, Y : \Omega \rightarrow \mathbb{R}^n$  v. a.

Ts:  $\exists !$  legge condizionale di  $Y$  data  $X$

L'unicità significa che  $\forall B \in \mathcal{B}(\mathbb{R}^n)$ , se  $Q$  e  $\tilde{Q}$  sono due leggi condizionali, allora  $Q(x, B) = \tilde{Q}(x, B)$  quasi certamente secondo  $P^X$ . Questo nel seguito giustificherà la possibilità di definire arbitrariamente la legge condizionale su insiemi che hanno misura nulla secondo  $P^X$ .

Si affrontano ora alcuni casi notevoli di legge condizionata.

**1 Componenti indipendenti** Partiamo dal caso estremo di v. a. indipendenti. Se  $X \perp Y$ ,  $P^{(X,Y)} = P^X \otimes P^Y$ .

L'intuizione suggerisce in tal caso  $\mathbf{P}(Y \in B | X = x) = \mathbf{P}(Y \in B)$ :  $Q$ , se la sua definizione è stata ben scelta, deve rispettare questa intuizione di  $P^{Y|X}(dy|x) = P^Y(dy)$ . Definisco allora  $Q(x, B) = P^Y(B)$  e mostro che è effettivamente una legge condizionale.

$\forall B \in \mathcal{F}$   $Q_B(x) = P^Y(B)$  è costante rispetto a  $x$ , quindi la controimmagine di qualsiasi boreliano in  $[0, 1]$  è vuota e  $Q_B(x)$  è misurabile rispetto a  $\mathcal{E}$ .  $Q(x, B)$  è una misura di probabilità perché coincide con  $P^Y(B)$ , che lo è in quanto legge di  $Y$ .

Per definizione di legge prodotto e per il teorema di Fubini-Tonelli  $P^{(X,Y)}(A \times B) = \int_{E \times F} I_{A \times B}(x,y) dP^{(X,Y)}(x,y) = \int_{E \times F} I_A(x) I_B(y) dP^{(X,Y)}(x,y) = \int_{E \times F} I_A(x) I_B(y) dP^X(x) \otimes P^Y(y) = \int_E I_A(x) \left( \int_F I_B(y) dP^Y(y) \right) dP^X(x) = \int_E I_A(x) P^Y(B) dP^X(x)$ , dato che  $P^Y(B) = E^Y(I_B(y))$ . Vale quindi anche C3, e  $P^Y(B)$  è l'unica legge condizionale. E' vero quindi che  $X \perp Y \iff Q(x, B) = P^Y(B)$ .

**2 Una componente funzione dell'altra** Il caso estremo opposto al precedente è il seguente. Sia  $Y = h(X)$  con

$$h : E \rightarrow F \text{ misurabile e non costante. Intuitivamente, } P^{Y|X}(B|x) = \begin{cases} 1 & \text{se } h(x) \in B \\ 0 & \text{se } h(x) \notin B \end{cases} = \delta_{h(x)}(B). \text{ Definisco allora } Q(x, B) = \delta_{h(x)}(B). Q_B(x) \text{ è } \mathcal{E}\text{-misurabile perché } \forall C \in \mathcal{F} Q_B^{-1}(C) = \begin{cases} h^{-1}(B) & \text{se } 1 \in C, 0 \notin C \\ (h^{-1}(B))^c & \text{se } 0 \in C, 1 \notin C \\ E & \text{se } \{0, 1\} \subseteq C \\ \emptyset & \text{se } \{0, 1\} \subseteq C^c \end{cases},$$

che è in ogni caso un elemento di  $\mathcal{E}$  perché  $h$  è misurabile.

$$\delta_{h(x)}(F) = 1 \text{ perché sicuramente } h(x) \in F; \text{ se } (F_i)_{i \geq 1} \text{ è una famiglia di elementi disgiunti di } \mathcal{F}, \delta_{h(x)}\left(\bigcup_{i=1}^{+\infty} F_i\right) = \begin{cases} 1 & \text{se } \exists i : h(x) \in F_i \\ 0 & \text{se } \nexists i : h(x) \in F_i \end{cases} = \sum_{i=1}^{+\infty} \delta_{h(x)}(F_i), \text{ quindi } \delta_{h(x)}(B) \text{ è una misura di probabilità su } (F, \mathcal{F}).$$

$$\int_E I_A(x) Q(x, B) dP^X(x) = \int_E I_A(x) \delta_{h(x)}(B) dP^X(x) = \int_E I_A(x) I_{h^{-1}(B)}(x) dP^X(x) = \int_{A \cap h^{-1}(B)} dP^X(x) = \mathbf{P}(X \in (A \cap h^{-1}(B))), \text{ che coincide con } \mathbf{P}(X \in A, h(X) \in B) = \mathbf{P}(X \in A, Y \in B).$$

**3 Vettore discreto** Considero  $\mathbf{X} : \Omega \rightarrow \mathbb{R}^k, \mathbf{Y} : \Omega \rightarrow \mathbb{R}^n$  vettori aleatori discreti e il vettore  $\begin{pmatrix} \mathbf{X} \\ \mathbf{Y} \end{pmatrix} : \Omega \rightarrow \mathbb{R}^{n+k},$

$$\text{con } \mathcal{E} = \mathcal{B}(\mathbb{R}^k), \mathcal{F} = \mathcal{B}(\mathbb{R}^n). \text{ Definisco } Q(\mathbf{x}, B) = \begin{cases} \mathbf{P}(\mathbf{Y} \in B | \mathbf{X} = \mathbf{x}) & \text{se } \mathbf{x} \in S_{\mathbf{X}} \\ \tilde{Q}(B) & \text{se } \mathbf{x} \notin S_{\mathbf{X}} \end{cases} \text{ dove } \tilde{Q} : \mathcal{F} \rightarrow [0, 1] \text{ è}$$

una misura di probabilità qualsiasi e  $S_{\mathbf{X}} = \{\mathbf{x} \in \mathbb{R}^k : p_{\mathbf{X}}(\mathbf{x}) > 0\}$ , e mostro che è effettivamente una legge condizionale.

$Q$  è una funzione misurabile nel suo primo argomento perché  $Q(\mathbf{x}, B) = \mathbf{P}(\mathbf{Y} \in B | \mathbf{X} = \mathbf{x}) I_{S_{\mathbf{X}}}(\mathbf{x}) + \tilde{Q}(B) I_{S_{\mathbf{X}}^c}(\mathbf{x})$  ha  $\mathbf{P}(\mathbf{Y} \in B | \mathbf{X} = \mathbf{x})$  misurabile dato che ogni insieme discreto è in  $\mathcal{B}(\mathbb{R}^k)$ , in quanto unione di singoletti.

$Q$  è una misura di probabilità nel suo secondo argomento perché lo è la probabilità condizionata e perché  $\tilde{Q}$  lo è per ipotesi. Inoltre, poiché  $\mathbf{X}$  è discreto, per la regola del valore atteso

$$\int_{\mathbb{R}^k} I_A(\mathbf{x}) Q(\mathbf{x}, B) dP^{\mathbf{X}}(\mathbf{x}) = \sum_{\mathbf{x} \in S_{\mathbf{X}}} I_A(\mathbf{x}) \mathbf{P}(\mathbf{Y} \in B | \mathbf{X} = \mathbf{x}) P^{\mathbf{X}}(\mathbf{x}) = \sum_{\mathbf{x} \in S_{\mathbf{X}} \cap A} \mathbf{P}(\mathbf{Y} \in B | \mathbf{X} = \mathbf{x}) \mathbf{P}(\mathbf{X} = \mathbf{x}),$$

che è, per la formula delle probabilità totali,  $\mathbf{P}(\mathbf{X} \in A, \mathbf{Y} \in B)$ . Si noti che la scelta di  $\tilde{Q}$  è irrilevante.

$$\mathbf{P}(\mathbf{Y} \in B | \mathbf{X} = \mathbf{x}) \text{ è una probabilità discreta perché } \mathbf{P}(\mathbf{Y} \in S_{\mathbf{Y}} | \mathbf{X} = \mathbf{x}) = \frac{\mathbf{P}(\mathbf{Y} \in S_{\mathbf{Y}}, \mathbf{X} = \mathbf{x})}{\mathbf{P}(\mathbf{X} = \mathbf{x})} = \frac{\mathbf{P}(\mathbf{X} = \mathbf{x})}{\mathbf{P}(\mathbf{X} = \mathbf{x})} = 1, \text{ con}$$

$$\text{densità discreta di } Y \text{ } p_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x}) = \mathbf{P}(\mathbf{Y} = \mathbf{y} | \mathbf{X} = \mathbf{x}) = \frac{p_{(\mathbf{X}, \mathbf{Y})}(\mathbf{x}, \mathbf{y})}{p_{\mathbf{X}}(\mathbf{x})}.$$

Nel caso discreto il valore atteso di  $Y$  rispetto alla nuova misura di probabilità  $P^{Y|X}(B|x)$ , che si indica con  $\mathbf{E}(Y|X=x)$ , è  $\int_{\Omega} Y(\omega) P(d\omega|X=x) = \int_{\mathbb{R}} y P^{Y|X}(dy|x) = \sum_{y \in S_Y} y p_{Y|X}(y|x)$ . In generale, data  $h : \mathbb{R} \rightarrow \mathbb{R}$  boreliana e limitata oppure nonnegativa,  $\mathbf{E}(h(Y)|X=x) = \sum_{y \in S_Y} h(y) p_{Y|X}(y|x)$ .

**4 Vettore assolutamente continuo** Considero  $\mathbf{X} : \Omega \rightarrow \mathbb{R}^k, \mathbf{Y} : \Omega \rightarrow \mathbb{R}^n$  v. a. tali che il vettore  $\begin{pmatrix} \mathbf{X} \\ \mathbf{Y} \end{pmatrix} :$

$\Omega \rightarrow \mathbb{R}^{n+k}$  è assolutamente continuo, con densità congiunta  $f_{(\mathbf{X}, \mathbf{Y})}$  e densità marginali  $f_{\mathbf{X}}, f_{\mathbf{Y}}$  (rappresentanti scelte arbitrariamente), con  $\mathcal{E} = \mathcal{B}(\mathbb{R}^k), \mathcal{F} = \mathcal{B}(\mathbb{R}^n)$ . Posto  $S_{\mathbf{X}} = \{\mathbf{x} \in \mathbb{R}^k : f_{\mathbf{X}}(\mathbf{x}) > 0\}$ , definisco la densità condizionale di  $\mathbf{Y}$  dato che  $\mathbf{X} = \mathbf{x}$   $f_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x}) = \begin{cases} \frac{f_{(\mathbf{X}, \mathbf{Y})}(\mathbf{x}, \mathbf{y})}{f_{\mathbf{X}}(\mathbf{x})} & \text{se } \mathbf{x} \in S_{\mathbf{X}} \\ \tilde{f}(\mathbf{y}) & \text{se } \mathbf{x} \notin S_{\mathbf{X}} \end{cases}$  dove  $\tilde{f} : \mathbb{R}^n \rightarrow \mathbb{R}$  è una qualsiasi densità continua (nonnegativa, borel, ecc.). Definisco allora  $Q(\mathbf{x}, B) = \int_B f_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x}) d\mathbf{y}$  e mostro che è effettivamente una legge condizionale.  $Q$  è una funzione misurabile nel suo primo argomento perché  $Q(\mathbf{x}, B) = \int_B \left( \frac{f_{(\mathbf{X}, \mathbf{Y})}(\mathbf{x}, \mathbf{y})}{f_{\mathbf{X}}(\mathbf{x})} I_{S_{\mathbf{X}}}(\mathbf{x}) + \tilde{f}(\mathbf{y}) I_{S_{\mathbf{X}}^c}(\mathbf{x}) \right) d\mathbf{y}$ : le  $f$  sono tutte misurabili in quanto densità continue, le funzioni indicatrici lo sono perché  $S_{\mathbf{X}}$  è un boreliano ( $f$  deve avere un insieme di punti di discontinuità di misura nulla); poiché somme, prodotti, integrali (si sta integrando rispetto a  $m \otimes m = m_2$ , quindi si applica Fubini-Tonelli) di funzioni misurabili sono misurabili,  $Q$  è misurabile. Inoltre, fissato  $\mathbf{x} \in \mathbb{R}^k$ ,  $Q(\mathbf{x}, B)$  è una misura di probabilità perché la funzione di  $\mathbf{y}$   $\frac{f_{(\mathbf{X}, \mathbf{Y})}(\mathbf{x}, \mathbf{y})}{f_{\mathbf{X}}(\mathbf{x})}$  è una densità continua su  $\mathbb{R}^n$  ( $\int_{\mathbb{R}^k} \left( \frac{f_{(\mathbf{X}, \mathbf{Y})}(\mathbf{x}, \mathbf{y})}{f_{\mathbf{X}}(\mathbf{x})} I_{S_{\mathbf{X}}}(\mathbf{x}) + \tilde{f}(\mathbf{y}) I_{S_{\mathbf{X}}^c}(\mathbf{x}) \right) d\mathbf{y} = I_{S_{\mathbf{X}}}(\mathbf{x}) + I_{S_{\mathbf{X}}^c}(\mathbf{x}) = 1$ ), e poi l'integrale è sempre una misura di probabilità. Vale poi  $\int_{\mathbb{R}^n} I_A(\mathbf{x}) Q(\mathbf{x}, B) P^{\mathbf{X}}(d\mathbf{x}) = \int_{\mathbb{R}^n} I_A(\mathbf{x}) Q(\mathbf{x}, B) f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x}$ , cioè, essendo nulla la densità in  $S_{\mathbf{X}}^c$ ,  $\int_{S_{\mathbf{X}}} I_A(\mathbf{x}) \left( \int_B \frac{f_{(\mathbf{X}, \mathbf{Y})}(\mathbf{x}, \mathbf{y})}{f_{\mathbf{X}}(\mathbf{x})} d\mathbf{y} \right) f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x} = \int_{S_{\mathbf{X}}} I_A(\mathbf{x}) \left( \int_B f_{(\mathbf{X}, \mathbf{Y})}(\mathbf{x}, \mathbf{y}) d\mathbf{y} \right) d\mathbf{x} = \int_{S_{\mathbf{X}} \times \mathbb{R}^n} I_A(\mathbf{x}) I_B(\mathbf{y}) f_{(\mathbf{X}, \mathbf{Y})}(\mathbf{x}, \mathbf{y}) d\mathbf{y} d\mathbf{x} = \int_{\mathbb{R}^k \times \mathbb{R}^n} I_A(\mathbf{x}) I_B(\mathbf{y}) f_{(\mathbf{X}, \mathbf{Y})}(\mathbf{x}, \mathbf{y}) d\mathbf{y} d\mathbf{x} = \mathbf{P}(\mathbf{X} \in A, \mathbf{Y} \in B)$ .

Così ho scoperto che se un vettore è a. c., allora la sua probabilità condizionale è assolutamente continua e le componenti sono continue. La seguente proposizione mostra anche che vale il viceversa.

Come nel caso discreto,  $\forall \mathbf{x} \in S_{\mathbf{X}} \mathbf{E}(\mathbf{Y}|\mathbf{X} = \mathbf{x}) = \int_{\mathbb{R}^n} \mathbf{y} Q(\mathbf{x}, d\mathbf{y}) = \int_{\mathbb{R}^n} \mathbf{y} f_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x}) d\mathbf{y}$ , mentre  $\mathbf{E}(h(\mathbf{Y})|\mathbf{X} = \mathbf{x}) = \int_{\mathbb{R}^n} h(\mathbf{y}) f_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x}) d\mathbf{y}$ .

**Prop 20.3 (densità congiunta da densità condizionale)**

Hp:  $\begin{pmatrix} \mathbf{X} \\ \mathbf{Y} \end{pmatrix} : \Omega \rightarrow \mathbb{R}^k \times \mathbb{R}^n$  ve. a.,  $\mathbf{X} : \Omega \rightarrow \mathbb{R}^k$  v. a. assolutamente

continuo con densità  $f_{\mathbf{X}}, P^{\mathbf{Y}|\mathbf{X}}(d\mathbf{y}|\mathbf{x})$  ha densità  $f_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x}) \forall \mathbf{x} \in S_{\mathbf{X}}$

Ts:  $\begin{pmatrix} \mathbf{X} \\ \mathbf{Y} \end{pmatrix}$  è assolutamente continuo con densità

$$f_{(\mathbf{X}, \mathbf{Y})}(\mathbf{x}, \mathbf{y}) = \begin{cases} f_{\mathbf{X}}(\mathbf{x}) f_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x}) & \forall (\mathbf{x}, \mathbf{y}) \in S_{\mathbf{X}} \times \mathbb{R}^n \\ 0 & \text{altrove} \end{cases}$$

Nel caso particolare in cui  $\mathbf{X} \perp \mathbf{Y}$  si ottiene, come già visto, che  $f_{(\mathbf{X}, \mathbf{Y})}(\mathbf{x}, \mathbf{y}) = f_{\mathbf{X}}(\mathbf{x}) f_{\mathbf{Y}}(\mathbf{y})$ . Questo permette di assegnare una densità congiunta avendo stabilito quella condizionale!

**5 Vettore misto** Sia  $\mathbf{X} : \Omega \rightarrow \mathbb{R}^k$  una v. a. discreto con supporto  $S_{\mathbf{X}}$  e densità discreta  $p_{\mathbf{X}}$  e  $\mathbf{Y} : \Omega \rightarrow \mathbb{R}^n$

un vettore aleatorio qualsiasi. Suppongo inoltre che  $\mathbf{P}(\mathbf{Y} \in B | \mathbf{X} = \mathbf{x})$ , fissato  $\mathbf{x}$ , sia una probabilità assolutamente continua, cioè  $\exists f_{(\mathbf{Y}|\mathbf{X})}^*(\mathbf{y}|\mathbf{x}) : \mathbf{P}(\mathbf{Y} \in B | \mathbf{X} = \mathbf{x}) = \int_B f_{(\mathbf{Y}|\mathbf{X})}^*(\mathbf{y}|\mathbf{x}) d\mathbf{y} \forall \mathbf{x} \in S_{\mathbf{X}}$ . Pongo allora

$$f_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x}) = \begin{cases} f_{\mathbf{Y}|\mathbf{X}}^*(\mathbf{y}|\mathbf{x}) & \text{se } \mathbf{x} \in S_{\mathbf{X}} \\ \tilde{f}(\mathbf{y}) & \text{se } \mathbf{x} \notin S_{\mathbf{X}} \end{cases}, \text{ dove } \tilde{f} : \mathbb{R}^n \rightarrow \mathbb{R} \text{ è una qualsiasi densità continua, e definisco}$$

$Q(\mathbf{x}, B) = \int_B f_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x}) d\mathbf{y}$ , che si può dimostrare essere effettivamente una legge condizionale (come abbiamo fatto prima). Le proprietà 1 e 2 si dimostrano in modo identico al caso assolutamente continuo. Devo poi

mostrare che  $\forall A \in \mathcal{B}(\mathbb{R}^k), \forall B \in \mathcal{B}(\mathbb{R}^n)$  vale  $\mathbf{P}(X \in A, Y \in B) = \int_{\mathbb{R}^k} I_A(x) Q(x, B) dP^X(x)$ . Essendo  $X$

discreta, per la formula di disintegrazione il lato sinistro è  $\mathbf{P}(X \in A, Y \in B) = \sum_{k \in A \cap S_{\mathbf{X}}} \mathbf{P}(X = k, Y \in B)$ .

$\mathbf{P}(X = k, Y \in B) = \mathbf{P}(Y \in B | X = k) \mathbf{P}(X = k) = \left( \int_B f_{(\mathbf{Y}|\mathbf{X})}^*(\mathbf{y}|k) d\mathbf{y} \right) P^X(k)$ , quindi il lato sinistro è

$\sum_{k \in A \cap S_{\mathbf{X}}} P^X(k) \int_B f_{(\mathbf{Y}|\mathbf{X})}^*(\mathbf{y}|k) d\mathbf{y}$ ; poiché la somma è fatta per  $k \in A \cap S_{\mathbf{X}}$ , vale  $\sum_{k \in A \cap S_{\mathbf{X}}} P^X(k) \int_B f_{(\mathbf{Y}|\mathbf{X})}^*(\mathbf{y}|k) d\mathbf{y} =$

$\sum_{k \in A \cap S_{\mathbf{X}}} P^X(k) \int_B f_{(\mathbf{Y}|\mathbf{X})}(\mathbf{y}|k) d\mathbf{y}$ . Il lato destro è  $\int_{\mathbb{R}^k} I_A(x) Q(x, B) dP^X(x) = \int_A Q(x, B) dP^X(x)$ , che

per la regola del valore atteso per le v. a. discrete è uguale a  $\sum_{x \in A \cap S_{\mathbf{X}}} Q(x, B) P^X(x)$ . Quindi effettivamente

lato sinistro e destro coincidono e  $\mathbf{P}(X \in A, Y \in B) = \int_{\mathbb{R}^k} I_A(x) Q(x, B) dP^X(x)$ .

Se in particolare  $A = \mathbb{R}^k$ , si ottiene  $\mathbf{P}(\mathbf{X} \in A, \mathbf{Y} \in B) = \mathbf{P}(Y \in B) = \sum_{\mathbf{k} \in S_{\mathbf{X}}} P^{\mathbf{X}}(k) \int_B f_{(\mathbf{Y}|\mathbf{X})}(\mathbf{y}|\mathbf{k}) d\mathbf{y} = \int_B \sum_{\mathbf{k} \in S_{\mathbf{X}}} P^{\mathbf{X}}(k) f_{(\mathbf{Y}|\mathbf{X})}(\mathbf{y}|\mathbf{k}) d\mathbf{y}$

dunque  $\mathbf{Y}$  è assolutamente continuo con densità  $f_{\mathbf{Y}}(\mathbf{y}) = \sum_{\mathbf{x} \in S_{\mathbf{X}}} p_{\mathbf{X}}(\mathbf{x}) f_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x}) \forall \mathbf{y} \in \mathbb{R}^n$  (si sta usando una formula di disintegrazione).

**6 Vettore gaussiano** Sia  $\begin{pmatrix} X \\ Y \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mu_X \\ \mu_Y \end{pmatrix}, \begin{bmatrix} \sigma_X^2 & \rho\sigma_X\sigma_Y \\ \rho\sigma_X\sigma_Y & \sigma_Y^2 \end{bmatrix}\right)$  un vettore gaussiano bidimensionale,

dove  $\sigma_X = \sqrt{\sigma_X^2}$  (deviazione standard). Cerco  $P^{Y|X}(dy|x)$ .

Se  $\sigma_X^2 = 0$ ,  $X \sim \delta_{\mu_X}$  e  $X \perp Y$  per ogni  $Y$  (una costante è indipendente da qualsiasi v. a.), quindi si ricade nel

caso 1 e  $P^{Y|X}(dy|x) \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$ . Il caso di  $\sigma_X^2 \neq 0$  è affrontato nella seguente proposizione.

**Prop 21.1 (legge condizionale di un vettore gaussiano)**

$$\text{Hp: } \begin{pmatrix} X \\ Y \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mu_X \\ \mu_Y \end{pmatrix}, \begin{bmatrix} \sigma_X^2 & \rho\sigma_X\sigma_Y \\ \rho\sigma_X\sigma_Y & \sigma_Y^2 \end{bmatrix}\right), \sigma_X^2 > 0$$

$$\text{Ts: } P^{Y|X}(dy|x) \sim \mathcal{N}(m(x), q^2), \text{ con } m(x) = \mu_Y + \rho \frac{\sigma_Y}{\sigma_X} (x - \mu_X) \text{ e } q^2 = \sigma_Y^2 (1 - \rho^2)$$

Si ha quindi  $m(x) = \mathbf{E}(Y|X=x) = \mu_Y + \rho \frac{\sigma_Y}{\sigma_X} (x - \mu_X)$ , mentre  $q^2 = \text{var}(Y|X=x) = \int_{\mathbb{R}} (y - m(x))^2 P^{Y|X}(dy|x) = \sigma_Y^2 (1 - \rho^2)$  è la varianza di  $Y$  condizionata. Se  $\sigma_Y^2 = 0$ ,  $X \perp Y$  e si ricade nel primo caso visto:  $P^{Y|X}(dy|x) \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$ . Se  $\rho^2 = 1$  c'è una relazione lineare affine e quindi deterministica tra  $X$  e  $Y$ : si ha ancora  $q^2 = 0$ , per cui  $Y|X=x \sim \delta_{m(x)}$ , cioè  $Y$  condizionata giace sempre sulla retta di regressione, quindi l'interpolazione è perfetta.  $q^2$ , pur funzione di  $x$  in teoria, è costante.

**Dim\*** Se  $\sigma_Y^2 = 0$ ,  $Y = \mu_Y$  q. c. e quindi  $X \perp Y$ : si ricade nel primo caso visto e  $P^{Y|X}(dy|x) \sim P^Y(dy|x) = P^Y(y) = \mathcal{N}(\mu_Y, 0)$ .

Se  $\rho^2 = 1$ , per 15.6  $Y = aX + b$  con  $a = \rho \frac{\sigma_Y}{\sigma_X}$ ,  $b = \mu_Y - \mu_X \rho \frac{\sigma_X}{\sigma_Y}$ , quindi si ricade nel secondo caso e  $P^{Y|X}(dy|x) \sim \delta_{ax+b}$ , cioè  $\mathcal{N}(m(x), 0)$ , con  $m(x) = ax + b = \rho \frac{\sigma_Y}{\sigma_X} x + \mu_Y - \mu_X \rho \frac{\sigma_X}{\sigma_Y}$ .

Se invece  $\det C = (1 - \rho^2) \sigma_X^2 \sigma_Y^2 > 0$ , il vettore è assolutamente continuo e si ottiene  $P^{Y|X}$  dal caso 4 affrontato sopra. E' noto che  $f_{(X,Y)}(x,y) = \frac{1}{2\pi\sqrt{(1-\rho^2)\sigma_X^2\sigma_Y^2}} e^{-\frac{1}{2(1-\rho^2)}\left(\frac{(x-\mu_X)^2}{\sigma_X^2} - 2\rho\frac{(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y} + \frac{(y-\mu_Y)^2}{\sigma_Y^2}\right)}$ , per cui si è nel caso assolutamente continuo e la densità condizionale è  $\frac{f_{(X,Y)}(x,y)}{f_X(x)} = \frac{1}{\sqrt{2\pi(1-\rho^2)\sigma_Y^2}} e^{-\frac{1}{2(1-\rho^2)}\left(\frac{(x-\mu_X)^2}{\sigma_X^2} - 2\rho\frac{(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y} + \frac{(y-\mu_Y)^2}{\sigma_Y^2}\right) - \frac{1}{2\sigma_X^2}(x-\mu_X)^2}$ . L'esponente è  $-\frac{1}{2(1-\rho^2)}\left(\frac{(x-\mu_X)^2}{\sigma_X^2} - 2\rho\frac{(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y} + \frac{(y-\mu_Y)^2}{\sigma_Y^2}\right) - \frac{1}{2\sigma_X^2}(x-\mu_X)^2 = -\frac{1}{2\sigma_Y^2(1-\rho^2)}\left(x - \left(\mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(x-\mu_X)\right)\right)^2$ , per cui la densità condizionale coincide con  $\frac{1}{\sqrt{2\pi(1-\rho^2)\sigma_Y^2}} e^{-\frac{1}{2\sigma_Y^2(1-\rho^2)}\left(x - \left(\mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(x-\mu_X)\right)\right)^2}$ . ■

Si affronta lo stesso problema per due vettori aleatori generici.

### Prop 21.1 (legge condizionale di un vettore gaussiano)

$$\text{Hp:} \quad \begin{pmatrix} \mathbf{X} \\ \mathbf{Y} \end{pmatrix} : \Omega \rightarrow \mathbb{R}^{k+n}, \quad \begin{pmatrix} \mathbf{X} \\ \mathbf{Y} \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mu_{\mathbf{X}} \\ \mu_{\mathbf{Y}} \end{pmatrix}, \begin{bmatrix} C_{\mathbf{X}} & C_{\mathbf{X}\mathbf{Y}} \\ C_{\mathbf{Y}\mathbf{X}} & C_{\mathbf{Y}} \end{bmatrix}\right), \quad \det C_{\mathbf{X}} > 0$$

$$\text{Ts:} \quad P^{\mathbf{Y}|\mathbf{X}}(d\mathbf{y}|\mathbf{x}) \sim \mathcal{N}(\mathbf{m}(\mathbf{x}), Q), \quad \text{con } \mathbf{m}(\mathbf{x}) = \mu_{\mathbf{Y}} + C_{\mathbf{Y}\mathbf{X}}C_{\mathbf{X}}^{-1}(\mathbf{x} - \mu_{\mathbf{X}}) \quad \text{e } Q = C_{\mathbf{Y}} - C_{\mathbf{Y}\mathbf{X}}C_{\mathbf{X}}^{-1}C_{\mathbf{X}\mathbf{Y}}$$

Nelle ipotesi si ha  $\begin{bmatrix} C_{\mathbf{X}} & C_{\mathbf{X}\mathbf{Y}} \\ C_{\mathbf{Y}\mathbf{X}} & C_{\mathbf{Y}} \end{bmatrix} \in M_{\mathbb{R}}(n+k, n+k)$  e  $C_{\mathbf{Y}\mathbf{X}} = C_{\mathbf{X}\mathbf{Y}}$ , dato che la matrice di covarianza dev'essere simmetrica.

## 13.1 Attese condizionate

Voglio studiare  $\mathbf{E}(Y|X=x)$  al variare di  $x$ , ragionando a priori e non più in medias res.

Considero  $(\Omega, \mathcal{A}, \mathbf{P})$  spazio di probabilità,  $(E, \mathcal{E})$  spazio misurabile,  $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ ,  $X : \Omega \rightarrow E$  v. a.,  $Y : \Omega \rightarrow \mathbb{R}^n$  v. a. limitata, quindi tale che  $\exists M > 0 : \|Y(\omega)\| \leq M \forall \omega \in \Omega$ . Si richiede che  $F = \mathbb{R}^n$  perché in tal caso, per 20.2,  $\exists ! Q(x, B)$ . La richiesta di limitatezza è più forte che richiedere che  $Y$  sia finita, cioè che  $\mathbf{P}(Y = +\infty) = 0$ : infatti se  $\|Y(\omega)\| \leq M \forall \omega \in \Omega$ ,  $\mathbf{P}(Y = +\infty) = 0 = \mathbf{P}\left(\bigcap_{n=1}^{+\infty} (Y \geq n)\right) = 0$  perché per  $n = M+1$  si ha un evento di probabilità nulla, ma esistono v. a. non limitate e q. c. finite. La limitatezza implica che  $Y \in L^1(\Omega, \mathcal{A}, \mathbf{P})$ ,

quindi  $\int_{\mathbb{R}} y P^{Y|X}(dy|x)$  esiste finito per la regola del valore atteso; invece con  $Y$  finita potremmo avere valore atteso infinito.

In base a quanto visto finora, se  $n = 1$  e si fissa a  $x$  il valore di  $X$ ,  $\mathbf{E}(Y|X = x) = \int_{\mathbb{R}} y P^{Y|X}(dy|x)$ , che si sceglie di indicare con  $m(x)$ ,  $m : E \rightarrow \mathbb{R}$ , che è una funzione di  $x$  misurabile rispetto a  $\mathcal{E}$  (si dimostra facilmente perché  $P^{Y|X}(dy|x)$  è  $\mathcal{E}$ -misurabile). In realtà, dato che si conosce il valore di  $X(\omega)$ , ma non tutto il risultato dell'esperimento,  $m(x)$  può anche essere vista come  $m(X(\omega))$ , con  $m \circ X : \Omega \rightarrow \mathbb{R}$ , e ha il significato di valore centrale assunto da  $Y$  conoscendo il valore di  $X$ ,  $\mathbf{E}(Y|X = X(\omega))$ .  $m(X(\omega))$  è composizione di funzioni misurabili e quindi  $\mathcal{A}$ -misurabile:  $m(X)$  è una variabile aleatoria detta attesa condizionata, o valore atteso condizionale, di  $Y$  data  $X$ .

1. Considero l'esperimento già visto del dado e della moneta:  $D \sim \mathcal{U}(\{1, \dots, 6\})$ ,  $T$  numero di teste uscite nei lanci della moneta,  $p_{T|D}(\cdot|d) = \text{bin}(d, \frac{1}{2})$ , con  $d \in \{1, \dots, 6\}$ . Calcolo  $\mathbf{E}(T|D)$ .  $\mathbf{E}(T|D = d) = \frac{1}{2}d$ , per cui  $\mathbf{E}(T|D) = \frac{D}{2}$ , in accordo con l'intuizione secondo cui, se si sono fatti  $D$  lanci, il numero di teste è circa la metà.

Se si conosce  $X(\omega)$ , è noto anche se si sono realizzati tutti gli eventi  $A$  che possono essere scritti tramite  $X$ , cioè tali che  $\exists B \in \mathcal{E} : A = (X \in B)$ . L'insieme  $\{(X \in B) : B \in \mathcal{E}\}$  è detto insieme degli eventi generati da  $X$  e si indica con  $\sigma(X)$ ; si è già dimostrato che è una  $\sigma$ -algebra su  $\Omega$ . Essendo  $X$  misurabile rispetto a  $\mathcal{A}$ ,  $\sigma(X) \subseteq \mathcal{A}$ , ed è quindi detta sotto  $\sigma$ -algebra di  $\mathcal{A}$ ; si dimostra che  $\sigma(X)$  è la più piccola  $\sigma$ -algebra su  $\Omega$  rispetto a cui  $X$  è misurabile.  $\sigma(X)$  rappresenta l'informazione "rivelata" da  $X$ , cioè l'insieme degli eventi su cui si sa tutto se si conosce il valore di  $X$ .

$X$  è  $\mathcal{A}$ -misurabile sse  $\sigma(X) \subseteq \mathcal{A}$ .

1. Sia  $(\{0, 1\}^{\mathbb{N}}, \mathcal{A}, \mathbf{P})$  lo spazio di Bernoulli,  $X_k : \Omega \rightarrow \{0, 1\}$ ,  $X_k(\omega) = \omega_k$ . Allora, dato  $k = 3$ ,  $\sigma(X_3) = \{A \subseteq \Omega : A = X^{-1}(B), B \in \{\emptyset, \{0, 1\}, \{0\}, \{1\}\}\}$ . Sapendo che  $X_3 = 1$ , so anche che se si sono verificati gli eventi  $(X_3 > 0)$  (si è verificato),  $(X_3 < \frac{1}{2})$  (non si è verificato), mentre non ho informazioni sull'evento  $(X_1 = 0)$ , che non appartiene a  $\sigma(X_3) = \{\Omega, \emptyset, X_3^{-1}(\{1\}), X_3^{-1}(\{0\})\}$ .

### Lemma 21.2 (lemma di misurabilità di Doob)

Hp:  $(\Omega, \mathcal{A}, \mathbf{P})$  spazio di probabilità,  $X : \Omega \rightarrow E, V : \Omega \rightarrow \mathbb{R}^n$  v. a.

Ts:  $V$  è  $\sigma(X)$ -misurabile  $\iff \exists h : E \rightarrow \mathbb{R}^n$  misurabile :  $V = h(X)$

Il senso del lemma è che, affinché ogni controimmagine di  $B \in \mathcal{B}(\mathbb{R}^n)$  attraverso  $V$  appartenga a  $\sigma(X)$ ,  $B$  deve poter essere scritto come evento generato da  $X$  ( $\sigma(V) \subseteq \sigma(X)$ ), e questo è possibile per ogni  $B$  se e solo se  $V = h(X)$ .

**Dim** Se  $\exists h : E \rightarrow \mathbb{R}^n$  misurabile :  $V = h(X)$ , allora  $\forall B \in \mathcal{B}(\mathbb{R}^n) V^{-1}(B) = X^{-1}(h^{-1}(B)) \in \sigma(X)$  perché ovviamente  $X$  è  $\sigma(X)$ -misurabile.

Sia  $V$   $\sigma(X)$ -misurabile e semplice. Suppongo  $V = a_1 I_{A_1} + \dots + a_N I_{A_N}$ : essendo  $\sigma(X)$ -misurabile,  $A_1, \dots, A_N \in \sigma(X)$  e  $\exists B_1, \dots, B_N : A_1 = X^{-1}(B_1), \dots, A_N = X^{-1}(B_N)$ , per cui  $h(x) = a_1 I_{B_1}(x) + \dots + a_N I_{B_N}(x)$  dà la tesi.

Sia  $V$  nonnegativa: esiste una successione  $V_n$  di v. a. semplici e  $\sigma(X)$ -misurabili che cresce a  $V$ ; per ciascuna di esse  $\exists h_n : V_n = h_n(X)$ . Allora  $\forall \omega \lim_{n \rightarrow +\infty} V_n(\omega) = V(\omega) = \lim_{n \rightarrow +\infty} h_n(X(\omega)) = h(X(\omega))$ : poiché  $h(x) = \lim_{n \rightarrow +\infty} h_n(x)$  è misurabile per i criteri di misurabilità, vale  $V = h(X)$ . ■

Segue da tale lemma che  $\mathbf{E}(Y|X) = m(X) = Z$  è  $\sigma(X)$ -misurabile.

La triplice notazione per il valore atteso condizionale evidenzia rispettivamente il significato modellistico relativo alla probabilità condizionata, l'essere una funzione di  $X$ , l'essere una variabile aleatoria.

### Prop 21.3

Hp:  $(\Omega, \mathcal{A}, \mathbf{P})$  spazio di probabilità,  $(E, \mathcal{E})$  spazio misurabile,

$X : \Omega \rightarrow E$  v. a.,  $Y : \Omega \rightarrow \mathbb{R}$  v. a. limitata,  $Z = m(X) = \mathbf{E}(Y|X)$

Ts:  $Z$  è l'unica (a meno di eventi trascurabili) v. a. reale

$\sigma(X)$ -misurabile tale che  $\forall A \in \sigma(X) \int_A Y d\mathbf{P} = \int_A Z d\mathbf{P}$

Nel caso di  $A = \Omega \int_{\Omega} Y d\mathbf{P} = \mathbf{E}(Y) = \int_{\Omega} \mathbf{E}(Y|X) d\mathbf{P} = \mathbf{E}(\mathbf{E}(Y|X)) = \mathbf{E}(Z)$ : questo è dovuto al fatto che  $Z$  è il valore atteso di  $Y$  condizionato a un certo valore di  $X$ , che però non è noto a priori e quindi varia su tutto il supporto di  $X$  senza fornire informazioni aggiuntive. Quindi  $\mathbf{E}(Y)$  può essere ottenuto con una prima "centratura" condizionata a  $X$ , che non dà nuova informazioni, e una seconda centratura: da questo si evince che  $\mathbf{E}(Y|X)$  dipende solo da  $\sigma(X)$ .  $\int_A Y d\mathbf{P} = \int_A Z d\mathbf{P}$ , se  $Y$  e  $Z$  fossero in  $L^2$ , verrebbe da  $\langle X, Y \rangle = \langle X, \Pi Y \rangle$  con  $X = I_A$ .

La tesi può essere scritta equivalentemente come  $\mathbf{E}(I_A Y) = \mathbf{E}(I_A Z) = \mathbf{E}(I_A \mathbf{E}(Y|X))$ . Il senso della proposizione è che se  $A$  è un evento generato da  $X$ , nel calcolare l'integrale ristretto ad  $A$  si possiede già tutta l'informazione rivelata da  $X$ , e si può quindi calcolare  $\mathbf{E}(I_A Y)$  invece che  $\mathbf{E}(I_A \mathbf{E}(Y|X))$ .

Da  $\mathbf{E}(Y) = \mathbf{E}(\mathbf{E}(Y|X))$  si deduce che, se  $X$  è discreta, per la regola del valore atteso  $\mathbf{E}(Y) = \int_{\Omega} \mathbf{E}(Y|X) d\mathbf{P}(\omega) = \int_E \mathbf{E}(Y|x) dP^X(x) = \sum_{x \in S_X} \mathbf{E}(Y|x) p_X(x)$  (!); se  $X$  è assolutamente continua,  $\mathbf{E}(Y) = \int_{\Omega} \mathbf{E}(Y|X) d\mathbf{P}(\omega) = \int_E \mathbf{E}(Y|x) dP^X(x) = \int_{\mathbb{R}^n} \mathbf{E}(Y|x) f_X(x) dx$ .

Nella tesi non compare direttamente  $X$ , ma solo  $\sigma(X)$ . Questo significa che se due v. a. generano la stessa  $\sigma$ -algebra avranno la stessa attesa condizionata: e. g. se  $h(X)$  è biunivoca e bimisurabile (cioè misurabile e con inversa misurabile), allora  $\sigma(h(X)) = \sigma(X)$  (non cambia l'informazione rivelata) e  $\mathbf{E}(Y|h(X)) = \mathbf{E}(Y|X)$ .

1. Sia  $h(X) = X^3$ . Allora  $\sigma(X^3) = \sigma(X)$ , per cui  $\mathbf{E}(Y|X) = \mathbf{E}(Y|X^3)$ .

Quindi la tesi, che racchiude il significato fondamentale di  $Z$ , può essere usata come definizione alternativa di attesa condizionata - cosa conveniente perché è indipendente dalla legge condizionale - per una generica  $\sigma$ -algebra.

Data  $\mathcal{A}$   $\sigma$ -algebra su  $\Omega$ , si indica con  $\mathcal{G} \subseteq \mathcal{A}$  una generica sotto  $\sigma$ -algebra di  $\mathcal{A}$ . Finora si è considerata  $\mathcal{G} = \sigma(X)$ .

**Def 21.4** Dato  $(\Omega, \mathcal{A}, \mathbf{P})$  spazio di probabilità,  $Y : \Omega \rightarrow \mathbb{R}$  variabile aleatoria tale che  $Y \in L^1(\Omega, \mathcal{A}, \mathbf{P})$ ,  $\mathcal{G} \subseteq \mathcal{A}$  sotto  $\sigma$ -algebra di  $\mathcal{A}$ , si dice attesa condizionale di  $Y$  data  $\mathcal{G}$ , e si indica con  $\mathbf{E}(Y|\mathcal{G})$ , una variabile aleatoria  $Z : \Omega \rightarrow \mathbb{R}$  tale che

**C1**  $Z \in L^1(\Omega, \mathcal{A}, \mathbf{P})$

**C2**  $Z$  è  $\mathcal{G}$ -misurabile

**C3**  $\forall G \in \mathcal{G} \int_G Z d\mathbf{P} = \int_G Y d\mathbf{P}$

Modellisticamente parlando,  $\mathbf{E}(Y|\mathcal{G})$  rappresenta la modifica di  $\mathbf{E}(Y)$  in base alle nuove informazioni ottenute in medias res, che sono il verificarsi o meno degli eventi di  $\mathcal{G}$ . C2 significa che  $\forall B \in \mathcal{B}(\mathbb{R}) Z^{-1}(B) \in \mathcal{G} \subsetneq \mathcal{A}$ . C3 può essere espressa equivalentemente dicendo che  $\mathbf{E}(I_G Y) = \mathbf{E}(I_G Z)$ , con  $I_G = I_G(\omega)$ .  $Z$  esprime come cambia  $Y$  sapendo se si sono verificati o no tutti gli eventi di  $\mathcal{G}$ . Si noti che in generale  $\mathbf{E}(Y|\mathcal{G}) \neq Y$ , se  $Y$  non è  $\mathcal{G}$ -misurabile.

Se  $\mathbf{Y} : \Omega \rightarrow \mathbb{R}^n$ ,  $\mathbf{E}(\mathbf{Y}|\mathcal{G})$  è definita come

$$\begin{pmatrix} \mathbf{E}(Y_1|\mathcal{G}) \\ \dots \\ \mathbf{E}(Y_n|\mathcal{G}) \end{pmatrix}.$$

C3 è equivalente alla condizione C3'  $\forall W$   $\mathcal{G}$ -misurabile e limitata vale  $\int_\Omega ZW d\mathbf{P} = \int_\Omega YW d\mathbf{P}$ , cioè  $\mathbf{E}(WZ) = \mathbf{E}(WY)$ . Infatti, se vale C3', nel caso particolare di  $W = I_G$  con  $G \in \mathcal{G}$  si ha  $\int_G Z d\mathbf{P} = \int_G Y d\mathbf{P}$ . Mostro il contrario: suppongo che valga C3, cioè  $\forall G \in \mathcal{G} \mathbf{E}(I_G Y) = \mathbf{E}(I_G Z)$ . Suppongo  $W$  sia  $\mathcal{G}$ -misurabile e semplice, per cui  $W = w_1 I_{A_1} + \dots + w_N I_{A_N}$ . Allora per linearità  $\mathbf{E}(WY) = \mathbf{E}((w_1 I_{A_1} + \dots + w_N I_{A_N})Y) = w_1 \mathbf{E}(I_{A_1} Y) + \dots + w_N \mathbf{E}(I_{A_N} Y)$  coincide con  $\mathbf{E}(WZ)$ . Sia ora  $W$   $\mathcal{G}$ -misurabile limitata non negativa: esiste una successione  $W_n$  di v. a. semplici che la approssima, per le quali vale  $\mathbf{E}(W_n Z) = \mathbf{E}(W_n Y)$ . Allora  $|W_n Z| \leq LZ$ , che è integrabile, e  $W_n Z$  converge a  $WZ$ , per cui  $\lim_{n \rightarrow +\infty} \mathbf{E}(W_n Z) = \mathbf{E}(WZ)$ ; ma d'altro canto, per motivi analoghi  $\lim_{n \rightarrow +\infty} \mathbf{E}(W_n Y) = \mathbf{E}(WY)$ , per cui  $\mathbf{E}(WZ) = \mathbf{E}(WY)$ .



Si può dimostrare che nelle ipotesi della definizione esiste  $Z$ ; inoltre vale il seguente teorema.

**Teo 21.5 (unicità quasi certa dell'attesa condizionata a una sotto  $\sigma$ -algebra)**

Hp:  $(\Omega, \mathcal{A}, \mathbf{P})$  spazio di probabilità,  $Y : \Omega \rightarrow \mathbb{R}$  v. a.,  $Y \in L^1(\Omega, \mathcal{A}, \mathbf{P})$ ,

$\mathcal{G} \subseteq \mathcal{A}$  sotto  $\sigma$ -algebra di  $\mathcal{A}$

Ts:  $\exists ! Z$  che soddisfa 21.4

Vale quindi l'unicità dell'attesa condizionale a una sotto  $\sigma$ -algebra, a meno di identità quasi certa: se  $Z, \tilde{Z}$  attese condizionali di  $Y$  data  $\mathcal{G}$ ,  $Z = \tilde{Z}$  quasi certamente. In realtà quindi si indica con  $\mathbf{E}(Y|\mathcal{G})$  una classe di equivalenza di v. a.

Se  $\mathcal{G} = \sigma(X)$ ,  $\mathbf{E}(Y|\sigma(X))$  coincide con  $m(X)$  definita sopra, che è infatti integrabile,  $\sigma(X)$ -misurabile (perché uguale a  $m(X)$ , per il lemma di Doob), e ha la proprietà C3 per la 21.3.

Non esiste una ricetta standard per trovare  $Z = \mathbf{E}(Y|\mathcal{G})$ ; tuttavia la 21.5 fa sì che se seguendo l'intuito si trova una  $Z$  che soddisfa le tre proprietà, allora quella è l'attesa condizionale.

1. Sia  $X \in L^1(\Omega, \mathcal{A}, \mathbf{P})$ ,  $\mathcal{G} = \sigma(X)$ ,  $Y = X$ . Per quanto visto sopra  $Z = \mathbf{E}(Y|\sigma(X))$  si indica in tal caso con  $\mathbf{E}(X|X)$ . Intuitivamente, è ragionevole che  $\mathbf{E}(X|X) = X$ . Ovviamente  $X \in L^1(\Omega, \mathcal{A}, \mathbf{P})$ , è  $\sigma(X)$ -misurabile perché  $\sigma(X)$  è la più piccola  $\sigma$ -algebra su  $\mathcal{A}$  che renda  $X$  misurabile, e vale ovviamente  $\forall G \in \sigma(X)$   $\int_G X d\mathbf{P} = \int_G X d\mathbf{P}$ .
2. Sia  $X$  gaussiana di parametri  $\mu, \sigma^2$  con  $\sigma^2 > 0$ .  $\mathbf{E}(X|X^2) = \mathbf{E}(X|\sigma(X^2))$ .  $h(X) = X^2$  non è biunivoca, quindi non soddisfa le ipotesi su  $h$  che servono per affermare  $\mathbf{E}(X|X^2) = \mathbf{E}(X|X)$ : infatti  $X^2$  rivela meno informazione di  $X$ ,  $\sigma(X^2) \subsetneq \sigma(X)$ . Per trovare  $\mathbf{E}(X|X^2)$  si può ragionare intuitivamente sul fatto che data  $X^2$  i valori possibili sono  $-X$  e  $X$ , il cui "centro" è 0 perché la densità normale è simmetrica. Rigorosamente, si vuole imporre  $\mathbf{E}(I_G Z) = \mathbf{E}(I_G X)$ , dove  $G \in \sigma(X^2)$  è sempre della forma  $G = (X^2 \in B)$ . Vale  $\mathbf{E}(I_G X) = \mathbf{E}(I_{(X^2 \in B)} X) = \mathbf{E}(I_B(X^2) X) = \int_{-\infty}^{+\infty} I_B(x^2) x \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2} dx = 0 \forall B$  perché la funzione integranda è integrabile in senso improprio e dispari. Allora la variabile aleatoria  $Z = 0$  soddisfa C1, C2, C3, quindi per unicità  $\mathbf{E}(X|X^2) = 0$ .

**Teo 22.1 (proprietà dell'attesa condizionata)**

Hp:  $(\Omega, \mathcal{A}, \mathbf{P})$  spazio di probabilità,  $Y : \Omega \rightarrow \mathbb{R}$  variabile aleatoria,

$$Y \in L^1(\Omega, \mathcal{A}, \mathbf{P}), \mathcal{G} \subseteq \mathcal{A} \text{ sotto } \sigma\text{-algebra di } \mathcal{A}$$

Ts: (1)  $\mathbf{E}(\cdot|\mathcal{G}) : L^1(\Omega, \mathcal{A}, \mathbf{P}) \rightarrow L^1(\Omega, \mathcal{G}, \mathbf{P})$  è una mappa lineare positiva

$$(2) \mathbf{E}(\mathbf{E}(Y|\mathcal{G})) = \mathbf{E}(Y)$$

$$(3) \text{ se } Y \text{ è } \mathcal{G}\text{-misurabile } \mathbf{E}(Y|\mathcal{G}) = Y \text{ q. c.}$$

$$(4) \text{ se } Y \perp \mathcal{G} \text{ allora } \mathbf{E}(Y|\mathcal{G}) = \mathbf{E}(Y) \text{ q. c.}$$

$$(5) \text{ se } \mathcal{H} \subseteq \mathcal{G} \subseteq \mathcal{A} \text{ è una sotto-sotto } \sigma\text{-algebra, } \mathbf{E}(Y|\mathcal{H}) = \mathbf{E}(\mathbf{E}(Y|\mathcal{G})|\mathcal{H}) \text{ q. c.}$$

$$(6) \text{ se } W \text{ è } \mathcal{G}\text{-misurabile e } WY \in L^1(\mathcal{A}), \text{ allora } \mathbf{E}(WY|\mathcal{G}) = W\mathbf{E}(Y|\mathcal{G}) \text{ q. c.}$$

$$(7) \text{ se } Y_1 = Y_2 \text{ q. c., allora } \mathbf{E}(Y_1|\mathcal{G}) = \mathbf{E}(Y_2|\mathcal{G}) \text{ q. c.}$$

$$(8) \text{ se } Y_n \geq 0 \forall n, Y_n \nearrow Y \text{ q. c., allora } \mathbf{E}(Y_n|\mathcal{G}) \nearrow \mathbf{E}(Y|\mathcal{G}) \text{ q. c.}$$

$$(9) \text{ se } |Y_n| \leq V \in L^1 \forall n, Y_n \rightarrow Y \text{ q. c., allora } \mathbf{E}(Y_n|\mathcal{G}) \rightarrow \mathbf{E}(Y|\mathcal{G}) \text{ q. c.}$$

$$(10) \text{ se } Y \in L^p(\Omega, \mathcal{A}, \mathbf{P}), \text{ allora } \mathbf{E}(Y|\mathcal{G}) \in L^p(\Omega, \mathcal{G}, \mathbf{P})$$

$$(11) \text{ se } h : \mathbb{R} \rightarrow \mathbb{R} \text{ è boreliana, } h(Y) \in L^1, \mathcal{G} = \sigma(X), \text{ allora } \mathbf{E}(h(Y)|\mathcal{G}) = n(X)$$

$$\text{con } n(X) = \int_{\mathbb{R}} h(y) P^{Y|X}(dy|x)$$

(1) significa che  $\forall Y_1, Y_2 \in L^1(\Omega, \mathcal{A}, \mathbf{P}), \forall a, b \in \mathbb{R}$ , vale  $\mathbf{E}(aY_1 + bY_2|\mathcal{G}) = a\mathbf{E}(Y_1|\mathcal{G}) + b\mathbf{E}(Y_2|\mathcal{G})$ , e che  $\forall Y \in L^1(\Omega, \mathcal{A}, \mathbf{P})$ , se  $Y \geq 0$   $\mathbf{E}(Y|\mathcal{G}) \geq 0$  q. c. (ottenere delle informazioni su  $Y$  non cambia il fatto che  $\mathbf{E}(Y) \geq 0$ , poiché  $Y \geq 0$ ). In (2) è fondamentale l'ipotesi  $Y \in L^1$ . (3) è sensata perché sapere se si sono verificati o no tutti gli eventi di  $\mathcal{G}$  permette di sapere se si sono verificati tutti gli eventi generati da  $Y$ , e quindi di conoscere il valore esatto di  $Y$ : tutta l'informazione presente in  $\mathcal{G}$  rivela anche  $Y$ . In questo la proiezione produce la migliore approssimazione possibile. In (4)  $Y \perp \mathcal{G}$  significa che  $\sigma(Y) \perp \mathcal{G}$ , cioè per  $\forall A \in \sigma(Y), B \in \mathcal{G}$  vale  $A \perp B$ . E' una sorta di ortogonalità: ottengo la peggiore approssimazione possibile, se ho informazioni slegate da  $Y$  e che quindi non sono di alcuna utilità. (5) è detta proprietà di inscatolamento. (6) discende da (3):  $\mathbf{E}(W|\mathcal{G}) = W$  q. c. (8) è il teorema di convergenza monotona. (9) è quello di convergenza dominata. In (11)  $h(Y) \in L^1$  serve per poter parlare di  $\mathbf{E}(Y|\mathcal{G})$ . Si capisce che  $n(X)$  è un'estensione di  $m(X)$ , per cui la nuova definizione di attesa condizionata è effettivamente un'estensione; coerenza.

Una conseguenza del teorema è che se considero  $\mathcal{G} = \{\emptyset, \Omega\}$   $\sigma$ -algebra banale, allora  $Y$  è indipendente da  $\mathcal{G}$  e per 22.1 (4) vale  $\mathbf{E}(Y|\mathcal{G}) = \mathbf{E}(Y)$ .

**Dim\*** (1) mostro che  $\forall Y_1, Y_2 \in L^1(\Omega, \mathcal{A}, \mathbf{P})$  e  $\forall a, b \in \mathbb{R}$   $\mathbf{E}(aY_1 + bY_2|\mathcal{G}) = a\mathbf{E}(Y_1|\mathcal{G}) + b\mathbf{E}(Y_2|\mathcal{G})$ .  $a\mathbf{E}(Y_1|\mathcal{G}) + b\mathbf{E}(Y_2|\mathcal{G})$  è in  $L^1(\Omega, \mathcal{A}, \mathbf{P})$  perché lo è ogni addendo; è  $\mathcal{G}$ -misurabile perché lo è ogni addendo. Inoltre  $\forall G \in \mathcal{G}$   $\int_G (aY_1 + bY_2) d\mathbf{P} = \int_G (a\mathbf{E}(Y_1|\mathcal{G}) + b\mathbf{E}(Y_2|\mathcal{G})) d\mathbf{P}$  perché per linearità dell'integrale  $\int_G (aY_1 + bY_2) d\mathbf{P} = a \int_G Y_1 d\mathbf{P} + b \int_G Y_2 d\mathbf{P} = a \int_G \mathbf{E}(Y_1|\mathcal{G}) d\mathbf{P} + b \int_G \mathbf{E}(Y_2|\mathcal{G}) d\mathbf{P}$ .

Se  $Y \in L^1(\Omega, \mathcal{A}, \mathbf{P})$  e  $Y \geq 0$  allora  $\mathbf{E}(Y|\mathcal{G}) \geq 0$  q. c. Infatti per la terza proprietà  $\forall G \in \mathcal{G}$   $\int_G Y d\mathbf{P} = \int_G \mathbf{E}(Y|\mathcal{G}) d\mathbf{P}$ , ma se  $Y \geq 0$ , per positività del valore atteso  $\int_G Y d\mathbf{P} \geq 0$  e di conseguenza  $\int_G \mathbf{E}(Y|\mathcal{G}) d\mathbf{P} \geq 0$  q. c.

(2) E' noto  $\mathbf{E}(I_G Y) = \mathbf{E}(I_G Z) \forall G \in \mathcal{G}$ : se si considera in particolare  $G = \Omega$ , si ha  $\mathbf{E}(Y) = \mathbf{E}(Z)$ .

(3) Se  $Y$  è  $\mathcal{G}$ -misurabile, poiché per ipotesi  $Y \in L^1(\Omega, \mathcal{A}, \mathbf{P})$  e banalmente  $\mathbf{E}(I_G Y) = \mathbf{E}(I_G Y) \forall G \in \mathcal{G}$ ,  $Y$  soddisfa le tre proprietà richieste all'attesa condizionata, per cui, per unicità - a meno di identità quasi certa - dell'attesa condizionata,  $Z = Y$ .

(4) Se  $Z = \mathbf{E}(Y)$  e  $\forall G \in \mathcal{G}, \forall B \in \sigma(Y)$   $\mathbf{P}(BG) = \mathbf{P}(B)\mathbf{P}(G)$ , le v. a.  $Y$  e  $I_G$  sono indipendenti. Vale allora  $\mathbf{E}(I_G Y) = \mathbf{E}(I_G Z) \forall G \in \mathcal{G}$ , perché si ha  $\mathbf{E}(I_G Y) = \mathbf{E}(Y)\mathbf{E}(I_G) = \mathbf{E}(I_G \mathbf{E}(Y)) \forall G \in \mathcal{G}$ , dato che il valore atteso per v. a. indipendenti si fattorizza. ■

Abbiamo visto il significato di  $\mathbf{E}(Y|\mathcal{G})$  se  $Y \in L^1(\Omega, \mathcal{A}, \mathbf{P})$ : è il valore atteso di  $Y$  acquisita l'informazione  $\mathcal{G}$ . Se però  $Y \in L^2(\Omega, \mathcal{A}, \mathbf{P})$ , dell'attesa condizionata si può dare un'ulteriore interpretazione, dovuta al fatto che  $L^2(\Omega, \mathcal{A}, \mathbf{P})$  è uno spazio di Hilbert: essa ha infatti un'importante proprietà di ottimizzazione.

**Prop 22.2 (proprietà di minimo dell'attesa condizionata)**

Hp :  $(\Omega, \mathcal{A}, \mathbf{P})$  spazio di probabilità,  $Y : \Omega \rightarrow \mathbb{R}$  v. a.,

$Y \in L^2(\Omega, \mathcal{A}, \mathbf{P}), \mathcal{G} \subseteq \mathcal{A}$  sotto  $\sigma$ -algebra di  $\mathcal{A}$

Ts :  $\min_{V \in L^2(\Omega, \mathcal{G}, \mathbf{P})} \mathbf{E}((Y - V)^2) = \mathbf{E}((Y - \mathbf{E}(Y|\mathcal{G}))^2)$

Il significato della tesi è che  $\mathbf{E}(Y|\mathcal{G})$  è la proiezione di  $Y$  sul sottospazio vettoriale, e spazio di Hilbert,  $L^2(\Omega, \mathcal{G}, \mathbf{P})$ , cioè l'attesa condizionata a  $\mathcal{G}$  è la migliore approssimazione di  $Y$  tramite una variabile aleatoria che è solo  $\mathcal{G}$ -misurabile (il senso di questo è che se in medias res si conosce  $\mathcal{G}$ , allora si conosce anche tale v.a.). Infatti  $\mathbf{E}((Y - V)^2)$ , in quanto norma al quadrato di  $Y - V$  (e quindi distanza al quadrato tra  $Y$  e  $V$ ), rappresenta la bontà dell'approssimazione di  $Y$  operata da  $V$ . Si sarebbe quindi potuta dare la definizione di attesa condizionata meno in generale, per  $Y \in L^2(\Omega, \mathcal{A}, \mathbf{P})$ , e dal fatto che  $\mathbf{E}(Y|\mathcal{G}) := \Pi_{L^2(\Omega, \mathcal{G}, \mathbf{P})}(Y)$  si sarebbero dedotte tutte le proprietà già viste più in generale.

$Y \in L^2$  implica  $\mathbf{E}(Y|\mathcal{G}) \in L^2$  per 22.1 (10), quindi il lato destro è ben posto. L'ipotesi implica peraltro che  $Y \in L^1(\Omega, \mathcal{A}, \mathbf{P})$ , per cui  $\mathbf{E}(Y|\mathcal{G})$  esiste ed è unico.

**Dim** Sia  $Z = \mathbf{E}(Y|\mathcal{G})$ .  $\mathbf{E}\left((Y - V)^2\right) = \mathbf{E}\left((Y - Z + Z - V)^2\right) = \mathbf{E}\left((Y - Z)^2\right) + \mathbf{E}\left((Z - V)^2\right) + 2\mathbf{E}\left((Z - V)(Y - Z)\right)$   
 $= \mathbf{E}\left((Y - \mathbf{E}(Y|\mathcal{G}))^2\right) + I_2 + I_1$ .  $\frac{1}{2}I_1 = \mathbf{E}\left((Z - V)(Y - Z)\right) = \mathbf{E}\left(\mathbf{E}\left((Z - V)(Y - Z)|\mathcal{G}\right)\right)$  per 22.1 (3). Ma  $Z - V$   
è  $\mathcal{G}$ -misurabile perché  $V$  lo è per ipotesi e  $Z$  lo è per costruzione, inoltre  $(Z - V)(Y - Z) \in L^1(\Omega, \mathcal{A}, \mathbf{P})$  perché  
ognuna delle due v. a. è in  $L^2$  ( $Z \in L^2$  perché  $Y \in L^2$ ), quindi per 22.1 (6) vale  $\mathbf{E}\left((Z - V)(Y - Z)|\mathcal{G}\right) =$   
 $(Z - V)\mathbf{E}(Y - Z|\mathcal{G})$ . Per linearità e 22.1 (3)  $\mathbf{E}(Y - Z|\mathcal{G}) = \mathbf{E}(Y|\mathcal{G}) - \mathbf{E}(Z|\mathcal{G}) = \mathbf{E}(Y|\mathcal{G}) - Z = 0$  q. c., quindi  
 $I_1 = 0$ .

Inoltre  $I_2 \geq 0$  per positività del valore atteso. Ne segue che  $\mathbf{E}\left((Y - V)^2\right) = \mathbf{E}\left((Y - \mathbf{E}(Y|\mathcal{G}))^2\right) + I_2 \geq$   
 $\mathbf{E}\left((Y - \mathbf{E}(Y|\mathcal{G}))^2\right) \forall V \in L^2(\Omega, \mathcal{G}, \mathbf{P})$ , da cui la tesi. L'uguaglianza vale se  $V = Z$ . ■

Per le proprietà della proiezione, vale inoltre  $Y - \mathbf{E}(Y|\mathcal{G}) \in (L^2(\Omega, \mathcal{G}, \mathbf{P}))^\perp$ , cioè  $\langle Y - \mathbf{E}(Y|\mathcal{G}), W \rangle_{L^2(\Omega, \mathcal{G}, \mathbf{P})} = 0$   
 $\forall W \in L^2(\Omega, \mathcal{G}, \mathbf{P})$ . Da questo segue in particolare  $\langle Y, W \rangle_{L^2(\Omega, \mathcal{G}, \mathbf{P})} = \int_\Omega YW d\mathbf{P}(\omega) = \langle \mathbf{E}(Y|\mathcal{G}), W \rangle_{L^2(\Omega, \mathcal{G}, \mathbf{P})} =$   
 $\int_\Omega \mathbf{E}(Y|\mathcal{G}) W d\mathbf{P}(\omega)$ . Se scelgo  $W = I_G$ , ritrovo la proprietà 3 richiesta in 21.4.

### 13.2 Varianza condizionata

Date  $X : (\Omega, \mathcal{A}, \mathbf{P}) \rightarrow (E, \mathcal{E})$ ,  $Y : \Omega \rightarrow \mathbb{R}$  v. a., suppongo che esista  $P^{Y|X}(dy|x)$ . Allora  $\text{var}(Y|X = x)$  può essere  
banalmente vista come  $\mathbf{E}(h(Y)|X = x)$  con  $h(Y) = (Y - m(Y))^2$ , che è già stato definito, per cui  $\text{var}(Y|X = x) =$   
 $\int_{\mathbb{R}} (y - m(x))^2 P^{Y|X}(dy|x)$ , che scegliamo di indicare con  $q^2 : E \rightarrow \mathbb{R}$ .  $q^2$  è  $\mathcal{E}$ -misurabile. Se si pone  $x = X(\omega)$ , si  
ottiene  $q^2(X(\omega))$ , che è una v. a.  $q^2 \circ X : \Omega \rightarrow \mathbb{R}$  detta varianza di  $Y$  data  $X$  e indicata con  $\text{var}(Y|X)$ . Come  
visto per l'attesa condizionata, la definizione può essere estesa alla varianza condizionata a una sotto  $\sigma$ -algebra  
generica e non  $\sigma(X)$ .

**Def 22.3** Data  $Y \in L^2(\Omega, \mathcal{A}, \mathbf{P})$  e  $\mathcal{G} \subseteq \mathcal{A}$  sotto  $\sigma$ -algebra di  $\mathcal{A}$ , si dice varianza condizionale di  $Y$  dato  $\mathcal{G}$ , e si  
indica con  $\text{var}(Y|\mathcal{G})$ , la variabile aleatoria  $\mathbf{E}\left((Y - \mathbf{E}(Y|\mathcal{G}))^2|\mathcal{G}\right)$ .

$\text{var}(Y|\mathcal{G})$  è ben definita perché  $Y \in L^2$  implica  $Y^2 \in L^1$  e  $\mathbf{E}(Y|\mathcal{G}) \in L^2$  per 22.1 (10), per cui  $(Y - \mathbf{E}(Y|\mathcal{G}))^2 \in$   
 $L^1$  ed esiste la sua attesa condizionata.

**Teo (proprietà della varianza condizionata)**

Hp:  $Y \in L^2(\Omega, \mathcal{A}, \mathbf{P})$  e  $\mathcal{G} \subseteq \mathcal{A}$  sotto  $\sigma$ -algebra di  $\mathcal{A}$

Ts: (1)  $\text{var}(Y|\mathcal{G}) \geq 0$

(2)  $\text{var}(Y|\mathcal{G}) = \mathbf{E}(Y^2|\mathcal{G}) - \mathbf{E}^2(Y|\mathcal{G})$  q. c.

(3)  $\mathbf{E}(\text{var}(Y|\mathcal{G})) = \mathbf{E}\left((Y - \mathbf{E}(Y|\mathcal{G}))^2\right)$

(4) se  $\mathcal{G} = \sigma(X)$ ,  $\text{var}(Y|\sigma(X)) = q(X)$  definita sopra

(5)  $\text{var}(Y) = \text{var}(\mathbf{E}(Y|\mathcal{G})) + \mathbf{E}(\text{var}(Y|\mathcal{G}))$

(4) mostra la coerenza con quanto fatto inizialmente. (5) è detta formula di scomposizione della varianza.

**Dim** (1)  $\text{var}(Y|\mathcal{G}) = \mathbf{E}\left[(Y - \mathbf{E}(Y|\mathcal{G}))^2|\mathcal{G}\right] \geq 0$  per la positività dell'attesa condizionata, in 22.1 (1).

(2)  $\text{var}(Y|\mathcal{G}) = \mathbf{E}\left[(Y - Z)^2|\mathcal{G}\right]$ .  $(Y - Z)^2 = Y^2 + \mathbf{E}^2(Y|\mathcal{G}) - 2YZ$ :  $\mathbf{E}\left[(Y^2 + Z^2 - 2YZ)|\mathcal{G}\right] = \mathbf{E}[Y^2|\mathcal{G}] + \mathbf{E}[Z^2|\mathcal{G}] - 2\mathbf{E}[YZ|\mathcal{G}]$ .  $\mathbf{E}[Z^2|\mathcal{G}] = Z^2$  (per 22.1 (3) e (6)),  $\mathbf{E}[YZ|\mathcal{G}] = Z\mathbf{E}[Y|\mathcal{G}] = Z^2$  per 22.1 (6), quindi si ottiene  $\text{var}(Y|\mathcal{G}) = \mathbf{E}[Y^2|\mathcal{G}] - Z^2$ .

(3) Per definizione  $\mathbf{E}(\text{var}(Y|\mathcal{G})) = \mathbf{E}\left(\mathbf{E}\left((Y - \mathbf{E}(Y|\mathcal{G}))^2|\mathcal{G}\right)\right)$ . Per 22.1 (2)  $\mathbf{E}\left(\mathbf{E}\left((Y - \mathbf{E}(Y|\mathcal{G}))^2|\mathcal{G}\right)\right) = \mathbf{E}\left((Y - \mathbf{E}(Y|\mathcal{G}))^2\right)$ .

(5) Sia  $Z = \mathbf{E}(Y|\mathcal{G})$ .  $\text{var}(Y) = \mathbf{E}\left((Y - \mathbf{E}(Y))^2\right)$ ,  $\mathbf{E}(\text{var}(Y|\mathcal{G})) = \mathbf{E}\left[\mathbf{E}\left((Y - Z)^2|\mathcal{G}\right)\right]$ ,  $\text{var}(Z) = \mathbf{E}\left((Z - \mathbf{E}(Z))^2\right)$ . Con lo stesso trucco di prima,  $\mathbf{E}\left((Y - \mathbf{E}(Y))^2\right) = \mathbf{E}\left((Y - Z)^2\right) + \mathbf{E}\left((Z - \mathbf{E}(Y))^2\right) + 2\mathbf{E}((Z - \mathbf{E}(Y))(Y - Z)) = I_1 + I_2 + I_3$ .  $2I_3 = \mathbf{E}(\mathbf{E}[(Z - \mathbf{E}(Y))(Y - Z)|\mathcal{G}])$ . Per quanto visto prima  $\mathbf{E}[(Z - \mathbf{E}(Y))(Y - Z)|\mathcal{G}] = (Z - \mathbf{E}(Y))\mathbf{E}(Y - Z|\mathcal{G})$ , quindi si ottiene  $2I_3 = \mathbf{E}[(Z - \mathbf{E}(Y))\mathbf{E}(Y - Z|\mathcal{G})]$ : applicando di nuovo 22.1 (6) si ha  $(Z - \mathbf{E}(Y))\mathbf{E}[\mathbf{E}((Y - Z)|\mathcal{G})]$ , che è nulla perché  $\mathbf{E}((Y - Z)|\mathcal{G}) = 0$ , come visto sopra. Si ha quindi  $\mathbf{E}\left((Y - \mathbf{E}(Y))^2\right) = \mathbf{E}\left((Y - Z)^2\right) + \mathbf{E}\left((Z - \mathbf{E}(Y))^2\right)$ , ma  $\mathbf{E}(Y) = \mathbf{E}(Z)$ , quindi vale la tesi, viste le uguaglianze iniziali. ■

Per la formula di decomposizione, dato che ogni varianza è nonnegativa,  $\text{var}(Y) \geq \text{var}(\mathbf{E}(Y|\mathcal{G}))$ : la dispersione dell'attesa condizionata è inferiore alla dispersione di  $Y$ , dato che si sta considerando la proiezione su uno spazio più piccolo; a livello modellistico, avendo acquisito nuove informazioni, la varianza deve diminuire. Il caso estremo si ha se  $Y \perp \mathcal{G}$ :  $\mathbf{E}(Y|\mathcal{G})$  è la costante  $\mathbf{E}(Y)$  e ha quindi varianza nulla.

Vale d'altronde  $\text{var}(Y) = \mathbf{E}\left((Y - \mathbf{E}(Y))^2\right) \geq \mathbf{E}(\text{var}(Y|\mathcal{G})) = \mathbf{E}\left((Y - \mathbf{E}(Y|\mathcal{G}))^2\right)$ . Questo ribadisce la proprietà di minimo dell'attesa condizionata, confrontandola in particolare con  $\mathbf{E}(Y)$ , che - essendo costante - è una v. a.  $\mathcal{G}$ -misurabile: l'attesa condizionata è un'approssimazione di  $Y$  migliore rispetto al valore atteso.

1.  $D \sim \mathcal{U}(\{1, \dots, 6\})$ ,  $T|D = d \sim \text{bin}(d, \frac{1}{2})$ , con  $d \in \{1, \dots, 6\}$ . Verifico la formula di decomposizione della varianza:  $\text{var}(T) = \text{var}(\mathbf{E}(T|D)) + \mathbf{E}(\text{var}(T|D))$ , per (5) nel caso particolare di  $\mathcal{G} = \sigma(D)$ .  $\mathbf{E}(T|D = d) =$

$\frac{d}{2}$  perché è il valore atteso di una binomiale, quindi  $\mathbf{E}(T|D) = \frac{D}{2}$ ; analogamente  $\text{var}(T|D=d) = \frac{d}{4}$  e  $\text{var}(T|D) = \frac{D}{4}$ . Allora  $\text{var}(T) = \text{var}\left(\frac{D}{2}\right) + \mathbf{E}\left(\frac{D}{4}\right) = \frac{1}{4}\frac{35}{12} + \frac{1}{4}\frac{7}{2} = \frac{77}{48}$ .

2.  $\begin{pmatrix} X \\ Y \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_X \\ \mu_Y \end{pmatrix}, \begin{bmatrix} \sigma_X^2 & \rho\sigma_X\sigma_Y \\ \rho\sigma_X\sigma_Y & \sigma_Y^2 \end{bmatrix}\right)$ . E' noto che  $\mathbf{E}(Y|X=x) = \mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(x - \mu_X)$ , mentre  $\text{var}(Y|X=x) = \sigma_Y^2(1 - \rho^2)$ . Vale effettivamente la formula di decomposizione della varianza, perché  $\text{var}(\mathbf{E}(Y|X)) + \mathbf{E}(\text{var}(Y|X)) = \text{var}\left(\mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(X - \mu_X)\right) + \mathbf{E}(\sigma_Y^2(1 - \rho^2)) = \rho^2\frac{\sigma_Y^2}{\sigma_X^2}\text{var}(X) + \sigma_Y^2(1 - \rho^2) = \sigma_Y^2$ .

## 14 Convergenza di successioni di variabili aleatorie

Dato  $(\Omega, \mathcal{A}, \mathbf{P})$  spazio di probabilità, sia  $(X_n)_{n \geq 1}$  una successione di variabili aleatorie, con  $X_n : \Omega \rightarrow \mathbb{R} \forall n$ . E' possibile dare diverse definizioni di convergenza di tale successione a un'altra variabile aleatoria  $X$  per  $n \rightarrow +\infty$ . Scegliamo come codominio  $\mathbb{R}$ , che è uno spazio metrico particolarmente semplice.

**Def 23.1** Dato  $(\Omega, \mathcal{A}, \mathbf{P})$  spazio di probabilità,  $(X_n)_{n \geq 1}$  successione di v. a. con  $X_n : \Omega \rightarrow \mathbb{R} \forall n$ ,  $X : \Omega \rightarrow \mathbb{R}$  v. a., si dice che  $X_n$  converge a  $X$  certamente, e si scrive  $X_n \rightarrow_{n \rightarrow +\infty} X$ , se  $\forall \omega \in \Omega \lim_{n \rightarrow +\infty} X_n(\omega) = X(\omega)$ .

[per v. a. non reali dovremmo metrizzare  $E$ ]

Fissato  $\omega$ ,  $X_n(\omega)$  è una successione numerica; la convergenza certa coincide con la convergenza puntuale vista in analisi. Tuttavia in probabilità la convergenza certa è una richiesta troppo forte nella maggior parte dei casi, che non cattura il significato modellistico della convergenza di variabili aleatorie: il fatto che un piccolo insieme di sperimentatori non osservi lo stesso fenomeno degli altri (cioè che per alcuni  $\omega$  non valga  $\lim_{n \rightarrow +\infty} X_n(\omega) = X(\omega)$ ) non è rilevante.

1. Considero lo spazio di Bernoulli, con  $\Omega = \{0, 1\}^{\mathbb{N}}$ , che rappresenti ad esempio il lancio ripetuto di una moneta. Siano  $X_1, \dots, X_n \sim i.i.d\left(\frac{1}{2}\right)$  le v. a. che valgono 1 se è uscita testa all'i-esimo lancio. Ogni  $\omega \in \Omega$  è un risultato dell'esperimento, cioè una successione di Bernoulli, e si può considerare svolto da un certo sperimentatore. Ci si aspetta che la media campionaria di tali v. a. sia tale che  $\lim_{n \rightarrow +\infty} \frac{X_1 + \dots + X_n}{n} = \frac{1}{2}$ . Tuttavia, in  $\Omega$  esiste anche  $\bar{\omega} = (1, 1, \dots)$ , per la quale  $\lim_{n \rightarrow +\infty} \frac{X_1(\bar{\omega}) + \dots + X_n(\bar{\omega})}{n} = 1$ ; oppure  $\tilde{\omega} = (0, 0, \dots)$ , per la quale  $\lim_{n \rightarrow +\infty} \frac{X_1(\tilde{\omega}) + \dots + X_n(\tilde{\omega})}{n} = 0$ : non è quindi vero che  $\frac{X_1 + \dots + X_n}{n}$  converge puntualmente a  $\frac{1}{2}$ . Vedremo che la convergenza è detta quasi certa, cioè il numero di sperimentatori per cui non vale  $\lim_{n \rightarrow +\infty} \frac{X_1 + \dots + X_n}{n} = \frac{1}{2}$  è trascurabile.

**Def 23.1** Dato  $(\Omega, \mathcal{A}, \mathbf{P})$  spazio di probabilità,  $(X_n)_{n \geq 1}$  successione di v. a. con  $X_n : \Omega \rightarrow \mathbb{R} \forall n$ ,  $X : \Omega \rightarrow \mathbb{R}$  v. a., si dice che  $X_n$  converge a  $X$  quasi certamente, e si scrive  $X_n \rightarrow_{n \rightarrow +\infty} X$  q. c., se  $A = \{\omega \in \Omega : \lim_{n \rightarrow +\infty} X_n(\omega) = X(\omega)\}$  è tale che  $\mathbf{P}(A) = 1$ .

$A \in \mathcal{A}$  perché  $\lim_{n \rightarrow +\infty} X_n$  è una v. a. (vedi criteri di misurabilità), per cui  $\mathbf{P}(A)$  è ben definita. La convergenza quasi certa si è già vista come ipotesi nei teoremi di convergenza monotona e dominata. Il significato modellistico è che quasi ogni sperimentatore osserva lo stesso fenomeno. Evidentemente, se  $X_n$  converge a  $X$  certamente, allora  $X_n$  converge a  $X$  quasi certamente, perché  $A = \Omega$ ; non vale il viceversa.

Per ogni  $\omega$  fissato l'insieme dei valori di  $X_n(\omega)$ , al variare di  $n$ , è detta traiettoria.

1. Sia  $X$  reale, per cui  $\mathbf{P}(X < +\infty) = 1$ . Definisco  $X_n(\omega) = \frac{X(\omega)}{n}$ :  $X_n \rightarrow 0$  certamente.
2. Considero  $X \geq 0$  e le  $X_n$  usate per dare la definizione di valore atteso:  $X_n \rightarrow X$  certamente per costruzione.
3. Dato  $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathcal{U}((0, 1)))$ , sia  $X_n(\omega) = I_{[0, \frac{1}{n}]}(\omega)$ . Per  $\omega \in [0, 1]^c$  evidentemente  $\lim_{n \rightarrow +\infty} X_n(\omega) = 0$ . Fissato  $\omega \in (0, 1]$ , vale  $\omega > \frac{1}{n}$  definitivamente per  $n \rightarrow +\infty$ , per cui  $X_n \rightarrow 0 \forall \omega \in \mathbb{R} \setminus \{0\}$ . Invece  $\lim_{n \rightarrow +\infty} X_n(0) = 1$ . Dunque  $X_n \rightarrow 0$  quasi certamente, essendo  $A = \mathbb{R} \setminus \{0\} : \mathbf{P}(A) = 1$ .
4. Dato  $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathcal{U}((0, 1)))$ , sia  $X_n(\omega) = nI_{[0, \frac{1}{n}]}(\omega)$ . Per  $\omega \in [0, 1]^c$   $\lim_{n \rightarrow +\infty} X_n(\omega) = 0$ . Fissato  $\omega \in (0, 1]$ , vale  $\omega > \frac{1}{n}$  definitivamente per  $n \rightarrow +\infty$ , per cui  $X_n \rightarrow 0 \forall \omega \in \mathbb{R} \setminus \{0\}$ . Invece  $\lim_{n \rightarrow +\infty} X_n(0) = +\infty$ . Dunque  $X_n \rightarrow 0$  quasi certamente, essendo  $A = \mathbb{R} \setminus \{0\} : \mathbf{P}(A) = 1$ .
5. Se  $(Y_n)_{n \geq 1}$  è una successione di v.a. tale che  $\forall \omega \in \Omega$   $Y_n(\omega)$  è crescente e superiormente limitata (e. g. data  $(X_n)_{n \geq 1}$ ,  $Y_n = \max\{X_1, \dots, X_n\}$ ), allora  $\forall \omega$  la successione reale  $(Y_n(\omega))_{n \geq 1}$  converge a  $\sup_n \{Y_n(\omega)\}$ , quindi  $Y_n(\omega)$  converge q. c. alla variabile aleatoria  $\sup_n Y_n$ . Quindi, qualora sia noto che  $Y_n \xrightarrow{\mathbf{P}} Y$  in probabilità, sicuramente  $Y_n \rightarrow Y$  q. c. per unicità del limite in probabilità. Questo si applica analogamente per successioni decrescenti e inferiormente limitate.

**Teo 23.3 (proprietà della convergenza quasi certa)**

Hp:  $(\Omega, \mathcal{A}, \mathbf{P})$  spazio di probabilità,  $(X_n)_{n \geq 1}, (Y_n)_{n \geq 1}$  successioni di v.a.,  $X_n, Y_n : \Omega \rightarrow \mathbb{R}$

$\forall n, X, Y : \Omega \rightarrow \mathbb{R}$  v.a.,  $a \in \mathbb{R}, h : \mathbb{R} \rightarrow \mathbb{R}$  continua,  $X_n \rightarrow X$  q. c.,  $Y_n \rightarrow Y$  q. c.

Ts: (1)  $aX_n \rightarrow aX$  q. c.,  $aY_n \rightarrow aY$  q. c.

(2) se  $X_n \rightarrow X$  q. c. e  $X_n \rightarrow \tilde{X}$  q. c.,  $X = \tilde{X}$  q. c.

(3)  $h(X_n) \rightarrow h(X)$  q. c.

(4)  $X_n + Y_n \rightarrow X + Y$  q. c.

(5)  $X_n Y_n \rightarrow XY$  q. c.

(6)  $\frac{X_n}{Y_n} \rightarrow \frac{X}{Y}$  q. c. se  $Y_n, Y \neq 0$

(2) significa che il limite quasi certo è unico a meno di identità quasi certa. (1) è un sottocaso di (3).

**Dim** (1) è una conseguenza delle proprietà dei limiti di successioni.

(2) Per definizione,  $A_1 = \{\omega \in \Omega : \lim_{n \rightarrow +\infty} X_n(\omega) = X(\omega)\}$  e  $A_2 = \{\omega \in \Omega : \lim_{n \rightarrow +\infty} X_n(\omega) = \tilde{X}(\omega)\}$  sono tali che  $\mathbf{P}(A_1) = \mathbf{P}(A_2) = 1$ . Poiché  $A_1 \cap A_2 \neq \emptyset$ , considero  $\omega \in A_1 \cap A_2$ : per tale  $\omega$  vale  $X(\omega) = \tilde{X}(\omega)$  per unicità del limite di successioni reali. Allora considero  $\{\omega \in \Omega : X(\omega) = \tilde{X}(\omega)\} \supseteq A_1 \cap A_2$ :  $\mathbf{P}(A_1 \cap A_2) = \mathbf{P}(A_1) + \mathbf{P}(A_2) - \mathbf{P}(A_1 \cup A_2) = 2 - 1 = 1$ , per cui  $\mathbf{P}\left(\left\{\omega \in \Omega : X(\omega) = \tilde{X}(\omega)\right\}\right) = 1$  e  $X = \tilde{X}$  q. c.

(3) deriva dalla continuità di  $h$ .

(4),(5),(6) si dimostrano come (2), usando anche le proprietà dei limiti di successioni. ■

**Def 23.4** Dato  $(\Omega, \mathcal{A}, \mathbf{P})$  spazio di probabilità,  $(X_n)_{n \geq 1}$  successione di v. a. con  $X_n : \Omega \rightarrow \mathbb{R} \forall n, X : \Omega \rightarrow \mathbb{R}$  v. a., si dice che  $X_n$  converge a  $X$  in  $L^p$ , con  $p \geq 1$ , e si scrive  $X_n \rightarrow^{L^p} X$ , se  $X_n \in L^p \forall n \geq 1, X \in L^p$  e  $\lim_{n \rightarrow +\infty} \mathbf{E}(|X_n - X|^p) = 0$ .

$\mathbf{E}(|X_n - X|^p)$  è una successione numerica (ben definita perché  $X_n \in L^p \forall n, X \in L^p$ ), che può essere scritta come  $\left(\|X_n - X\|_p\right)^p$  (elevare alla  $p$  non cambia la convergenza a 0): di fatto la convergenza in  $L^p$  è una convergenza in norma  $p$ -esima (in particolare metrica), che è una proprietà in media, in cui i singoli sperimentatori (cioè i singoli  $\omega$ , le traiettorie) non sono considerati.

Se  $X_n \rightarrow^{L^p} X$ , ogni sottosuccessione di  $X_n$  converge a  $X$  in  $L^p$ .



**Prop 23.5 (proprietà della convergenza in  $L^p$ )**

Hp:  $(\Omega, \mathcal{A}, \mathbf{P})$  spazio di probabilità,  $(X_n)_{n \geq 1}, (Y_n)_{n \geq 1}$  successioni di v.a.,

$$X_n, Y_n : \Omega \rightarrow \mathbb{R} \forall n, X, Y : \Omega \rightarrow \mathbb{R} \text{ v.a.}, a \in \mathbb{R}, X_n \xrightarrow{L^p} X, Y_n \xrightarrow{L^p} Y$$

Ts: (1) se  $X_n \xrightarrow{L^p} X$  e  $X_n \xrightarrow{L^p} \tilde{X}$ ,  $X = \tilde{X}$  q. c.

$$(2) aX_n \xrightarrow{L^p} aX$$

$$(3) X_n + Y_n \xrightarrow{L^p} X + Y$$

$$(4) \mathbf{E}(|X_n|^p) \rightarrow \mathbf{E}(|X|^p)$$

(1) è l'unicità quasi certa del limite. Si noti che, mentre è lecito considerare la convergenza in  $L^p$  di  $X_n + Y_n$  perché gli  $L^p$  sono spazi vettoriali, non lo è considerare quella di  $X_n Y_n$ , che in generale non appartiene a  $L^p$ . (4) significa che la successione numerica delle norme  $p$ -esime converge alla norma  $p$ -esima del limite. La convergenza in  $L^p$  serve ad approssimare i momenti. (!!)

1. Dimostro che se  $X_n \xrightarrow{L^2} X$  allora  $\text{var}(X_n) \rightarrow \text{var}(X) = \mathbf{E}(X^2) - \mathbf{E}^2(X)$ . So che  $X_n \in L^2 \forall n \geq 1, X \in L^2$  e  $\lim_{n \rightarrow +\infty} \mathbf{E}(|X_n - X|^2) = 0$ . Poiché  $\text{var}(X) = \mathbf{E}(X^2) - \mathbf{E}^2(X)$  e  $\text{var}(X_n) = \mathbf{E}(X_n^2) - \mathbf{E}^2(X_n)$ , Voglio dimostrare che  $\lim_{n \rightarrow +\infty} \mathbf{E}(X_n) = \mathbf{E}(X)$  e  $\lim_{n \rightarrow +\infty} \mathbf{E}(X_n^2) = \mathbf{E}(X^2)$ . La seconda uguaglianza è vera per 23.5 (4) con  $p = 2$ . Per la prima basterebbe usare il fatto che  $X_n \xrightarrow{L^1} X$ , così anche la successione dei momenti primi converge; tuttavia, si dimostra direttamente.

Per la disuguaglianza di Cauchy-Schwarz con  $Y = 1$ ,  $|\mathbf{E}(X_n - X)| \leq \sqrt{\mathbf{E}((X_n - X)^2)}$ , e il lato destro tende a 0 per  $n \rightarrow +\infty$ : per il teorema dei carabinieri,  $\lim_{n \rightarrow +\infty} |\mathbf{E}(X_n - X)| = 0$ , cioè  $\lim_{n \rightarrow +\infty} \mathbf{E}(X_n) = \mathbf{E}(X)$  per linearità del valore atteso e continuità del valore assoluto. Quindi  $\lim_{n \rightarrow +\infty} \text{var}(X_n) = \lim_{n \rightarrow +\infty} (\mathbf{E}(X_n^2) - \mathbf{E}^2(X_n)) = \mathbf{E}(X^2) - \mathbf{E}^2(X)$ , cioè  $\text{var}(X)$ .

**Def 23.6** Dato  $(\Omega, \mathcal{A}, \mathbf{P})$  spazio di probabilità,  $(X_n)_{n \geq 1}$  successione di v. a. con  $X_n : \Omega \rightarrow \mathbb{R} \forall n, X : \Omega \rightarrow \mathbb{R}$  v. a., si dice che  $X_n$  converge a  $X$  in probabilità, e si scrive  $X_n \xrightarrow{\mathbf{P}} X$ , se  $\forall \varepsilon > 0 \exists \lim_{n \rightarrow +\infty} \mathbf{P}(|X_n - X| > \varepsilon) = 0$ .

La definizione, significativa per  $\varepsilon$  piccoli, indica che la percentuale di sperimentatori per cui la distanza tra  $X_n$  e  $X$  supera  $\varepsilon$  tende a 0 per  $n \rightarrow +\infty$ .  $|X_n - X| > \varepsilon \in \mathcal{A}$  perché  $X_n, X$  sono v. a., il valore assoluto è continuo e la somma di v.a. è una v.a.

1. Se  $X_n \xrightarrow{\mathbf{P}} X$ ,  $\lim_{n \rightarrow +\infty} \mathbf{P}(|X_n - X| = x) = 0 \forall x \neq 0$ , perché se si applica 23.6 con  $\varepsilon < x$  si ha  $(|X_n - X| = x) \subseteq (|X_n - X| > \varepsilon)$ , quindi  $\mathbf{P}(|X_n - X| = x) \leq \mathbf{P}(|X_n - X| > \varepsilon)$  e  $\mathbf{P}(|X_n - X| = x)$  tende a 0 per il teorema dei carabinieri.

**Prop 23.7 (caratterizzazione della convergenza in probabilità)**

Hp:  $(\Omega, \mathcal{A}, \mathbf{P})$  spazio di probabilità,  $(X_n)_{n \geq 1}$  successione di v. a. r.,  $X : \Omega \rightarrow \mathbb{R}$  v. a.

Ts:  $X_n \rightarrow^{\mathbf{P}} X \iff \mathbf{E} \left( \frac{|X_n - X|}{1 + |X_n - X|} \right) \rightarrow 0$  per  $n \rightarrow +\infty$

$\mathbf{E} \left( \frac{|X_n - X|}{1 + |X_n - X|} \right)$  è ben definito per qualsiasi  $X_n, X$  perché l'argomento è limitato. La proposizione chiarisce che anche la convergenza in probabilità può essere vista come convergenza metrica:  $\mathbf{E} \left( \frac{|X_n - X|}{1 + |X_n - X|} \right)$  definisce una metrica sull'insieme delle v. a. r., che diventa così uno spazio metrico completo. Infatti  $d(X_n, X) = \mathbf{E} \left( \frac{|X_n - X|}{1 + |X_n - X|} \right)$  è nulla se e solo se  $X_n = X$  q. c., è chiaramente simmetrica, e per disuguaglianza triangolare del valore assoluto, monotonia e linearità del valore atteso  $\mathbf{E} \left( \frac{|X - Y|}{1 + |X - Y|} \right) = \mathbf{E} \left( 1 - \frac{1}{1 + |X - Y|} \right) = \mathbf{E} \left( 1 - \frac{1}{1 + |X - Z| + |Z - Y|} \right) \leq \mathbf{E} \left( \frac{|X - Z| + |Z - Y|}{1 + |X - Z| + |Z - Y|} \right) = \mathbf{E} \left( \frac{|X - Z|}{1 + |X - Z| + |Z - Y|} \right) + \mathbf{E} \left( \frac{|Z - Y|}{1 + |X - Z| + |Z - Y|} \right) \leq \mathbf{E} \left( \frac{|X - Z|}{1 + |X - Z|} \right) + \mathbf{E} \left( \frac{|Z - Y|}{1 + |Z - Y|} \right)$ . Quindi  $X_n$  converge in probabilità a  $X$  se e solo se  $X_n$  converge a  $X$  secondo la metrica data (il che implica che sia una successione di Cauchy secondo tale metrica).

Questo teorema facilita molto la verifica della convergenza in probabilità, riducendola alla verifica della convergenza di una successione numerica: non è la convergenza in  $L^1$ , ma una proprietà più debole, dato che il denominatore controlla le traiettorie per cui il numeratore diventa grande e rende più facile la convergenza.

Valgono le stesse proprietà della convergenza quasi certa.

**Teo 23.8 (proprietà della convergenza in probabilità)**

Hp:  $(\Omega, \mathcal{A}, \mathbf{P})$  spazio di probabilità,  $(X_n)_{n \geq 1}, (Y_n)_{n \geq 1}$  successioni di v.a.r.,

$X, Y : \Omega \rightarrow \mathbb{R}$  v.a.,  $a \in \mathbb{R}, h : \mathbb{R} \rightarrow \mathbb{R}$  continua,  $X_n \rightarrow^{\mathbf{P}} X, Y_n \rightarrow^{\mathbf{P}} Y$

Ts: (1)  $aX_n \rightarrow^{\mathbf{P}} aX, aY_n \rightarrow^{\mathbf{P}} aY$

(2) se  $X_n \rightarrow^{\mathbf{P}} X$  e  $X_n \rightarrow^{\mathbf{P}} \tilde{X}, X = \tilde{X}$  q. c.

(3)  $h(X_n) \rightarrow^{\mathbf{P}} h(X)$

(4)  $X_n + Y_n \rightarrow^{\mathbf{P}} X + Y$

(5)  $X_n Y_n \rightarrow^{\mathbf{P}} XY$

(6)  $\frac{X_n}{Y_n} \rightarrow^{\mathbf{P}} \frac{X}{Y}$  se  $Y_n, Y \neq 0$

(2) ancora q. c. Si dimostrano usando la seconda definizione.

**Teo 23.9 (relazioni tra convergenze)**

Hp:  $(\Omega, \mathcal{A}, \mathbf{P})$  spazio di probabilità,  $(X_n)_{n \geq 1}$  successione di v.a. r.,  $X : \Omega \rightarrow \mathbb{R}$  v.a

Ts: (1)  $X_n \rightarrow X$  q. c.  $\implies X_n \rightarrow^{\mathbf{P}} X$

(2)  $X_n \rightarrow^{L^p} X \implies X_n \rightarrow^{L^q} X$  e  $X_n \rightarrow^{\mathbf{P}} X \forall 1 \leq q \leq p$

La tesi rende evidente che la convergenza in legge è la proprietà più debole. Usando (1), dalla convergenza certa si può dedurre la convergenza in probabilità. La convergenza in  $L^p$  non implica né è implicata dalla convergenza quasi certa.

**Dim** (1) L'idea è di usare il teorema di convergenza monotona o dominata per sfruttare la caratterizzazione della convergenza in probabilità. Dato che  $X_n \rightarrow X$  q. c.,  $Z_n = \frac{|X_n - X|}{1 + |X_n - X|} \rightarrow 0$  q. c. per le proprietà del limite puntuale e la continuità del valore assoluto. Inoltre  $|Z_n| \leq 1 \forall n$ , e la v. a. costante 1 è integrabile, per cui, per il teorema di convergenza dominata,  $\lim_{n \rightarrow +\infty} \mathbf{E}(Z_n) = \mathbf{E}(0) = 0$ , perciò  $X_n \rightarrow^{\mathbf{P}} X$  per 23.7.

(2) \*Mostro che  $X_n \rightarrow^{L^2} X \implies X_n \rightarrow^{L^1} X$ . Per la disuguaglianza di Cauchy-Schwarz con  $Y = 1$ ,  $|\mathbf{E}(X_n - X)| \leq \sqrt{\mathbf{E}((X_n - X)^2)}$ , e il lato destro tende a 0 per  $n \rightarrow +\infty$ : per il teorema dei carabinieri,  $\lim_{n \rightarrow +\infty} |\mathbf{E}(X_n - X)| = 0$ , cioè  $\lim_{n \rightarrow +\infty} \mathbf{E}(X_n) = \mathbf{E}(X)$  per linearità del valore atteso e continuità del valore assoluto.

Mostro solo che  $X_n \rightarrow^{L^1} X \implies X_n \rightarrow^{\mathbf{P}} X$ . So che  $\lim_{n \rightarrow +\infty} \mathbf{E}(|X_n - X|) = 0$ ; voglio mostrare che  $\forall \varepsilon > 0$   $\lim_{n \rightarrow +\infty} \mathbf{P}(|X_n - X| > \varepsilon) = 0$ . Dalla disuguaglianza di Markov è noto che  $\mathbf{P}(|X_n - X| > \varepsilon) \leq \frac{\mathbf{E}(|X_n - X|)}{\varepsilon}$ , quindi per il teorema dei carabinieri  $\lim_{n \rightarrow +\infty} \mathbf{P}(|X_n - X| > \varepsilon) = 0$ . ■

Se  $X_n \rightarrow X$  q. c. e  $X_n \rightarrow Y$  in  $L^p$  allora  $X = Y$  q. c. per unicità q. c. del limite in probabilità.

1. Se  $X_n \sim$ , le  $X_n$  sono iid e  $T_n = \max\{X_1, \dots, X_n\} \rightarrow^{\mathbf{P}} c$ , allora  $T_n \rightarrow c$  q. c.. Infatti, è noto che  $\forall \varepsilon > 0$   $\lim_{n \rightarrow +\infty} \mathbf{P}(|T_n - c| < \varepsilon) = 1$ : ma l'evento  $A = (T_n \rightarrow c)$  è uguale a  $\bigcap_{\varepsilon > 0} A_\varepsilon$ , con  $A_\varepsilon = \bigcap_{n=\nu}^{+\infty} (|T_n - c| < \varepsilon)$ , e  $A_{n+1} = (|T_{n+1} - c| < \varepsilon) \subseteq (|T_n - c| < \varepsilon) = A_n$ , cioè la successione degli eventi  $A_n$  è decrescente. Dunque  $\mathbf{P}(A) = 1$  perché  $\mathbf{P}(A_\varepsilon) = \lim_{n \rightarrow +\infty} \mathbf{P}(|T_n - c| < \varepsilon) = 1 \forall \varepsilon > 0$ .

**Teo 23.10 (la rivincita della convergenza in probabilità)**

Hp:  $(\Omega, \mathcal{A}, \mathbf{P})$  spazio di probabilità,  $(X_n)_{n \geq 1}$  successione

di v.a. r.,  $X : \Omega \rightarrow \mathbb{R}$  v.a,  $X_n \rightarrow^{\mathbf{P}} X$

(1)  $\exists (n_k)_{k \geq 1} : X_{n_k} \rightarrow X$  q. c. per  $k \rightarrow +\infty$

(2) se inoltre  $\exists Y \in L^p : |X_n| \leq Y \forall n$ , allora  $X_n \in L^p, X \in L^p, X_n \rightarrow^{L^p} X$

Se  $p = 1$ , (2) diventa il solito teorema di convergenza dominata: infatti, se  $\lim_{n \rightarrow +\infty} \mathbf{E}(|X_n - X|) = 0$ , allora  $\lim_{n \rightarrow +\infty} |\mathbf{E}(X_n - X)| = 0$  perché  $0 \leq |\mathbf{E}(X_n - X)| \leq \mathbf{E}(|X_n - X|)$  e dunque  $\lim_{n \rightarrow +\infty} \mathbf{E}(X_n - X) = 0$ .

**Dim\*** (1) Per ipotesi  $\lim_{n \rightarrow +\infty} \mathbf{E}\left(\frac{|X_n - X|}{1 + |X_n - X|}\right) = 0$ : quindi  $\exists (n_k)_{k \geq 1} : \lim_{k \rightarrow +\infty} \mathbf{E}\left(\frac{|X_{n_k} - X|}{1 + |X_{n_k} - X|}\right) = 0$  e  $\mathbf{E}\left(\frac{|X_{n_k} - X|}{1 + |X_{n_k} - X|}\right) \leq \frac{1}{2^k}$ . Allora per 9.4 (1), essendo il termine generale della serie una v. a. nonnegativa,  $\mathbf{E}\left(\sum_{k=1}^{+\infty} \frac{|X_{n_k} - X|}{1 + |X_{n_k} - X|}\right) = \sum_{k=1}^{+\infty} \mathbf{E}\left(\frac{|X_{n_k} - X|}{1 + |X_{n_k} - X|}\right) \leq \sum_{k=1}^{+\infty} \frac{1}{2^k} < +\infty$ . Se  $\mathbf{E}\left(\sum_{k=1}^{+\infty} \frac{|X_{n_k} - X|}{1 + |X_{n_k} - X|}\right)$  è finito, allora  $\sum_{k=1}^{+\infty} \frac{|X_{n_k} - X|}{1 + |X_{n_k} - X|} < +\infty$  q. c. (una v. a. con valore atteso finito è quasi certamente finita). Questo implica che il termine generale  $\frac{|X_{n_k} - X|}{1 + |X_{n_k} - X|} \rightarrow 0$  per  $k \rightarrow +\infty$  q. c., cioè  $|X_{n_k} - X| \rightarrow 0$  q. c., cioè  $X_{n_k} \rightarrow X$  q. c. ■

1. Si dà un esempio del fatto che la convergenza in  $L^1$  non implica la convergenza quasi certa, detto esempio della macchina da scrivere. Considero lo spazio di probabilità  $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathcal{U}((0, 1)))$  e la v. a.  $X_{n,k}(\omega) = I_{[\frac{k-1}{n}, \frac{k}{n})}(\omega)$  con  $k = 1, \dots, n$  (si fa corrispondere a ogni  $(n, k)$  una  $j$  in modo da avere un solo indice:  $Y_1 := X_{1,1}, Y_2 := X_{2,1}, Y_3 := X_{2,2}, \dots$ ). Fissato  $n$ , al variare di  $k$   $[\frac{k-1}{n}, \frac{k}{n})$  è una partizione di  $[0, 1)$  mediante intervalli di ampiezza  $\frac{1}{n}$ , che si infittisce al crescere di  $n$ :  $X_{1,1}(\omega) = I_{[0,1)}(\omega)$ ,  $X_{2,1}(\omega) = I_{[0, \frac{1}{2})}(\omega)$ ,  $X_{2,2}(\omega) = I_{[\frac{1}{2}, 1)}(\omega), \dots$   $\mathbf{E}(|X_{n,k}|) = \mathbf{E}\left(I_{[\frac{k-1}{n}, \frac{k}{n})}(\omega)\right) = \mathbf{P}\left([\frac{k-1}{n}, \frac{k}{n})\right) = \frac{1}{n}$ , quindi  $\lim_{n \rightarrow +\infty} \mathbf{E}(|X_{n,k}|) = 0$ , dalla qual cosa si intuisce che  $X_{n,k} \xrightarrow{L^1} 0$ : ogni  $X_{n,k}$  è integrabile, 0 lo è e  $\lim_{n \rightarrow +\infty} \mathbf{E}(|X_{n,k} - 0|) = 0$ . Allora, se  $X_n$  ha un limite quasi certo, questo deve essere 0, per unicità quasi certa del limite!. Non vale però che  $X_{n,k} \rightarrow 0$  q. c., perché, fissato  $\bar{\omega} \in [0, 1)$ ,  $X_{n,k}(\bar{\omega})$  oscilla tra 0 e 1:  $\forall n \geq 1 \exists \bar{k} : X_{n,\bar{k}}(\bar{\omega}) = 1$  e per  $k \neq \bar{k}$  vale  $X_{n,k}(\bar{\omega}) = 0$ . Non è quindi possibile che  $\lim_{n \rightarrow +\infty} X_n(\bar{\omega}) = X(\bar{\omega}) = 0$ , per nessun  $\bar{\omega}$ , e la convergenza non è quasi certa. Questo si può dedurre argomentando che la convergenza a  $X = 0$  in  $L^1$  implica la convergenza in probabilità, che implica che  $\exists (n_k)_{k \geq 1} : X_{n_k} \rightarrow 0$  q. c. per  $k \rightarrow +\infty$ . Tuttavia  $\{\omega : \lim_{n \rightarrow +\infty} X_n(\omega) \neq 0\} = [0, 1)$ , per cui  $X_n \not\rightarrow 0$  q. c.

2. Considero una sottosuccessione di  $X_{n,k}$ :  $Y_n := X_{n,1} \forall n \geq 1$ .  $X_{n,1}(\omega) = I_{[0, \frac{1}{n})}(\omega)$ :  $\lim_{n \rightarrow +\infty} X_{n,1}(\omega) = 0 \forall \omega \neq 0$ , quindi  $Y_n \rightarrow 0$  q. c. Inoltre  $Y_n \xrightarrow{L^1} 0$ : ogni  $Y_n$  è integrabile, 0 lo è e  $\lim_{n \rightarrow +\infty} \mathbf{E}(|Y_n - 0|) = \lim_{n \rightarrow +\infty} \frac{1}{n} = 0$ , coerentemente con 23.10 (1)
3. Dato lo spazio di probabilità  $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathcal{U}((0, 1)))$ , si è già visto che  $X_n = nI_{[0, \frac{1}{n})} \rightarrow 0$  q. c. Non vale però  $X_n \xrightarrow{L^1} 0$ : ogni  $X_n$  è integrabile perché limitata e 0 lo è, ma  $\mathbf{E}(|X_n|) = \mathbf{E}\left(nI_{[0, \frac{1}{n})}\right) = n\mathbf{P}\left([0, \frac{1}{n})\right) = 1$ , quindi  $\lim_{n \rightarrow +\infty} \mathbf{E}(|X_n - 0|) = 1$ .

## 14.1 Convergenza debole

**Def 24.1** Data  $(Q_n)_{n \geq 1}$  successione di misure di probabilità,  $Q$  misura di probabilità su  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ , si dice che  $Q_n$  converge debolmente a  $Q$  se  $\forall h : \mathbb{R} \rightarrow \mathbb{R}$  continua e limitata vale  $\lim_{n \rightarrow +\infty} \int_{\mathbb{R}} h(x) dQ_n(x) = \int_{\mathbb{R}} h(x) dQ(x)$ .

$\int_{\mathbb{R}} h dQ_n$  è una successione numerica.  $h$  è detta funzione test; la classe delle funzioni test  $\{h : \mathbb{R} \rightarrow \mathbb{R} : h \text{ è continua e limitata}\}$  è indicata con  $C_b(\mathbb{R})$ . La definizione richiede di testare un limite su tutte le  $h$  siffatte, che grazie alle ipotesi aggiuntive costituiscono un insieme più piccolo delle funzioni solo misurabili. La continuità di  $h$  ne implica la misurabilità, la limitatezza l'integrabilità, per cui  $\int_{\mathbb{R}} h dQ_n, \int_{\mathbb{R}} h dQ$  sono ben definiti.

1. La convergenza debole di misure è una generalizzazione della consueta convergenza di una successione numerica. Considero una successione a valori reali  $x_n \rightarrow x$ , e  $Q_n(x) = \delta_{x_n}(x)$ .  $\int_{\mathbb{R}} h(x) dQ_n(x) = h(x_n)$ , dato che si sta calcolando il valore atteso di  $h(X_n(\omega))$ , dove la legge di  $X_n$  è una delta di Dirac  $\delta_{x_n}$ .  $\lim_{n \rightarrow +\infty} h(x_n) = h(x) = \int_{\mathbb{R}} h(x) dQ(x)$ , grazie alla continuità di  $h$  (la sola misurabilità non è sufficiente). Allora  $Q_n = \delta_{x_n}$  converge debolmente a  $Q = \delta_x$ . Vale anche il viceversa.

La definizione di convergenza debole serve a poter dare una nozione di convergenza anche per una successione di v. a. r. non tutte definite sullo stesso spazio di probabilità: in tal caso infatti una scrittura del tipo  $X_n(\omega) - X(\omega)$  non ha significato, perché non è chiaro a quale insieme appartenga  $\omega$ . E' solo possibile confrontare le leggi delle v. a.

**Def 24.2** Dato  $(\Omega_n, \mathcal{A}_n, \mathbf{P}_n)$  spazio di probabilità  $\forall n$ ,  $(\Omega, \mathcal{A}, \mathbf{P})$  spazio di probabilità,  $(X_n)_{n \geq 1}$  successione di v. a. con  $X_n : \Omega_n \rightarrow \mathbb{R} \forall n$ ,  $X : \Omega \rightarrow \mathbb{R}$  v. a., si dice che  $X_n$  converge a  $X$  in legge, e si scrive  $X_n \rightarrow X$  o  $X_n \rightarrow^{\mathcal{D}} X$ , se la successione di misure di probabilità  $(P^{X_n})_{n \geq 1}$  converge debolmente a  $P^X$ .

D'ora in poi, quando si scrive  $X_n \rightarrow X$ , si suppone implicitamente che siano v. a. r. definite su spazi di probabilità come sopra.

1. Se  $(X_n)_{n \geq 1}$  è una successione di v. a. costante rispetto a  $n$ , cioè tale che  $X_n = X_1 \forall n$ , allora  $X_n \rightarrow X$  certamente, q. c., in  $L^p \forall p$ , in probabilità, in legge.
2. Se  $(X_n)_{n \geq 1}$  è una successione di v. a. costante rispetto a  $\omega$ , cioè tale che  $\forall n$   $X_n$  non dipende da  $\omega$ , cioè è una successione di v. a. costanti (a valori reali), allora  $\lim_{n \rightarrow +\infty} X_n$  è il limite certo, q. c., in  $L^p \forall p$ , in probabilità, in legge di  $X_n$ .

**Prop 24.3 (caratterizzazione della convergenza in legge)**

Hp:  $(X_n)_{n \geq 1}$  successione di v. a.,  $X : \Omega \rightarrow \mathbb{R}$  v. a.

Ts:  $X_n \rightarrow X \iff \lim_{n \rightarrow +\infty} \mathbf{E}_n(h(X_n)) = \mathbf{E}(h(X)) \quad \forall h \in C_b(\mathbb{R})$

Questo significa che se  $X_n \rightarrow X$  e  $Y$  ha la stessa legge di  $X$ , allora vale anche  $X_n \rightarrow Y$ , perché il valore atteso dipende solo dalla legge. Non implica  $\lim_{n \rightarrow +\infty} \mathbf{E}_n(X_n) = \mathbf{E}(X)$ .

**Dim** Per definizione  $X_n \rightarrow X \iff \forall h \in C_b(\mathbb{R})$  vale  $\lim_{n \rightarrow +\infty} \int_{\mathbb{R}} h(x) dP^{X_n}(x) = \int_{\mathbb{R}} h(x) dP^X(x)$ . Per la regola del valore atteso  $\int_{\mathbb{R}} h(x) dP^{X_n}(x) = \int_{\Omega} h(X_n(\omega)) d\mathbf{P}_n(\omega)$  e  $\int_{\mathbb{R}} h(x) dP^X(x) = \int_{\Omega} h(X(\omega)) d\mathbf{P}(\omega)$ , quindi quanto scritto è equivalente a  $\lim_{n \rightarrow +\infty} \mathbf{E}_n(h(X_n)) = \mathbf{E}(h(X))$ , dove  $\mathbf{E}_n$  indica l'integrale rispetto alla misura  $\mathbf{P}_n$ . ■

Vale l'unicità in legge del limite in legge: se  $X_n \rightarrow X$  e  $X_n \rightarrow Y$ , allora  $P^X = P^Y$ , che non implica affatto  $X = Y$  q. c.:  $X$  e  $Y$  potrebbero anche essere definite su spazi diversi.

1. Considero  $(X_n)_{n \geq 1}$  su  $(\Omega, \mathcal{A}, \mathbf{P})$ ,  $X \sim b(\frac{1}{2})$  e  $X_n = X \quad \forall n \geq 1$ : ovviamente  $X_n \rightarrow X$ . Considero  $Y = 1 - X$ : anche  $Y \sim b(\frac{1}{2})$  (si può ricavare facilmente con la funzione caratteristica) e  $X_n \rightarrow Y$ , ma  $\mathbf{P}(X = Y) = \mathbf{P}(X = Y = \frac{1}{2}) = 0$ , quindi, pur essendo  $P^X = P^Y$ , non è vero che  $X = Y$  q. c.

**Prop 24.4 (convergenza in legge nel caso discreto)**

Hp:  $(X_n)_{n \geq 1}$  successione di v.a.r. discrete,  $X : \Omega \rightarrow \mathbb{R}$  v. a. discreta

con supporti  $S_{X_n} \quad \forall n, S_X$  e densità discrete  $p_{X_n} \quad \forall n, p_X$ ;  $S := \bigcup_{n=1}^{+\infty} S_{X_n} \cup S$

Ts: (1) se  $p_{X_n}(s) \rightarrow p_X(s) \quad \forall s \in S$ ,  $X_n \rightarrow X$

(2) se  $\exists A : \mathcal{D}A = \emptyset$  e  $S_{X_n}, S \subseteq A \quad \forall n$ ,

$p_{X_n}(s) \rightarrow p_X(s) \quad \forall s \in S \iff X_n \rightarrow X$

Essendo i supporti tutti discreti, anche  $S$  è discreto. (1) afferma che la convergenza puntuale della successione delle densità è sufficiente per la convergenza in legge. (2) richiede che i supporti siano tutti contenuti in un insieme privo di punti di accumulazione: e. g.  $\mathbb{N}$  o  $\mathbb{Z}$ . La valutazione di una convergenza puntuale è molto più semplice della verifica della definizione.

1. Sia  $X_n \sim \text{bin}(n, \frac{\lambda}{n})$ ,  $\lambda > 0$ ,  $n \geq \lambda$ .  $S_{X_n} = \{0, \dots, n\} \subseteq \mathbb{N}$ . Considero  $X \sim \text{poiss}(\lambda)$ , con  $S_X = \mathbb{N}$ : dunque  $S = \mathbb{N}$  e posso applicare (2).  $p_{X_n}(k) = \binom{n}{k} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k}$ , che converge a  $e^{-\lambda} \frac{\lambda^k}{k!}$  per  $n \rightarrow +\infty$ . Infatti

$\frac{n!}{k!(n-k)!} \frac{\lambda^k}{n^k} \frac{(n-\lambda)^{n-k}}{n^{n-k}} = \frac{\lambda^k}{k!} \frac{n!}{(n-k)!} \frac{(n-\lambda)^{n-k}}{n^n} \sim \frac{\lambda^k}{k!} \frac{n^n e^{-n} \sqrt{2\pi n}}{(n-k)^{n-k} e^{-(n-k)} \sqrt{2\pi(n-k)}} \frac{(n-\lambda)^{n-k}}{n^n}$  per l'approssimazione di Sterling: quindi si ha  $\frac{\lambda^k}{k!} \sqrt{\frac{n}{n-k}} e^{-k} \left(\frac{n-\lambda}{n-k}\right)^{n-k}$ , e in particolare  $\left(\frac{n-\lambda}{n-k}\right)^{n-k} = \left(1 + \frac{k-\lambda}{n-k}\right)^{n-k} = \left(1 + \frac{1}{\frac{n-k}{k-\lambda}}\right)^{n-k} = \left(\left(1 + \frac{1}{\frac{n-k}{k-\lambda}}\right)^{\frac{n-k}{k-\lambda}}\right)^{k-\lambda}$ , che per  $n \rightarrow +\infty$  tende a  $e^{k-\lambda}$ , per cui il risultato finale è  $\frac{\lambda^k}{k!} e^{-k} e^{k-\lambda} = e^{-\lambda} \frac{\lambda^k}{k!}$  e  $X_n$  converge in legge a  $X$ .

2. Sia  $X_n \sim \delta_{\frac{1}{n}}$ .  $S_{X_n} = \{0, 1\} \subseteq \mathbb{N}$ . Considero  $X \sim \delta_0$ : so che  $X_n \rightarrow X$  perché si sta studiando la convergenza di una successione numerica, come nell'esempio sopra. Studio però anche la convergenza puntuale della successione delle densità;  $S = \{0\} \cup \{\frac{1}{n} : n \in \mathbb{N}^*\}$ .  $p_{X_n}(\frac{1}{n}) = 1$ ,  $p_{X_n}(0) = 0 \forall n$ :  $\lim_{n \rightarrow +\infty} p_{X_n}(0) = 0$ , mentre  $p_X(0) = 1$ , quindi non è vero che  $p_{X_n}(s) \rightarrow p_X(s) \forall s \in S$ , nonostante  $X_n \rightarrow X$ . Infatti non si può applicare (2) perché  $\mathcal{D}S = \{0\}$ .

**Prop 24.5 (lemma di Scheffé: convergenza in legge nel caso continuo)**

Hp:  $(X_n)_{n \geq 1}$  successione di v.a.r. assolutamente continue,  $X : \Omega \rightarrow \mathbb{R}$

v. a. assolutamente continua con densità  $(f_{X_n})_{n \geq 1}, f_X$  rispettivamente

Ts: se  $f_{X_n} \rightarrow f_X$  quasi ovunque, allora  $X_n \rightarrow X$

Il teorema è del tutto analogo a 24.4 (1).

**Dim** So che  $\{x : \lim_{n \rightarrow +\infty} f_{X_n}(x) \neq f_X(x)\}$  ha misura 0 secondo Lebesgue. Devo mostrare che  $\forall h \in C_b(\mathbb{R})$  vale  $\lim_{n \rightarrow +\infty} \mathbf{E}_n(h(X_n)) = \mathbf{E}(h(X))$ , cioè  $\lim_{n \rightarrow +\infty} \int_{\mathbb{R}} h(x) f_{X_n}(x) dm = \int_{\mathbb{R}} h(x) f_X(x) dm$ , quindi che  $\lim_{n \rightarrow +\infty} \int_{\mathbb{R}} h(x) [f_{X_n}(x) - f_X(x)] dm = 0$ .

Dimostro che  $\lim_{n \rightarrow +\infty} \int_{\mathbb{R}} |f_{X_n} - f_X| dm = 0$ .  $\int_{\mathbb{R}} |f_{X_n} - f_X| dm = 2 \int_{\mathbb{R}} (f_{X_n} - f_X)_+ dm$

1. Sia  $X_n \sim \mathcal{N}(\mu_n, \sigma_n^2)$  con  $\mu_n \rightarrow \mu$ ,  $\sigma_n^2 \rightarrow \sigma^2 > 0$ . Allora  $f_{X_n}(x) = \frac{1}{\sqrt{2\pi\sigma_n^2}} e^{-\frac{1}{2} \frac{(x-\mu_n)^2}{\sigma_n^2}}$  e  $\lim_{n \rightarrow +\infty} f_{X_n}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2} \frac{(x-\mu)^2}{\sigma^2}} = f_X(x) \forall x$ , con  $X \sim \mathcal{N}(\mu, \sigma^2)$ . Dunque  $X_n \rightarrow X$ .

I criteri visti hanno l'evidente limite di essere applicabili solo a particolari classi di v. a. Si vedono nel seguito criteri basati su altre due funzioni caratterizzanti una v. a., cioè funzione di ripartizione e funzione caratteristica.

In generale  $X_n \rightarrow X$  non implica che  $P^{X_n}(B) \rightarrow P^X(B) \forall B \in \mathcal{B}(\mathbb{R})$ . Infatti dalla convergenza in legge non si può dedurre che  $\mathbf{E}_n(I_B(X_n)) \rightarrow \mathbf{E}(I_B(X))$ , perché in generale  $I_B(x)$  non è continua. Se però  $B$  è del tipo  $(-\infty, t]$ , le cose cambiano.

**Teo 24.6 (convergenza in legge e convergenza delle funzioni di ripartizione)**

Hp:  $(X_n)_{n \geq 1}$  successione di v.a.r.,  $X : \Omega \rightarrow \mathbb{R}$  v. a. con funzioni

di ripartizione  $(F_{X_n})_{n \geq 1}, F_X$  rispettivamente

Ts:  $X_n \rightarrow X \iff \lim_{n \rightarrow +\infty} F_{X_n}(t) = F_X(t) \quad \forall t \in \{t : F_X \text{ è continua in } t\}$

Quindi  $X_n \rightarrow X \iff P^{X_n}((-\infty, t]) \rightarrow P^X((-\infty, t])$  per  $t$  che sta nell'insieme dei punti di continuità di  $F_X$ :

$F_X$  è continua in  $t$  se e solo se  $\mathbf{P}(X = t) = 0$ .

Se in particolare  $X$  è assolutamente continua, la convergenza in legge equivale alla convergenza puntuale di  $F_{X_n}$  a  $F_X$ . In tal caso vale  $P^{X_n}(B) \rightarrow P^X(B) \quad \forall B$  del tipo  $(-\infty, s], [a, b], (a, b), [a, b), (a, b), \dots$

1. Considero  $(X_n)_{n \geq 1}$  successione di v.a.r., con  $F_{X_n}(s) = \begin{cases} 0 & \text{se } s < -\frac{1}{n} \\ \frac{1}{2} + \frac{n}{2}s & \text{se } s \in [-\frac{1}{n}, \frac{1}{n}) \\ 1 & \text{se } s \geq \frac{1}{n} \end{cases}$ , e valuto la convergenza

in legge della successione a qualche v. a..  $(F_{X_n})$  converge puntualmente a  $G(s) = \begin{cases} 0 & \text{se } s < 0 \\ \frac{1}{2} & \text{se } s = 0 \\ 1 & \text{se } s > 0 \end{cases}$ , che non

è una funzione di ripartizione perché non è continua da destra in 0. Questo fa intuire che se si considera

$F_X(s) = \begin{cases} 0 & \text{se } s < 0 \\ 1 & \text{se } s \geq 0 \end{cases}$  - che è la funzione di ripartizione di  $X \sim \delta_0$  -, allora è vero che  $\lim_{n \rightarrow +\infty} F_{X_n}(t) =$

$F_X(t) \quad \forall t \in \{t : F_X \text{ è continua in } t\} = \mathbb{R} \setminus \{0\}$ . Dunque  $X_n \rightarrow X$ .

2. Sia  $X_n \sim \mathcal{U}(\{\frac{1}{n}, \frac{2}{n}, \dots, 1\})$ . Data  $X \sim \mathcal{U}((0, 1))$ ,  $X_n \rightarrow X$ . La funzione di ripartizione di  $X_n$  è  $F_{X_n}(t) =$

$\begin{cases} 0 & \text{se } t < \frac{1}{n} \\ \frac{1}{n} & \text{se } t \in [\frac{1}{n}, \frac{2}{n}) \\ \dots & \\ \frac{n-1}{n} & \text{se } t \in [\frac{n-1}{n}, \frac{n}{n}) \\ 1 & \text{se } t \geq 1 \end{cases} : F_{X_n}(t) \text{ converge puntualmente a } F_X(t) = \begin{cases} 0 & \text{se } t < 0 \\ t & \text{se } t \in (0, 1) \\ 1 & \text{se } t \geq 1 \end{cases} \quad \forall t. \text{ Infatti, fissato}$

$t \in (0, 1)$ ,  $\forall n \in \mathbb{N} \exists k : t \in [\frac{k}{n}, \frac{k+1}{n})$ , che è equivalente  $nt \in [k, k+1)$ . Questo significa che si può scrivere

$k = [nt]$ , per cui  $t \in [\frac{[nt]}{n}, \frac{[nt]+1}{n})$ .  $\lim_{n \rightarrow +\infty} \frac{[nt]}{n} = \lim_{n \rightarrow +\infty} \frac{[nt]+1}{n} = t$ .

**Def** Data  $M$  collezione di misure di probabilità  $\mu$  su  $\mathbb{R}$ ,  $M$  si dice tight se  $\forall \varepsilon > 0$  esiste un insieme compatto

$K : \inf_{\mu \in M} \mu(K) > 1 - \varepsilon$ .

La definizione è ben posta perché gli insiemi compatti sono un sottinsieme degli insiemi chiusi, che appartengono a  $\mathcal{B}(\mathbb{R})$ ; applicheremo la definizione nel caso di  $M$  rappresentata da una successione  $(Q_n)_{n \geq 1}$ . Il significato è che,



stabilito un numero arbitrariamente vicino a 1, è possibile trovare un compatto che abbia misura maggiore secondo tutte le  $\mu$  della famiglia.

Vale il seguente teorema:

**Teo**

Hp:  $(Q_n)_{n \geq 1}$  successione tight di misure di probabilità su  $\mathbb{R}$

Ts: esiste  $n_k : Q_{n_k}$  converge debolmente

**Teo 24.7 (teorema di continuità di Levy)**

Hp:  $(X_n)_{n \geq 1}$  successione di v.a.r.,  $X : \Omega \rightarrow \mathbb{R}$  v. a. con funzioni

caratteristiche  $(\phi_{X_n})_{n \geq 1}, \phi_X$  rispettivamente

Ts: (1)  $X_n \rightarrow X \implies \lim_{n \rightarrow +\infty} \phi_{X_n}(u) = \phi_X(u) \quad \forall u \in \mathbb{R}$

(2) se  $\phi_{X_n}(u) \rightarrow \psi(u) \quad \forall u \in \mathbb{R}$  e  $\psi(u)$  è continua in  $u = 0$ , allora

$\exists Y$  v.a.r. tale che  $\phi_Y = \psi$  e  $X_n \rightarrow Y$

(1) è una conseguenza dell'unicità della legge limite e della reciproca caratterizzazione tra una legge e la sua funzione caratteristica.

(2) permette di caratterizzare il limite in legge della successione mediante la sua funzione caratteristica, pur non riconoscendola e non avendo un candidato limite.

**Corollari**

(1)  $X_n \rightarrow X \iff \lim_{n \rightarrow +\infty} \phi_{X_n}(u) = \phi_X(u) \quad \forall u \in \mathbb{R}$

(2)  $X_n$  converge in legge  $\iff \phi_{X_n}(u) \rightarrow \psi(u)$  con  $\psi$  continua in 0

Sia (1) che (2) sono una conseguenza di 24.7 (1),(2).

(1, 2) permette di controllare agevolmente, avendo la candidata  $X$ , se  $X_n \rightarrow X$  usando le funzioni caratteristiche. E' evidente l'unicità in legge del limite: la funzione caratteristica del limite ne caratterizza la legge. Se invece  $\phi_{X_n}(u) \rightarrow \psi(u)$  e  $\psi$  non è continua in 0, allora  $(X_n)_{n \geq 1}$  non ha limite in legge.

1.  $X_n \sim \mathcal{N}(\mu_n, \sigma_n^2)$  con  $\mu_n \rightarrow \mu, \sigma_n^2 \rightarrow \sigma^2 \geq 0$ .  $\phi_{X_n}(u) = e^{iu\mu_n - \frac{1}{2}\sigma_n^2 u^2}$ , che converge puntualmente a  $\phi_X(u) = e^{iu\mu - \frac{1}{2}\sigma^2 u^2}$ , con  $X \sim \mathcal{N}(\mu, \sigma^2)$ . Dunque  $X_n \rightarrow X$ .

2.  $X_n \sim \mathcal{U}((-n, n))$  ha  $\phi_{X_n}(u) = \begin{cases} 1 & \text{se } u = 0 \\ \frac{\sin(nu)}{nu} & \text{se } u \neq 0 \end{cases} : \phi_{X_n}(u) \rightarrow 0 \quad \forall u \neq 0$ , quindi  $\psi(u) = \begin{cases} 1 & \text{se } u = 0 \\ 0 & \text{se } u \neq 0 \end{cases}$ , che non è continua in 0: quindi  $X_n$  non converge in legge (intuitivamente, infatti, sarebbe un'uniforme su  $\mathbb{R}$ ).

3.  $X_n \sim \text{bin}(n, \frac{\lambda}{n})$ , con  $n \geq \lambda$ , ha  $\phi_{X_n}(u) = (1 - p + pe^{iu})^n = (1 - \frac{\lambda}{n} + \frac{\lambda}{n}e^{iu})^n = \left(1 + \frac{\lambda(e^{iu}-1)}{n}\right)^n \rightarrow e^{\lambda(e^{iu}-1)}$ , che è la  $\phi$  di  $X \sim \text{poiss}(\lambda)$ .

---

Aside: logaritmo complesso

Sia  $z \in \mathbb{C}, z \neq 0$ . Si definisce logaritmo complesso di  $z$ , e si indica con  $w = \log z$ , un numero complesso tale che  $e^w = z$ .

Posto  $w = x + iy$ , essendo  $z = |z|e^{i \arg z}$ , affinché si abbia  $e^w = e^x e^{iy} = |z|e^{i \arg z}$  è necessario e sufficiente che  $e^x = |z|$  e  $y = \arg z + 2k\pi$ , per cui  $\log z = |z| + i(\arg z + 2k\pi)$ . Questo significa che ogni numero complesso non nullo ha infiniti logaritmi: la funzione logaritmo è polidroma. Si dice logaritmo principale di  $z$ , e si indica con  $\text{Log} z$  il logaritmo che si ottiene con  $k = 0$  e  $\arg z \in [-\pi, \pi]$ . Il logaritmo complesso (principale) ha le seguenti proprietà:

$$(1) \exp(\text{Log} z) = z$$

$$(2) \text{Log}(z_1 z_2) \neq \text{Log}(z_1) + \text{Log}(z_2)$$

$$(3) \exp(\text{Log}(z_1 z_2)) = \exp(\text{Log} z_1 + \text{Log} z_2)$$

$$(4) \exp(w + v) = \exp(w) \exp(v)$$

$$(5) \text{Log}(1 + z) = z - \frac{1}{2}z^2 + \frac{1}{3}z^3 - \frac{1}{4}z^4 + o(z^4)$$

$$\forall z \in \mathbb{C} : |z| < 1; \lim_{z \rightarrow 0} \frac{\text{Log}(1 + z)}{z} = 1$$

Vale poi che se  $z_n \rightarrow z$ ,  $(1 + \frac{z_n}{n})^n \rightarrow e^z$ : infatti  $(1 + \frac{z_n}{n})^n = e^{n \log(1 + \frac{z_n}{n})} = e^{n(\frac{z_n}{n} + o(\frac{z_n}{n}))} = e^{z_n} e^{o(\frac{z_n}{n})} \rightarrow e^z$ .

**Teo 25.1 (c'è sempre uno messo peggio)**

Hp:  $(\Omega, \mathcal{A}, \mathbf{P})$  spazio di probabilità,  $(X_n)_{n \geq 1}$  successione

di v.a. r.,  $X : \Omega \rightarrow \mathbb{R}$  v.a.,  $X_n \rightarrow^{\mathbf{P}} X$

Ts:  $X_n \rightarrow X$

Quindi è la convergenza in legge a essere il tipo di convergenza più debole.

**Dim** Voglio mostrare che  $\lim_{n \rightarrow +\infty} \mathbf{E}(h(X_n)) = \mathbf{E}(h(X)) \forall h \in C_b(\mathbb{R})$ : l'idea è usare qualcosa di simile alla convergenza dominata. Per le proprietà della convergenza in probabilità,  $\forall h$  continua  $h(X_n) \rightarrow^{\mathbf{P}} h(X)$ ; se inoltre  $\exists M > 0 : |h(x)| \leq M \forall x$ , allora  $|h(X_n)| \in L^1(\Omega, \mathcal{A}, \mathbf{P}) \forall n$ ,  $h(X_n) \in L^1(\Omega, \mathcal{A}, \mathbf{P}) \forall n$  e per 23.10 (2), versione più forte della convergenza dominata,  $\lim_{n \rightarrow +\infty} \mathbf{E}(h(X_n)) = \mathbf{E}(h(X))$ . ■ !

In generale non vale l'implicazione opposta perché l'unicità del limite in legge è in legge, non è quasi certa, quindi non sarebbe ben identificata la v. a. a cui la successione converge in probabilità; tranne nel caso che segue.

**Teo 25.2** (l'implicazione si inverte se la convergenza in legge è a una costante)

Hp:  $(\Omega, \mathcal{A}, \mathbf{P})$  spazio di probabilità,  $(X_n)_{n \geq 1}$

successione di v.a. r.,  $c \in \mathbb{R}$  v.a.,  $X_n \rightarrow c$

Ts:  $X_n \rightarrow^{\mathbf{P}} c$

Occorre che le v. a. siano definite tutte sullo spazio per parlare di convergenza in probabilità.

**Dim** So che  $\lim_{n \rightarrow +\infty} \mathbf{E}(h(X_n)) = \mathbf{E}(h(c)) \forall h \in C_b(\mathbb{R})$ ; voglio mostrare che  $\lim_{n \rightarrow +\infty} \mathbf{E}\left(\frac{|X_n - c|}{1 + |X_n - c|}\right) = 0$ .

Allora considero  $\bar{h}(x) = \frac{|x - c|}{1 + |x - c|}$ : è continua perché rapporto di funzioni continue, e superiormente limitata da 1.

Quindi  $\lim_{n \rightarrow +\infty} \mathbf{E}(\bar{h}(X_n)) = \lim_{n \rightarrow +\infty} \mathbf{E}\left(\frac{|X_n - c|}{1 + |X_n - c|}\right) = \mathbf{E}(\bar{h}(c)) = 0$ . ■

**Prop 25.3** (proprietà della convergenza in legge)

Hp:  $(X_n)_{n \geq 1}, (Y_n)_{n \geq 1}$  successioni di v.a. r.,  $X : \Omega \rightarrow \mathbb{R}$ ,  $c \in \mathbb{R}$  v.a.,

$X_n \rightarrow X$ ,  $Y_n \rightarrow c$ ,  $h : \mathbb{R} \rightarrow \mathbb{R}$  è continua

Ts: (1)  $P^X$  è unica

(2)  $aX_n \rightarrow aX$

(3)  $h(X_n) \rightarrow h(X)$

(4)  $X_n + Y_n \rightarrow X + c$

(5)  $X_n Y_n \rightarrow cX$

(6)  $\frac{X_n}{Y_n} \rightarrow \frac{X}{c}$  se  $Y_n, c \neq 0$

Nelle ipotesi si suppone implicitamente che ogni  $X_n$  sia definita su un diverso spazio di probabilità. (4),(5),(6) vengono anche detti teorema di Slutsky.

**Dim** (1) è una conseguenza del teorema di Levy.

(2) Poiché  $\lim_{n \rightarrow +\infty} \mathbf{E}_n(h(X_n)) = \mathbf{E}(h(X)) \forall h \in C_b(\mathbb{R})$ , allora  $\lim_{n \rightarrow +\infty} \mathbf{E}_n(g(aX_n)) = \mathbf{E}(g(aX)) \forall g \in C_b(\mathbb{R})$ , poiché  $g$  è limitata e può inoltre essere vista come la composizione di  $g$  e della funzione continua  $ax$ , per cui è ancora continua e  $g(ax)$  può essere presa come  $h$ .

(3) So che  $\lim_{n \rightarrow +\infty} \mathbf{E}_n(f(X_n)) = \mathbf{E}(f(X)) \forall f \in C_b(\mathbb{R})$ . Voglio mostrare che  $\lim_{n \rightarrow +\infty} \mathbf{E}_n(g(h(X_n))) = \mathbf{E}(g(h(X))) \forall g \in C_b(\mathbb{R})$ . Ma se  $g$  è limitata, qualsiasi sia  $h$ ,  $g(h(x))$  è limitata; inoltre  $g(h(x))$  è continua perché composizione di funzioni continue, quindi  $g \circ h$  può essere presa come  $f$  e vale effettivamente  $\lim_{n \rightarrow +\infty} \mathbf{E}_n(g(h(X_n))) = \mathbf{E}(g(h(X)))$ .

■

Dalla dimostrazione di (3) si ricava che non occorre chiedere  $h$  limitata.

Il teorema non vale se  $Y_n \not\rightarrow c$ , come si evince dal seguente esempio.

1. Siano  $Z_1, Z_2 \sim iid\mathcal{N}(0, 1)$ . Definisco  $X_n = Z_1 \forall n$ ,  $Y_n = Z_1 \forall n$ : ovviamente  $X_n \rightarrow Z_1$ ,  $Y_n \rightarrow Z_1$ , ma è anche vero  $Y_n \rightarrow Z_2$ .  $X_n + Y_n \sim Z_1 + Z_1 = 2Z_1 \sim \mathcal{N}(0, 4)$ , mentre  $Z_1 + Z_2 \sim \mathcal{N}(0, 2)$ . Quindi non è possibile che  $X_n + Y_n$ , avendo sempre legge  $\mathcal{N}(0, 4)$ , converga in legge a una normale  $\mathcal{N}(0, 2)$ . La convergenza in legge delle componenti non è quindi abbastanza forte: occorrerebbe la convergenza in legge del vettore.

Se aggiungessi anche hp di indipendenza, varrebbe.

## 15 Teoremi limite

Un teorema che fa un'affermazione sulla convergenza della media campionaria è detto legge dei grandi numeri. Se la convergenza è quasi certa, la legge è detta forte.

**Teo 25.2 (legge forte dei grandi numeri 1)**

$$\begin{aligned} \text{Hp: } (X_n)_{n \geq 1} \text{ successione di v.a. r. su } (\Omega, \mathcal{A}, \mathbf{P}) \text{ spazio di probabilità,} \\ (X_n)_n \text{ iid, } X_n \in L^1(\Omega, \mathcal{A}, \mathbf{P}) \forall n, \mathbf{E}(X_n) = \mu \forall n, \bar{X}_n = \frac{X_1 + \dots + X_n}{n} \\ \text{Ts: } \bar{X}_n \rightarrow \mu \text{ q. c.} \end{aligned}$$

Serve che le v. a. siano definite tutte sullo stesso spazio per poter dire che sono indipendenti. Si noti che il teorema richiede delle ipotesi su  $(X_n)_{n \geq 1}$ , ma la tesi riguarda la convergenza di un'altra successione, la media campionaria; richiede delle ipotesi di integrabilità, quindi sulla legge delle  $X_n$  (che  $\mathbf{E}(|X_n|) = \int_{\mathbb{R}} |x| dP^{X_n}(x)$  esista finito), ma la tesi riguarda una proprietà quasi certa secondo  $\mathbf{P}$  dello spazio di partenza. Il significato del teorema è che la media campionaria, o media empirica, converge alla media teorica di ogni v. a. (cioè l'andamento orizzontale, delle traiettorie, tende a essere coerente con "la media che si calcola in verticale"): questo giustifica l'approccio frequentista alla probabilità. Si è già visto in un esempio passato che il risultato non può valere certamente. La tesi implica la convergenza anche in probabilità.

Vale in realtà un risultato più forte: vale anche l'implicazione inversa.

**Teo 25.2 (legge forte dei grandi numeri 2)**

$$\begin{aligned} \text{Hp: } (X_n)_{n \geq 1} \text{ successione di v.a. r. su } (\Omega, \mathcal{A}, \mathbf{P}) \text{ spazio di} \\ \text{probabilità, } (X_n)_n \text{ iid, } \mu \in \mathbb{R}, \bar{X}_n = \frac{X_1 + \dots + X_n}{n} \\ \text{Ts: } X_n \in L^1(\Omega, \mathcal{A}, \mathbf{P}) \ \forall n \text{ e } \mathbf{E}(X_n) = \mu \ \forall n \iff \bar{X}_n \rightarrow \mu \text{ q. c.} \end{aligned}$$

Si dimostra il teorema nella seguente versione:

**Teo 25.2 (legge forte dei grandi numeri 3)**

$$\begin{aligned} \text{Hp: } (X_n)_{n \geq 1} \text{ successione di v.a. r. su } (\Omega, \mathcal{A}, \mathbf{P}) \text{ spazio di probabilità, } (X_n)_n \text{ iid,} \\ X_n \in L^2(\Omega, \mathcal{A}, \mathbf{P}) \ \forall n, \mathbf{E}(X_n) = \mu \ \forall n, \text{var}(X_n) = \sigma^2, \bar{X}_n = \frac{X_1 + \dots + X_n}{n} \\ \text{Ts: } \bar{X}_n \rightarrow \mu \text{ q. c. e } \bar{X}_n \xrightarrow{L^2} \mu \end{aligned}$$

**Dim** Posso supporre  $\mu = 0$  senza perdita di generalità: per le proprietà delle convergenze basta considerare  $\tilde{X}_n = X_n - \mu$ . Questo implica  $\mathbf{E}(X_n) = 0$  e  $\text{var}(X_n) = \mathbf{E}(X_n^2)$ .

(1) Mostro che  $\bar{X}_n \xrightarrow{L^2} 0$ . Ogni elemento della successione  $\bar{X}_n$  è in  $L^2$  perché è uno spazio vettoriale,  $\mu$  ovviamente è in  $L^2$ , e  $\lim_{n \rightarrow +\infty} \mathbf{E}((\bar{X}_n - 0)^2) = \lim_{n \rightarrow +\infty} \text{var}(\bar{X}_n) = \lim_{n \rightarrow +\infty} \frac{\sigma^2}{n} = 0$ .

(2) Costruisco una sottosuccessione di  $\bar{X}_n$  che converge a 0 q. c. Prendo  $\bar{X}_{n^2} = \frac{X_1 + \dots + X_{n^2}}{n^2}$  (costituita dagli elementi  $\bar{X}_1, \bar{X}_4, \bar{X}_9, \dots$ ):  $\mathbf{E}(\bar{X}_{n^2}) = \mu = 0$ ,  $\text{var}(\bar{X}_{n^2}) = \frac{1}{n^4} n^2 \sigma^2 = \frac{\sigma^2}{n^2}$  per indipendenza. Considero  $\sum_{n=1}^{+\infty} \mathbf{E}(\bar{X}_{n^2}^2) = \sum_{n=1}^{+\infty} \frac{\sigma^2}{n^2} < +\infty$  (qui sta il senso di scegliere proprio  $\bar{X}_{n^2}$  come sottosuccessione!); ma per 9.4  $\sum_{n=1}^{+\infty} \mathbf{E}(\bar{X}_{n^2}^2) = \mathbf{E}(\sum_{n=1}^{+\infty} \bar{X}_{n^2}^2) < +\infty$ . Allora  $\sum_{n=1}^{+\infty} \bar{X}_{n^2}^2 < +\infty$  q. c., essendo  $\sum_{n=1}^{+\infty} \bar{X}_{n^2}^2$  una v. a. nonnegativa: quindi  $\lim_{n \rightarrow +\infty} \bar{X}_{n^2} = 0$  q. c. e  $\lim_{n \rightarrow +\infty} \bar{X}_n = 0$ , e ho trovato quello che cercavo.

(3) Mostro, a partire da tale sottosuccessione, che anche  $\bar{X}_n$  converge a 0 q. c.  $\forall n \geq 1$  chiamo  $p(n)$  un numero naturale tale che  $p^2(n) \leq n \leq (p(n) + 1)^2$ ; ad esempio  $p(n) = \lfloor \sqrt{n} \rfloor$ . La validità delle disuguaglianze implica peraltro  $(\sqrt{n} - 1)^2 \leq p^2(n) \leq n \iff \frac{(\sqrt{n}-1)^2}{n} \leq \frac{p^2(n)}{n} \leq 1$ , per cui  $\lim_{n \rightarrow +\infty} \frac{p^2(n)}{n} = 1$ .

Data la sottosuccessione  $\bar{X}_{n^2}$ , considero la sottosottosuccessione  $\bar{X}_{p^2(n)}$ . Poiché  $\bar{X}_{n^2} \rightarrow 0$  q. c., anche ogni sua sottosuccessione converge a 0 q. c., per cui  $\lim_{n \rightarrow +\infty} \bar{X}_{p^2(n)} = 0$  q. c.; devo mostrare che  $\lim_{n \rightarrow +\infty} \bar{X}_n = 0$  q.c., quindi mostro che  $\lim_{n \rightarrow +\infty} \left( \bar{X}_n - \frac{p^2(n)}{n} \bar{X}_{p^2(n)} \right) = 0$  q. c.  $\bar{X}_n - \frac{p^2(n)}{n} \bar{X}_{p^2(n)} = \frac{1}{n} \sum_{i=1}^n X_i - \frac{p^2(n)}{n} \frac{1}{p^2(n)} \sum_{i=1}^{p^2(n)} X_i = \frac{1}{n} \sum_{i=p^2(n)+1}^n X_i$  perché  $n \geq p^2(n)$ : vale quindi  $\mathbf{E}\left(\left(\bar{X}_n - \frac{p^2(n)}{n} \bar{X}_{p^2(n)}\right)^2\right) = \mathbf{E}\left(\left(\frac{1}{n} \sum_{i=p^2(n)+1}^n X_i\right)^2\right) = \frac{1}{n^2} \mathbf{E}\left(\left(\sum_{i=p^2(n)+1}^n X_i\right)^2\right) = \frac{1}{n^2} \text{var}\left(\sum_{i=p^2(n)+1}^n X_i\right) = \frac{1}{n^2} \sum_{i=p^2(n)+1}^n \text{var}(X_i) = \frac{1}{n^2} \sigma^2 (n - p^2(n)) \leq \frac{\sigma^2((p(n)+1)^2 - p^2(n))}{n^2} = \frac{\sigma^2(1+2p(n))}{n^2} \leq \frac{\sigma^2 3p(n)}{n^2} \leq \frac{3\sigma^2 \sqrt{n}}{n^2} = \frac{3\sigma^2}{n^{\frac{3}{2}}}$ , la cui serie converge. Allora si può concludere che  $\sum_{n=1}^{+\infty} \mathbf{E}\left(\left(\bar{X}_n - \frac{p^2(n)}{n} \bar{X}_{p^2(n)}\right)^2\right) < +\infty$ , per cui anche  $\mathbf{E}\left(\sum_{n=1}^{+\infty} \left(\bar{X}_n - \frac{p^2(n)}{n} \bar{X}_{p^2(n)}\right)^2\right) < +\infty$ , vale a dire che  $\sum_{n=1}^{+\infty} \left(\bar{X}_n - \frac{p^2(n)}{n} \bar{X}_{p^2(n)}\right)^2$

è finita q. c., per cui  $\lim_{n \rightarrow +\infty} \left( \bar{X}_n - \frac{p^2(n)}{n} \bar{X}_{p^2(n)} \right)^2 = 0$  q.c.,  $\lim_{n \rightarrow +\infty} \left( \bar{X}_n - \frac{p^2(n)}{n} \bar{X}_{p^2(n)} \right) = 0$  q. c. e  $\lim_{n \rightarrow +\infty} \bar{X}_n = \lim_{n \rightarrow +\infty} \left( \bar{X}_n - \frac{p^2(n)}{n} \bar{X}_{p^2(n)} \right) + \lim_{n \rightarrow +\infty} \frac{p^2(n)}{n} \bar{X}_{p^2(n)} = 0$  perché era già noto che  $\bar{X}_{p^2(n)} \rightarrow 0$  q. c. e  $\frac{p^2(n)}{n}$  è limitata da 1. ■

**Teo 26.1 (teorema centrale del limite)**

Hp:  $(X_n)_{n \geq 1}$  successione di v.a. r. su  $(\Omega, \mathcal{A}, \mathbf{P})$  spazio di probabilità,  $(X_n)_n$  iid,  $X_n \in L^2(\Omega, \mathcal{A}, \mathbf{P}) \forall n$ ,  $\mathbf{E}(X_n) = \mu \forall n$ ,  
 $\text{var}(X_n) = \sigma^2 > 0$ ,  $\bar{X}_n = \frac{X_1 + \dots + X_n}{n}$ ,  $T_n = \frac{\bar{X}_n - \mu}{\sqrt{\sigma^2/n}}$ ,  $Z \sim \mathcal{N}(0, 1)$   
Ts:  $T_n \rightarrow Z$

$T_n$  è la media campionaria standardizzata; vale l'unicità del limite solo in legge. Come per la LGN, la tesi riguarda la convergenza di una successione che non è quella su cui sono state fatte le ipotesi ( $(X_n)_{n \geq 1}$  converge in legge a qualsiasi v. a. che abbia la stessa legge delle  $X_n$ , dato che queste sono identicamente distribuite).

Il teorema implica che la funzione di ripartizione della media campionaria è  $\mathbf{P}(\bar{X}_n \leq x) = \mathbf{P}\left(\frac{\bar{X}_n - \mu}{\sqrt{\sigma^2/n}} \leq \frac{x - \mu}{\sqrt{\sigma^2/n}}\right) \simeq \Phi\left(\frac{x - \mu}{\sqrt{\sigma^2/n}}\right)$  per  $n \rightarrow +\infty$ . Quindi il TCL, rispetto alla LGN, dice qualcosa anche sulla legge della v. a. limite.

Il TCL implica che per  $n \rightarrow +\infty$   $\bar{X}_n$  ha legge  $\mathcal{N}(\mu, \sigma^2)$ . Inoltre  $T_n = \frac{n(\bar{X}_n - \mu)}{n\sqrt{\sigma^2/n}} = \frac{X_1 + \dots + X_n - n\mu}{\sqrt{n\sigma^2}} = \frac{S_n - n\mu}{\sqrt{n\sigma^2}}$ :  $S_n$  è la somma parziale delle  $X_n$ ,  $\mathbf{E}(S_n) = n\mu$ ,  $\text{var}(S_n) = n\sigma^2$ , quindi  $T_n$  è anche la standardizzata della somma parziale. Il TCL fornisce anche la legge asintotica delle  $S_n$ , perché la funzione di ripartizione delle  $S_n$  è  $\mathbf{P}(S_n \leq x) = \mathbf{P}\left(T_n \leq \frac{x - n\mu}{\sqrt{n\sigma^2}}\right) \simeq \Phi\left(\frac{x - n\mu}{\sqrt{n\sigma^2}}\right)$  per  $n \rightarrow +\infty$ .

Quando sono verificate le ipotesi del TCL, valgono anche quelle della LGN, per cui è anche noto che  $\bar{X}_n \rightarrow \mu$  q. c.: da questo è noto che  $\bar{X}_n - \mu \rightarrow 0$  q. c., però non è nota la velocità di convergenza. Questa è data dal TCL, che afferma che  $\frac{\bar{X}_n - \mu}{\sqrt{\sigma^2/n}} \rightarrow Z \sim \mathcal{N}(0, 1)$ , fornendo anche il confronto con  $\sqrt{\frac{\sigma^2}{n}}$ .

Se  $\sigma^2 = 0$ , ogni  $X_n$  è una v. a. costante con legge  $\delta_\mu$ , quindi converge in legge a  $\mu$ , normale di media  $\mu$  e varianza 0.

**Dim** Suppongo, senza perdita di generalità, che  $\mu = 0$ .  $\mathbf{E}(\bar{X}_n) = 0$ , quindi  $\text{var}(\bar{X}_n) = \mathbf{E}(\bar{X}_n^2)$ . Voglio usare il teorema di Levy, quindi calcolo la funzione caratteristica di  $X_n$ : poiché  $X_n \in L^2$ ,  $\phi_{X_n} \in C^2(\mathbb{R}, \mathbb{C}) \forall n$  e vale  $\phi'_{X_n}(0) = i\mathbf{E}(X_n) = 0$ ,  $\phi''_{X_n}(0) = -\mathbf{E}(X_n^2) = -\sigma^2$ . Allora, per il teorema di Taylor,  $\phi_{X_n}(v) = \phi_{X_n}(0) + \phi'_{X_n}(0)v + \frac{1}{2}\phi''_{X_n}(0)v^2 + o(v^2) = 1 - \frac{1}{2}\sigma^2 v^2 + o(v^2)$ .  $\phi_{T_n}(u) = \mathbf{E}(e^{iuT_n}) = \mathbf{E}\left(e^{iu \frac{\bar{X}_n - \mu}{\sqrt{\sigma^2/n}}}\right) = \mathbf{E}\left(e^{iu \frac{X_1 + \dots + X_n}{\sigma\sqrt{n}}}\right)$ . Questa è la funzione caratteristica di  $S_n$  calcolata in  $\frac{u}{\sigma\sqrt{n}}$ , cioè  $\phi_{X_1 + \dots + X_n}\left(\frac{u}{\sigma\sqrt{n}}\right) = \left(\phi_{X_1}\left(\frac{u}{\sigma\sqrt{n}}\right)\right)^n$  perché le  $X_n$  sono indipendenti e identicamente distribuite. Allora si ha  $\phi_{T_n}(u) = \left(\phi_{X_1}\left(\frac{u}{\sigma\sqrt{n}}\right)\right)^n = \left(1 - \frac{1}{2}\frac{u^2}{n} + o\left(\frac{u^2}{n}\right)\right)^n = \left(1 + \frac{z_n}{n}\right)^n$  con  $z_n = -\frac{1}{2}u^2 + no\left(\frac{u^2}{n}\right)$ , che, se  $\lim_{n \rightarrow +\infty} z_n = z$ , per  $n \rightarrow +\infty$  converge a  $e^z$ . Poiché  $z = -\frac{1}{2}u^2$ ,

$\phi_{T_n}(u)$  converge puntualmente a  $e^{-\frac{1}{2}u^2}$ , funzione caratteristica di una normale standard. Allora  $T_n$  converge in legge a una normale standard per il corollario al teorema di Levy. ■

## 15.1 Applicazioni

1 Sia  $S_n \sim \chi^2(n)$ .  $S_n =^D Z_1^2 + \dots + Z_n^2$ , con  $Z_1, \dots, Z_n \sim iid \mathcal{N}(0, 1)$ .  $Z_k^2 \in L^2 \forall k$  (perché  $\int_{-\infty}^{+\infty} x^2 e^{-\frac{1}{2}x^2} dx < +\infty \forall$ ), quindi posso applicare il TLC alla successione  $(Z_k^2)_{k \geq 1}$ :  $\frac{S_n - \mathbf{E}(S_n)}{\sqrt{\text{var}(S_n)}}$  converge in legge a una normale standard.  $\mathbf{E}(S_n) = n$  perché  $\mathbf{E}(Z_1^2) = 1$ ,  $\text{var}(S_n) = \text{var}(Z_1^2) + \dots + \text{var}(Z_n^2)$ .  $\text{var}(Z_i^2) = 2$  perché è la varianza di una  $\chi^2(1)$ , quindi  $\text{var}(S_n) = 2n$ . Allora  $\mathbf{P}(S_n \leq x) = \Phi\left(\frac{x-n}{\sqrt{2n}}\right)$ : in questo modo si possono calcolare facilmente i quantili della chi-quadro.

2 Sia  $V_n \sim t(n)$ . Si è già visto che  $f_{V_n}(t) \rightarrow f_Z(t) \forall t$ , dove  $Z \sim \mathcal{N}(0, 1)$ : allora, per il lemma di Scheffé,  $V_n \rightarrow Z$ . Si può dimostrare anche considerando che  $V_n =^D \frac{Z}{\sqrt{S_n/n}}$  con  $Z \sim \mathcal{N}(0, 1)$ ,  $S_n \sim \chi^2(n)$ ,  $Z \perp S_n$ . Poiché  $\frac{S_n}{n} = \bar{X}_n \rightarrow \mu = 1$  q. c. per la LGN applicata a  $(Z_n^2)_{n \geq 1}$ , che ha  $\mathbf{E}(Z_i^2) = 1$ , per le proprietà della convergenza quasi certa  $\frac{1}{\sqrt{S_n/n}} \rightarrow 1$  q. c. e  $\frac{Z}{\sqrt{S_n/n}} \rightarrow \frac{Z}{1}$  q. c. Poiché però  $Z$  è definita solo limitatamente alla legge e  $V_n = \frac{Z}{\sqrt{S_n/n}}$  solo in legge, si può dedurre che  $V_n = \frac{Z}{\sqrt{S_n/n}} \rightarrow Z$ .

3 Sia  $X_n \sim bin(n, p)$ ,  $Y_k \sim b(p)$  iid. So che  $X_n =^D Y_1 + \dots + Y_n = S_n$ : allora, applicando il TLC alle  $Y_k$  ( $Y_k \in L^2$ ,  $\mathbf{E}(Y_k) = p$ ,  $\text{var}(Y_k) = p(1-p) \forall k$ ), si ha che  $\frac{S_n - np}{\sqrt{np(1-p)}} \rightarrow Z$ , cioè  $S_n$  ha distribuzione approssimativamente  $\mathcal{N}(np, np(1-p))$  per  $n$  grandi, e  $\mathbf{P}(X_n \leq x) \simeq \Phi\left(\frac{x-np}{\sqrt{np(1-p)}}\right)$ .

## 15.2 Funzione di ripartizione empirica

Suppongo di avere una v. a.  $X$  con legge sconosciuta e di voler stimare la funzione di ripartizione della sua legge: è un tipo di inferenza molto più complesso dell'inferenza parametrica, condotto sulla base di un campione casuale estratto dalla legge di  $X$ .

**Def** Dato  $(\Omega, \mathcal{A}, \mathbf{P})$  spazio di probabilità,  $X : \Omega \rightarrow \mathbb{R}$  variabile aleatoria con legge  $P^X$ ,  $X_1, \dots, X_n \sim iid P^X$ , si dice funzione di ripartizione empirica  $F_n(t) = \frac{1}{n} \sum_{k=1}^n I_{(-\infty, t]}(X_k)$ .

La funzione di ripartizione empirica approssima  $F_X(t)$  con la frequenza relativa di variabili aleatorie del campione che stanno effettivamente in  $(-\infty, t]$ .  $F_n$  non è una funzione deterministica, ma stocastica, perché dipende dal campione casuale:  $F_n(t) = \frac{1}{n} \sum_{k=1}^n I_{(-\infty, t]}(X_k(\omega))$ .

Si noti che  $F_n(t)$  è effettivamente una funzione di ripartizione, di puro salto.

1. Sia  $n = 4$ ,  $x_1, x_2, x_3, x_4$  la realizzazione di 4 v. a. estratte dalla legge di  $X$ , e siano  $x_4 = -1, x_2 =$

$$1, x_1 = 2, x_3 = 4. \text{ Con questo campione posso costruire una realizzazione della fdr empirica: } F_n(t) = \begin{cases} 0 & \text{se } t < -1 \\ \frac{1}{4} & \text{se } t \in [-1, 1) \\ \frac{1}{2} & \text{se } t \in [1, 2) \\ \frac{3}{4} & \text{se } t \in [2, 4) \\ 1 & \text{se } t \geq 4 \end{cases}.$$

Fissato  $t$ ,  $(F_n(t))_{n \geq 1}$  è una successione di variabili aleatorie che mira a stimare  $F_X(t)$ . Definisco  $V_k = I_{(-\infty, t]}(X_k)$ :  $V_k \sim b(p)$  con  $p = \mathbf{P}(X_k \leq t) = F_{X_k}(t) = F_X(t)$  perché le v. a. sono identicamente distribuite. Allora  $F_n(t) = \frac{1}{n} \sum_{k=1}^n V_k = \bar{V}_k$ : è una media campionaria di bernoulli.  $F_n(t)$  ha le seguenti proprietà:

**P1**  $\mathbf{E}(F_n(t)) = F_X(t)$ : fissato  $t$ , il valore atteso dello stimatore coincide con il valore da stimare, quindi si dice che lo stimatore è corretto per  $F_X(t) \forall t$ .

**P2**  $\text{var}(F_n(t)) = \frac{F_X(t)(1-F_X(t))}{n}$ : fissato  $t$ ,  $\lim_{n \rightarrow +\infty} \text{var}(F_n(t)) = 0$ , quindi si dice che lo stimatore è consistente.

**P3** Per la LGN  $F_n$  converge a  $F_X(t)$  q. c. Per il TCL  $\frac{F_n(t) - F_X(t)}{\sqrt{\text{var}(F_n(t))}}$  converge in legge a una normale standard.

Tutte queste proprietà però valgono per  $t$  fissato: non riguardano la funzione globalmente. Ad esempio, la convergenza q. c. in P3 significa che l'insieme  $A = \{\omega : \lim_{n \rightarrow +\infty} F_n(t) = F_X(t)\} : \mathbf{P}(A) = 1$  dipende da  $t$ , cioè  $A = A(t)$ . E' possibile che non si riesca a trovare un insieme che sia di probabilità 1  $\forall t$ . Il seguente teorema fornisce informazioni globali sulla funzione.

### Teorema di Glivenko-Cantelli

Hp:  $(X_n)_{n \geq 1}$  successione di v.a. r. su  $(\Omega, \mathcal{A}, \mathbf{P})$  spazio di probabilità,

$$(X_n)_n \text{ iid, } F_n(t) = \frac{1}{n} \sum_{k=1}^n I_{(-\infty, t]}(X_k)$$

$$\text{Ts: } \lim_{n \rightarrow +\infty} \sup_{t \in \mathbb{R}} |F_n(t) - F_X(t)| = 0 \text{ q. c.}$$

Il teorema afferma che l'insieme degli  $\omega$  per cui  $F_n(t)$  converge uniformemente a  $F_X(t)$  è un evento di probabilità

1.



## 16 Catene di Markov

Data successione di v. a.  $(X_n)_{n \geq 1}$ , l'indice  $n$  può essere interpretato come indice temporale, per cui la successione modella l'evoluzione nel tempo di un fenomeno aleatorio. La convergenza di  $(X_n)_{n \geq 1}$  è già stata studiata nel caso in cui le  $X_n$  siano iid: ora togliamo questa ipotesi.

**Def 27.1** Dato  $(\Omega, \mathcal{A}, \mathbf{P})$  spazio di probabilità,  $(E, \mathcal{E})$  spazio misurabile e una famiglia di variabili aleatorie  $(X_t)_{t \in T}$  con  $T = \mathbb{R}^+$  o  $T = I$  intervallo,  $(X_t)_{t \in T}$  si dice processo stocastico (o aleatorio) se  $\forall t X_t : (\Omega, \mathcal{A}, \mathbf{P}) \rightarrow (E, \mathcal{E})$ . In tal caso  $E$  è detto spazio degli stati e ogni suo elemento è detto stato. Se in particolare  $T = \mathbb{N}$  e  $E$  è discreto, il processo stocastico  $(X_n)_{n \geq 0}$  è detto catena.

Il processo stocastico è una nozione generalizzata rispetto alle successioni di v. a. già viste: infatti l'indice dell'insieme  $t$  può anche essere continuo, a differenza dell'indice delle successioni. Una catena è quindi una successione di v. a. - indicizzata su  $\mathbb{N}$  - definite su uno stesso spazio a valori in uno spazio discreto.

Fissato  $\omega$ ,  $(X_t(\omega))_{t \in T}$  è una funzione da  $T$  in  $\mathbb{R}$ . In tal caso il grafico di  $(X_t(\omega))_{t \in T}$  come funzione di  $t$  è detto traiettoria della catena.

1. Suppongo di lanciare una moneta infinite volte; sia  $X_n(\omega)$  una v. a. che vale 1 se all' $n$ -esimo lancio esce testa, 0 altrimenti. Quindi  $E = \{0, 1\}$ ; uno spazio campionario plausibile è  $\Omega = \{0, 1\}^{\mathbb{N}}$ .  $(X_n)_{n \geq 1}$  è una catena, ma le  $X_n$  sono ancora iid.
2. Sia  $X_n$  il meteo in una certa località al giorno  $n$ .  $E = \{\text{sole, pioggia, nuvole}\}$ , quindi  $(X_n)_{n \geq 1}$  è una catena. In questo caso le  $X_n$  non sono indipendenti, perché il meteo in un certo giorno non è indipendente dal meteo del giorno precedente.
3. Suppongo che un giocatore con un capitale iniziale  $X_0 = 6$  punti 1 sull'uscita del 6 in un certo gioco. Sia  $X_n$  la v. a. che descrive il capitale del giocatore dopo l' $n$ -esimo lancio;  $E \subseteq \mathbb{N}$ , quindi  $(X_n)_{n \geq 1}$  è una catena. La traiettoria della catena dipende da  $X_0$  e dal meccanismo del gioco: in particolare, conoscendo  $X_n$  è possibile fare previsioni su  $X_{n+1}$ .

**Def 27.2** Dato un insieme  $E : |E| < +\infty$  e  $P \in M_{\mathbb{R}}(|E|, |E|)$ ,  $P$  si dice matrice di transizione (o stocastica) se  $p_{ij} \geq 0 \forall i, j \in E$  e  $\forall i \in E$  vale  $\sum_{j \in E} p_{ij} = 1$ . Se inoltre  $\forall j \in E \sum_{i \in E} p_{ij} = 1$ ,  $P$  si dice bistocastica.

1. Se  $P$  è stocastica, anche  $P^n$  è stocastica  $\forall n > 1$ . Infatti  $(P^2)_{ij} = \langle R_i, C_j \rangle$ , e  $\sum_{j \in E} (P^2)_{ij} = \langle R_i, C_1 + \dots + C_{|E|} \rangle = \langle R_i, (1 | \dots | 1)^T \rangle = \sum_{j \in E} p_{ij} = 1$ . Di conseguenza il prodotto di due matrici stocastiche è stocastico e  $P^n$  è stocastica.

Se  $P_i$  è stocastica  $\forall i = 1, \dots, n$ ,  $\sum_{i=1}^n P_i$  non è una matrice stocastica, ma lo è  $\frac{1}{n} \sum_{i=1}^n P_i$ .

2. Se  $P$  è stocastica ed è simmetrica, allora è anche bistocastica.

C'è una relazione biunivoca tra  $P$  e un grafo: una matrice di transizione  $P$  codifica le stesse informazioni di un grafo orientato avente  $|E|$  nodi, numerati da 1 a  $|E|$ , ciascuno collegato a se stesso e a tutti gli altri mediante archi orientati in modo che sulla freccia che collega il nodo  $i$  al nodo  $j$  è scritto il valore  $p_{ij}$ . Per ogni nodo la somma dei valori scritti su tutti gli archi uscenti dal nodo deve essere 1. Ogni arco assente è equivalente a un arco di valore 0; ogni arco senza scritte reca valore 1.  $P$  ha ovviamente un significato probabilistico.

**Def 27.3** Dato  $(\Omega, \mathcal{A}, \mathbf{P})$  spazio di probabilità,  $(E, \mathcal{E})$  spazio misurabile con  $|E| < +\infty$  e  $\mathcal{E} = 2^E$ ,  $v$  densità discreta su  $E$ ,  $P$  matrice di transizione e  $(X_n)_{n \geq 1}$  processo stocastico con  $X_n : \Omega \rightarrow E \forall n$ , si dice che  $(X_n)_{n \geq 0}$  è una catena di Markov omogenea con legge iniziale  $v$  e matrice di transizione  $P$ , e si scrive  $(X_n)_{n \geq 0} \sim CM(v, P)$ , se

$$\mathbf{M1} \quad \mathbf{P}(X_0 = i) = v_i \quad \forall i = 1, \dots, |E|$$

$$\mathbf{M2} \quad \mathbf{P}(X_{n+1} = j | X_0 = i_0, X_1 = i_1, \dots, X_n = i) = \mathbf{P}(X_{n+1} = j | X_n = i) = p_{ij} \quad \forall i_0, i_1, \dots, i, j \in E, \forall n \geq 0$$

$v$  densità discreta su  $E$  implica che  $v$  possa essere scritta come vettore in  $\mathbb{R}^{|E|}$ , avendo ordinato gli elementi di  $E$  con cardinalità finita, per cui  $v = (v_1, \dots, v_k)$ ,  $v_i \geq 0 \forall i \in E$ ,  $\sum_{i \in E} v_i = 1$ . La prima proprietà significa che  $X_0 \sim v$ , cioè la legge di  $X_0$  è discreta con densità  $v$  (la definizione implica peraltro che tutte le  $X_n$  sono discrete). La seconda, detta proprietà di Markov, significa che la previsione dello stato al tempo  $n+1$  dipende solo dallo stato al tempo  $n$  e non è influenzato da tutti gli stati precedenti: in questo sta il significato modellistico delle catene di Markov, il quale le rende adatte a descrivere fenomeni in cui "il passato non influenza il futuro".  $p_{ij}$  esprime quindi la probabilità di passare dallo stato  $i$  allo stato  $j$  (in un passo).

La proprietà di Markov vale  $\forall n$ , per cui  $\mathbf{P}(X_{n+1} = j | X_n = i) = \mathbf{P}(X_1 = j | X_0 = i) = p_{ij} \forall n$ : la proprietà per cui probabilità di transizione non dipende da  $n$  è detta omogeneità nel tempo.

$v$  e  $P$  sono sufficienti a descrivere il processo.

1. Dati  $a, b \in \mathbb{N}$ , Albus ha  $a$  galeoni, Bellatrix ne ha  $b$ . Viene lanciata una moneta: se esce testa (con probabilità  $p$ ) Albus riceve un galeone da Bellatrix, altrimenti accade il contrario. Il gioco prosegue lanciando monete finché uno dei due non esaurisce i galeoni; per semplicità si suppone che le monete siano lanciate infinite volte, ma che il processo ci interessi fino all'esaurimento dei galeoni (altrimenti il tempo di durata del gioco  $T$  sarebbe aleatorio). Sia  $X_n$  la v. a. che descrive il numero di galeoni di Albus:  $E = \{0, \dots, a+b\}$ ;  $X_0 = a$

q. c., quindi  $X_0 \sim \delta_a = v$ .  $\mathbf{P}(X_{n+1} = j|X_n = i) = 0$  se  $j \neq i+1, i-1$ ;  $\mathbf{P}(X_{n+1} = i+1|X_n = i) = p$ , mentre  $\mathbf{P}(X_{n+1} = i-1|X_n = i) = 1-p, \forall i, j \neq 0$ ; nei casi in cui il gioco finisce si ha invece  $\mathbf{P}(X_{n+1} = 0|X_n = 0) = 1 = \mathbf{P}(X_{n+1} = a+b|X_n = a+b)$ . Quindi la matrice di transizione è  $P \in M_{\mathbb{R}}(a+b+1, a+b+1)$ ,  $P =$

$$\begin{bmatrix} 1 & p & 0 & \dots & 0 \\ 1-p & 0 & p & \dots & 0 \\ 0 & 1-p & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1-p & 1 \end{bmatrix}, \text{ tridiagonale.}$$

2. Considero un'urna con  $N$  palline bianche e nere; a ogni istante pesco una pallina e la sostituisco con una pallina dell'altro colore. Sia  $X_n$  il numero di palline nere nell'urna dopo l' $n$ -esima estrazione:  $E = \{0, \dots, N\}$ .  $X_0 \sim v$ ?  $P(X_{n+1} = j|X_n = i) = 0$  se  $j \neq i+1, i-1$ ;  $\mathbf{P}(X_{n+1} = 1|X_n = 0) = 1$  perché le palline sono tutte bianche,  $\mathbf{P}(X_{n+1} = 2|X_n = 1) = \frac{N-1}{N}, \dots, \mathbf{P}(X_{n+1} = N|X_n = N-1) = \frac{1}{N}$ , mentre viceversa  $\mathbf{P}(X_{n+1} = 0|X_n = 1) = \frac{1}{N}$ ,  $\mathbf{P}(X_{n+1} = 1|X_n = 2) = \frac{2}{N}, \dots, \mathbf{P}(X_{n+1} = N-1|X_n = N) = 1$  perché le palline sono tutte nere.  $P \in M_{\mathbb{R}}(N+1, N+1)$ ,  $P =$
- $$\begin{bmatrix} 0 & 1 & 0 & \dots & 0 \\ \frac{1}{N} & 0 & \frac{N-1}{N} & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & \dots & \frac{N-1}{N} & 0 & \frac{1}{N} \\ 0 & 0 & \dots & 1 & 0 \end{bmatrix}.$$
- Lo stesso processo descrive la situazione fisica di due contenitori comunicanti, con molecole di gas all'interno, in cui a ogni istante si sposta una molecola;  $X_n$  descrive il numero di molecole nel primo recipiente.

3. Considero una successione di v.a. iid  $(X_n)_{n \geq 1}$  a valori in  $E : |E| < +\infty$ , con  $P^{X_n} = v \forall n$ ;  $X_0 \sim v$  è la distribuzione iniziale.  $(X_n)_{n \geq 1}$  è una catena di Markov perché  $\mathbf{P}(X_{n+1} = j|X_0 = i_0, X_1 = i_1, \dots, X_n = i) = \frac{\mathbf{P}(X_{n+1}=j, X_0=i_0, X_1=i_1, \dots, X_n=i)}{\mathbf{P}(X_0=i_0, X_1=i_1, \dots, X_n=i)} = \mathbf{P}(X_{n+1} = j) = \mathbf{P}(X_{n+1} = j|X_n = i)$  perché la probabilità si fattorizza per v. a. indipendenti.  $P =$
- $$\begin{bmatrix} v(1) & \dots & v(|E|) \\ \dots & \dots & \dots \\ v(1) & \dots & v(|E|) \end{bmatrix}.$$

**Teo 27.4 (legge congiunta dei primi  $n$  elementi di una catena)**

$$\text{Hp: } (X_n)_{n \geq 0} \sim CM(v, P)$$

$$\text{Ts: } \mathbf{P}(X_0 = i_0, X_1 = i_1, \dots, X_n = i_n) = v_{i_0} p_{i_0 i_1} \dots p_{i_{n-1} i_n}$$

Quindi conoscere  $v$  e  $P$  è sufficiente per fare previsioni sui primi  $n$  elementi di una catena: la legge congiunta  $P^{(X_0, \dots, X_n)}(i_0, \dots, i_n)$  dei primi  $n$  elementi è caratterizzata dalla legge iniziale e dalla matrice di transizione. Intuitivamente infatti è la probabilità di essere nello stato  $i_0$ , di passare da  $i_0$  a  $i_1$ , ecc.

**Dim** Sia  $\mathbf{P}(X_0 = i_0, X_1 = i_1, \dots, X_{n-1} = i_{n-1}) > 0$ . Per la formula di moltiplicazione  $\mathbf{P}(X_0 = i_0, X_1 = i_1, \dots, X_n = i_n) = \mathbf{P}(X_n = i_n | X_0 = i_0, \dots, X_{n-1} = i_{n-1}) \mathbf{P}(X_{n-1} = i_{n-1} | X_0 = i_0, \dots, X_{n-2} = i_{n-2}) \dots \mathbf{P}(X_1 = i_1 | X_0 = i_0) \mathbf{P}(X_0 = i_0)$ . Questo, per la proprietà di Markov, è  $\mathbf{P}(X_n = i_n | X_{n-1} = i_{n-1}) \mathbf{P}(X_{n-1} = i_{n-1} | X_{n-2} = i_{n-2}) \dots \mathbf{P}(X_1 = i_1 | X_0 = i_0) \mathbf{P}(X_0 = i_0) = p_{i_{n-1}i_n} p_{i_{n-2}i_{n-1}} \dots p_{i_0i_1} v_{i_0}$ .

Se invece  $\mathbf{P}(X_0 = i_0, \dots, X_{n-1} = i_{n-1}) = 0$ ,  $\mathbf{P}(X_0 = i_0, X_1 = i_1, \dots, X_n = i_n) = 0$ . Se  $\mathbf{P}(X_0 = i_0) = 0$ , si ha la tesi perché allora  $\mathbf{P}(X_1 = i_1 | X_0 = i_0) = 0$ . Se invece  $\mathbf{P}(X_0 = i_0) \neq 0$ ,  $\mathbf{P}(X_0 = i_0, X_1 = i_1, \dots, X_n = i_n) = \mathbf{P}(X_1 = i_1, \dots, X_n = i_n | X_0 = i_0) \mathbf{P}(X_0 = i_0) = 0$  implica  $\mathbf{P}(X_1 = i_1, \dots, X_n = i_n | X_0 = i_0) = \mathbf{P}(X_1 = i_1 | X_0 = i_0) = 0$ , cioè  $p_{i_0i_1} = 0$ , da cui la tesi. ■

Da 27.4 è evidente che se  $v$  cambia la legge congiunta cambia. Se  $v = \delta_i$ , la partenza è deterministica perché  $\mathbf{P}(X_0 = i) = 1$  e l'aleatorietà è tutta governata da  $P$  (questo accade e. g. se  $X_i$  rappresenta il capitale al tempo  $i$ ); altrimenti, sia la partenza che l'evoluzione sono casuali.

Suppongo ora, data  $(X_n)_{n \geq 0} \sim CM(v, P)$ , di osservare la catena al tempo  $n = 0$ : si verifica l'evento  $X_0 = i$ , con  $i \in E$ . E' naturale che, se si calcolano probabilità relative alla catena tenendo conto di questa informazione, non ci sia alcuna differenza rispetto alle probabilità relative a una catena  $(\tilde{X}_n)_{n \geq 0} \sim CM(\delta_i, P)$ . Si introduce una notazione atta a descrivere questa situazione.

**Def** Data  $(X_n)_{n \geq 0} \sim CM(v, P)$  con  $i \in E$  e  $v_i = \mathbf{P}(X_0 = i) > 0$ , si definisce  $\mathbf{P}_i : \mathcal{A} \rightarrow [0, 1]$ ,  $\mathbf{P}_i(A) := \mathbf{P}(A | X_0 = i) \forall A \in \mathcal{A}$ .

**Corollario 28.1 (catena di Markov condizionata allo stato iniziale)**

Hp:  $(X_n)_{n \geq 0} \sim CM(v, P)$  definita su  $(\Omega, \mathcal{A}, \mathbf{P})$ ,  $v_i > 0$

Ts:  $(X_n)_{n \geq 0} \sim CM(\delta_i, P)$  su  $(\Omega, \mathcal{A}, \mathbf{P}_i)$

La tesi non è scontata, perché in generale sostituendo  $\mathbf{P}$  con una legge qualsiasi non si ha alcuna garanzia sul fatto che si ottenga ancora una catena di Markov.

**Dim** Se si è venuti in possesso dell'informazione  $X_0 = i$ , effettivamente  $P^{X_0} = \delta_i$ . Dimostro che  $\mathbf{P}_i(X_{n+1} = j | X_0 = i_0, \dots, X_n = i_n) = \mathbf{P}_i(X_{n+1} = j | X_n = i_n) = p_{ij}$ . Infatti, poiché  $\mathbf{P}((A|B) | C) = \mathbf{P}(A | (B \cap C))$ ,  $\mathbf{P}_i(X_{n+1} = j | X_0 = i_0, \dots, X_n = i_n) = \mathbf{P}(X_{n+1} = j | X_0 = i_0, \dots, X_n = i_n, X_0 = i) = \mathbf{P}(X_{n+1} = j | X_0 = i, \dots, X_n = i_n)$ , che per la proprietà di Markov è  $\mathbf{P}(X_{n+1} = j | X_n = i_n) = p_{ij}$ : quindi  $(X_n)_{n \geq 0} \sim CM(\delta_i, P)$  anche su  $(\Omega, \mathcal{A}, \mathbf{P}_i)$ . ■

Se  $|E| = k$ , si indica la legge iniziale  $v$  con il vettore riga  $(v_1, \dots, v_k)$ . Vale  $(vP)_j = \sum_{i \in E} v_i P_{ij} = \sum_{i \in E} \mathbf{P}(X_0 = i) \mathbf{P}(X_1 = j | X_0 = i)$ , questa è la formula delle probabilità totali per  $\mathbf{P}(X_1 = j)$ , disintegrando  $(X_1 = j)$  con gli eventi  $(X_0 = i)$ , al variare di  $i \in E$ . Una formula analoga vale per  $\mathbf{P}(X_n = j)$  con  $n$  generico, come mostra 28.2 (1).

Per convenzione  $P^0 = Id$ ;  $P^1 = P$ . Vale  $P^2 \in M_{\mathbb{R}}(k, k)$ ,  $P_{ij}^2 = \sum_{k \in E} p_{ik} p_{kj}$ , ...,  $P_{ij}^n = \sum_{k_1, \dots, k_{n-1} \in E} p_{ik_1} p_{k_1 k_2} \dots p_{k_{n-1} j}$ . Si indica con  $p_{ij}^{(n)}$  il numero  $(P^{(n)})_{ij}$ , con  $v^{(n)}$  la legge della catena al tempo  $n$   $P^{X_n}$ .

**Teorema 28.2 (legge delle  $X_n$ , legge congiunta)**

Hp:  $(X_n)_{n \geq 0} \sim CM(v, P)$  definita su  $(\Omega, \mathcal{A}, \mathbf{P})$ ,

$$n, m \geq 0, n_1, \dots, n_l > 0, i_1, \dots, i_l \in E$$

$$\text{Ts: (1) } v^{(n)} = v^{(0)} P^n$$

$$(2) p_{ij}^{(n)} = \mathbf{P}(X_{n+m} = j | X_m = i) = \mathbf{P}_i(X_n = j)$$

$$(3) \mathbf{P}(X_{n_1} = i_1, \dots, X_{n_l} = i_l) = \sum_{i \in E} v_i^{(0)} p_{ii_1}^{(n_1)} p_{i_1 i_2}^{(n_2 - n_1)} \dots p_{i_{l-1} i_l}^{(n_l - n_{l-1})}$$

Il teorema è enunciato nell'ipotesi che le probabilità condizionate abbiano senso.

$v_n$  è la legge di  $X_n$ , che essendo definita su  $E$  di cardinalità finita può anche essere scritta come un vettore di  $\mathbb{R}^{|E|}$ , che è il risultato del prodotto di  $v^{(0)}$ , vettore della legge iniziale, e  $P^n$ . (1) è, per componenti,  $\mathbf{P}(X_n = j) = (v P^n)_j \forall j \in E$ .

(2) con  $m = 0$ , cioè la seconda uguaglianza della tesi, deriva da (1) vedendo la catena come  $(X_n)_{n \geq 0} \sim CM(\delta_i, P)$ , per cui  $\mathbf{P}(X_n = j | X_0 = i) = \delta_i P^n$ .

Mentre  $P$  contiene le probabilità che regolano il passaggio da un certo stato in  $X_m$  a un certo in  $X_{m+1}$ , le probabilità sul passaggio da uno stato in  $X_m$  a un altro in  $X_{m+n}$  sono contenute in  $P^n$ , detta matrice di transizione a  $n$  passi:  $(P^n)_{ij}$  è la probabilità di passare da  $i$  a  $j$  in  $n$  passi.

Il significato di (3) è calcolare la probabilità congiunta di un sottinsieme della successione (non necessariamente di elementi contigui) come probabilità di arrivare a  $i_1$  in  $n_1$  passi  $(p_{i_1}^{(n_1)})$ , a  $i_2$  in  $n_2 - n_1$  passi, ... ecc., senza conoscere il valore di  $X_0$  e quindi sommando su  $v_i^{(0)}$  al variare di  $i \in E$ .

**Dim** (1) Dimostro che  $\forall j \in E \mathbf{P}(X_n = j) = \sum_{i \in E} \mathbf{P}(X_0 = i) P_{ij}^n$ . E' noto che  $P_{ij}^n = \sum_{k_1, \dots, k_{n-1} \in E} p_{ik_1} p_{k_1 k_2} \dots p_{k_{n-1} j} = \sum_{k_1, \dots, k_{n-1} \in E} \mathbf{P}(X_1 = k_1 | X_0 = i) \mathbf{P}(X_2 = k_2 | X_1 = k_1) \dots \mathbf{P}(X_n = j | X_{n-1} = k_{n-1})$ , dato che la proprietà di Markov vale  $\forall n$ . Quindi la sommatoria iniziale è un'iterazione della formula delle probabilità totali:  $\sum_{i \in E} \mathbf{P}(X_0 = i) P_{ij}^n = \sum_{i, k_1, \dots, k_{n-1} \in E} \mathbf{P}(X_0 = i) \mathbf{P}(X_1 = k_1 | X_0 = i) \mathbf{P}(X_2 = k_2 | X_1 = k_1) \dots \mathbf{P}(X_n = j | X_{n-1} = k_{n-1}) = \sum_{k_{n-1} \in E} \dots \sum_{k_1 \in E} \sum_{i \in E} \mathbf{P}(X_0 = i) \mathbf{P}(X_1 = k_1 | X_0 = i) = \mathbf{P}(X_1 = k_1)$  ecc., quindi la sommatoria scritta coincide con  $\sum_{k_{n-1} \in E} \dots \sum_{k_1 \in E} \mathbf{P}(X_1 = k_1) = \sum_{k_{n-1} \in E} \dots \sum_{k_2 \in E} \mathbf{P}(X_2 = k_2) \dots \mathbf{P}(X_n = j | X_{n-1} = k_{n-1}) = \dots = \mathbf{P}(X_n = j)$ .

(3) In generale la strategia per calcolare una probabilità congiunta relativa a elementi di una catena di Markov è disintegrare rispetto a tutti gli elementi della catena partendo da  $X_0$ , e riscrivere tutto usando le probabilità condizionate sfruttando la proprietà di Markov. Per semplicità si dimostra nel caso particolare di  $n_1 = 1, n_2 = 3, n_3 = 6$ .

$$\begin{aligned} \mathbf{P}(X_1 = i_1, X_3 = i_3, X_6 = i_6) &= \sum_{k_0, k_2, k_4, k_5 \in E} \mathbf{P}(X_0 = k_0, X_1 = i_1, X_2 = k_2, X_3 = i_3, X_4 = k_4, X_5 = k_5, X_6 = i_6) \\ &= \sum_{k_0, k_2, k_4, k_5 \in E} v_{k_0} p_{k_0 i_1} \dots p_{k_5 i_6} = \sum_{k_0 \in E} v_{k_0} p_{k_0 i_1} \sum_{k_2 \in E} p_{i_1 k_2} p_{k_2 i_3} \sum_{k_4, k_5 \in E} p_{i_3 k_4} p_{k_4 k_5} p_{k_5 i_6} = \sum_{k_0 \in E} v_{k_0} p_{k_0 i_1} p_{i_1 i_3}^{(2)} p_{i_3 i_6}^{(3)}. \end{aligned}$$

(2)  $\mathbf{P}(X_{n+m} = j | X_m = i) = \frac{\mathbf{P}(X_{n+m}=j, X_m=i)}{\mathbf{P}(X_m=i)}$ . Il numeratore è  $\sum_{i_0 \in E} v_{i_0}^{(0)} p_{i_0 i}^{(m)} p_{ij}^{(n)}$ , mentre il denominatore è  $\sum_{i_0 \in E} v_{i_0}^{(0)} p_{i_0 i}^{(n)}$ : operando la semplificazione si ottiene  $\mathbf{P}(X_{n+m} = j | X_m = i) = p_{ij}^{(n+m)}$ . ■

$P^{n+m} = P^n P^m$ , quindi  $p_{ij}^{n+m} = \sum_{k \in E} p_{ik}^n p_{kj}^m$ . Per 28.2 (2)  $p_{ij}^{n+m} = \mathbf{P}(X_{n+m} = j | X_0 = i) = \mathbf{P}_i(X_{n+m} = j)$ , che è quindi uguale a  $\sum_{k \in E} p_{ik}^n p_{kj}^m = \sum_{k \in E} \mathbf{P}(X_n = k | X_0 = i) \mathbf{P}(X_m = j | X_n = k) = \sum_{k \in E} \mathbf{P}_i(X_n = k) \mathbf{P}(X_{n+m} = j | X_n = k)$ . Vale quindi la

### Equazione di Chapman-Kolmogorov

$$\mathbf{P}(X_{n+m} = j | X_0 = i) = \sum_{k \in E} \mathbf{P}_i(X_n = k) \mathbf{P}(X_{n+m} = j | X_n = k)$$

Il senso dell'uguaglianza è considerare tutte le possibili traiettorie con cui, avendo  $X_0 = i$ , può essere  $X_{n+m} = j$ : si sommano le probabilità di tutte le traiettorie al variare di  $k \in E$ , supponendo che al tempo  $n$  la catena sia in  $k$ .

**Dim** Per 28.2 (2)  $\mathbf{P}(X_{n+m} = j | X_0 = i) = p_{ij}^{(n+m)} = \mathbf{P}_i(X_{n+m} = j)$ .  $P^{n+m} = P^n P^m$ , quindi  $p_{ij}^{(n+m)} = \sum_{k \in E} p_{ik}^{(n)} p_{kj}^{(m)} = \sum_{k \in E} \mathbf{P}(X_n = k | X_0 = i) \mathbf{P}(X_{n+m} = j | X_n = k) = \sum_{k \in E} \mathbf{P}_i(X_n = k) \mathbf{P}(X_{n+m} = j | X_n = k)$ . ■

## 16.1 Classificazione degli stati

**Def 28.3** Data  $(X_n)_{n \geq 0} \sim CM(v, P)$  definita su  $(\Omega, \mathcal{A}, \mathbf{P})$ ,  $X_n : \Omega \rightarrow E \forall n \geq 0, j \in E$ , si dice istante di primo passaggio per  $j$ , e si indica con  $S_j$ , la variabile aleatoria  $S_j : \Omega \rightarrow \mathbb{N} \cup \{+\infty\}$ ,  $S_j(\omega) := \begin{cases} \min \{n \geq 0 : X_n(\omega) = j\} \\ +\infty \text{ se } X_n(\omega) \neq j \forall n \geq 0 \end{cases}$ ;

si dice istante di primo arrivo per  $j$ , e si indica con  $T_j$ , la variabile aleatoria  $T_j : \Omega \rightarrow \mathbb{N} \cup \{+\infty\}$ ,  $T_j(\omega) := \begin{cases} \min \{n \geq 1 : X_n(\omega) = j\} \\ +\infty \text{ se } X_n(\omega) \neq j \forall n \geq 1 \end{cases}$ .

Se  $X_0 = j$ ,  $S_j = 0$ , mentre  $T_j$  avrà sicuramente un valore diverso:  $T_j > 0$ . Se invece  $X_0 \neq j$ ,  $S_j = T_j$ . Se  $\exists n \geq 1 : X_n = j$ ,  $T_j = +\infty$ . Si può quindi scrivere  $S_j = T_j I_{(X_0 \neq j)}(\omega)$ . !

**Prop 28.4** (istante di primo arrivo e passaggio sono v. a.)

Hp:  $(X_n)_{n \geq 0} \sim CM(v, P)$  definita su  $(\Omega, \mathcal{A}, \mathbf{P})$ ,  $E$  spazio degli stati

Ts:  $S_j, T_j$  sono variabili aleatorie

**Dim** Mostro che  $T_j$  è una v. a.: allora anche  $S_j$  lo sarà, essendo  $S_j = T_j I_{(X_0 \neq j)}(\omega)$ .  $Im T_j = \mathbb{N} \cup \{+\infty\}$ , quindi è sufficiente mostrare che  $(T_j = n), (T_j = +\infty) \in \mathcal{A} \forall n$ , perché poi la controimmagine di ogni boreliano sarà la controimmagine di un sottinsieme di  $\mathbb{N} \cup \{+\infty\}$ , che può essere scritto come unione di eventi del tipo  $(T_j = n), (T_j = +\infty)$ . Ma  $(T_j = n) = (X_1 \neq j, \dots, X_{n-1} \neq j, X_n = j) = \bigcap_{i=1}^{n-1} (X_i \neq j) \cap (X_n = j) \in \mathcal{A}$  perché intersezione di eventi, dato che le  $X_n$  sono v. a.  $(T_j = +\infty) = \bigcap_{n=1}^{+\infty} (X_n \neq j) \in \mathcal{A}$  perché intersezione di eventi, dato che le  $X_n$  sono v. a. (oppure  $(T_j = +\infty) = (T_j < +\infty)^c = \left(\bigcup_{l=1}^{+\infty} (T_j = l)\right)^c \in \mathcal{A}$  per quanto detto sopra su  $(T_j = n)$ ). ■

La legge di  $S_j, T_j$  dipende solo da  $v, P$ .

1. Sia  $E = \{a, b\}$  con  $P = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ : si passa certamente dallo stato A allo stato B, per cui l'evoluzione è deterministica.  $S_a(\omega) = 0 \cdot I_{(X_0=a)}(\omega) + 1 \cdot I_{(X_0 \neq a)}(\omega)$ ,  $S_b(\omega) = 1 \cdot I_{(X_0=a)}(\omega) + 0 \cdot I_{(X_0 \neq a)}(\omega)$ ,  $T_a(\omega) = 1 \cdot I_{(X_0=b)}(\omega) + 2 \cdot I_{(X_0=a)}(\omega)$ ,  $T_b(\omega) = 1 \cdot I_{(X_0=a)}(\omega) + 2 \cdot I_{(X_0=b)}(\omega)$ . Se  $X_0 \sim v$ ,  $\mathbf{P}(S_a = 1) = \mathbf{P}(X_0 = b) = v(b)$ ;  $\mathbf{P}(S_a = 0) = \mathbf{P}(X_0 = a) = v(a)$ .  $\mathbf{E}(S_a) = v(b)$ .  $\mathbf{P}(T_a = 1) = \mathbf{P}(X_0 = b) = v(b)$ ;  $\mathbf{P}(T_a = 2) = \mathbf{P}(X_0 = a) = v(a)$ .
2. Sia  $E = \{a, b\}$  con  $P = \begin{bmatrix} 0 & 1 \\ \alpha & 1-\alpha \end{bmatrix}$  e  $v = \delta_a = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ .  $\mathbf{P}(X_0 = a) = 1$ . La legge dell'istante di primo passaggio è  $P^{S_a} = \delta_0$  con  $\mathbf{P}(S_a = 0) = \mathbf{P}_a(S_a = 0) = 1$ ,  $\mathbf{P}(S_a = 1) = 0$ ; la legge dell'istante di primo arrivo è  $P^{T_a}$  con  $\mathbf{P}(T_a = 1) = 0$ ,  $\mathbf{P}(T_a = k) = \mathbf{P}_a(T_a = k) = \alpha(1-\alpha)^{k-2}$  se  $k > 1$ .  $\mathbf{P}(T_a < +\infty) = \sum_{k=1}^{+\infty} \mathbf{P}(T_a = k) = \sum_{k=2}^{+\infty} \alpha(1-\alpha)^{k-2} = \frac{\alpha}{1-(1-\alpha)} = 1$ .  $\mathbf{E}(T_a) = \sum_{k=2}^{+\infty} k\alpha(1-\alpha)^{k-2} = \frac{\alpha}{1-\alpha} \sum_{k=2}^{+\infty} k(1-\alpha)^{k-1} = \frac{1+\alpha}{\alpha}$ .

Se quindi  $\mathbf{P}_i(S_j < +\infty) > 0$ , significa che lo stato  $j$  prima o poi viene visitato, dato che  $X_0 = i$ . Se  $i = j$   $\mathbf{P}_i(S_j < +\infty) = 1$ ; se  $i \neq j$   $\mathbf{P}_i(S_j < +\infty) = \mathbf{P}_i(T_j < +\infty)$ .

**Def** Data  $(X_n)_{n \geq 0} \sim CM(v, P)$  definita su  $(\Omega, \mathcal{A}, \mathbf{P})$ ,  $E$  spazio degli stati,  $i, j \in E$ , si dice che  $i$  conduce a  $j$ , e si scrive  $i \rightarrow j$ , se  $\mathbf{P}_i(S_j < +\infty) > 0$ !

Vale ovviamente  $i \rightarrow i$ .

**Prop 28.7 (condizioni equivalenti di conduzione)**

Hp:  $(X_n)_{n \geq 0} \sim CM(v, P)$  definita su  $(\Omega, \mathcal{A}, \mathbf{P})$ ,

$E$  spazio degli stati,  $i, j \in E, i \neq j$

Ts: (i)  $i \rightarrow j \iff \exists n \geq 0 : p_{ij}^{(n)} > 0$

(ii)  $i \rightarrow j \iff \exists j_1, \dots, j_{n-1} : p_{ij_1} p_{j_1 j_2} \dots p_{j_{n-1} j_n} > 0$

(i) è intuitivamente ovvia, dato che  $p_{ij}^{(n)} = \mathbf{P}(X_n = j | X_0 = i)$ . (ii) significa che  $i$  conduce a  $j$  se e solo se esiste un cammino, di probabilità non nulla - cioè un percorso di archi - che porta da  $i$  a  $j$ .

**Dim** (i) Se  $i \rightarrow j$ ,  $\mathbf{P}_i(S_j < +\infty) > 0$ .  $(S_j < +\infty) = \bigcup_{n=0}^{+\infty} (X_n = j)$ , ma allora, per subadditività della probabilità,  $\mathbf{P}_i\left(\bigcup_{n=0}^{+\infty} (X_n = j)\right) \leq \sum_{n=0}^{+\infty} \mathbf{P}_i(X_n = j)$ : quindi  $\sum_{n=0}^{+\infty} \mathbf{P}_i(X_n = j) > 0$ . Essendo la serie a termini nonnegativi, deve esistere  $\bar{n} : \mathbf{P}_i(X_{\bar{n}} = j) > 0$ , dunque  $p_{ij}^{(\bar{n})} > 0$ .

Se  $\exists \bar{n} : \mathbf{P}_i(X_{\bar{n}} = j) = p_{ij}^{(\bar{n})} > 0$ , allora  $\mathbf{P}_i(S_j < +\infty) > 0$  perché l'evento  $(X_{\bar{n}} = j) \subseteq (S_j < +\infty)$  (se l'istante di primo passaggio è  $j$ , allora esso è finito): per monotonia della probabilità,  $\mathbf{P}_i(S_j < +\infty) \geq \mathbf{P}_i(X_{\bar{n}} = j) > 0$ .

(ii)  $\exists n \geq 0 : p_{ij}^{(n)} > 0 \iff \exists j_1, \dots, j_{n-1} : p_{ij_1} p_{j_1 j_2} \dots p_{j_{n-1} j_n} > 0$ , dato che  $p_{ij}^{(n)} = \sum_{j_1, \dots, j_n \in E} p_{ij_1} \dots p_{j_n j}$ . ■

**Corollario 28.8 (transitività della conduzione)**

Hp:  $(X_n)_{n \geq 0} \sim CM(v, P)$  definita su  $(\Omega, \mathcal{A}, \mathbf{P})$ ,  $E$

spazio degli stati,  $i, j, k \in E$ ,  $i \rightarrow j$  e  $j \rightarrow k$

Ts:  $i \rightarrow k$

**Dim\*** E' noto che  $\exists n \geq 0 : p_{ij}^{(n)} > 0$ ,  $\exists m \geq 0 : p_{jk}^{(m)} > 0$ .  $p_{ik}^{(n+m)} = \sum_{h \in E} p_{ih}^{(n)} p_{hk}^{(m)} \geq p_{ij}^{(n)} p_{jk}^{(m)} > 0$  per ipotesi, quindi  $\exists n + m \geq 0 : p_{ik}^{(n+m)} > 0$  e  $i \rightarrow k$ . ■

1. Considero  $(X_n)_{n \geq 0}$  con  $P = \begin{bmatrix} 0 & 1 & 0 \\ 0 & \alpha & 1 - \alpha \\ 1 - \beta & 0 & \beta \end{bmatrix}$ .  $1 \rightarrow 2$  perché  $p_{12} > 0$ , cioè esiste un percorso di archi da 1 a 2; analogamente  $2 \rightarrow 3$ , quindi anche  $1 \rightarrow 3$ . In questo caso inoltre vale anche  $3 \rightarrow 1$ : da ogni stato è possibile raggiungere ogni altro stato.

2. Considero  $(X_n)_{n \geq 0}$  con  $P = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ . In questo caso  $1 \rightarrow 2, 2 \rightarrow 1$ , ma lo stato 3 è isolato.

**Def 28.9** Data  $(X_n)_{n \geq 0} \sim CM(v, P)$  definita su  $(\Omega, \mathcal{A}, \mathbf{P})$ ,  $E$  spazio degli stati,  $i, j \in E$ , si dice che  $i$  comunica con  $j$ , e si scrive  $i \leftrightarrow j$ , se  $i \rightarrow j$  e  $j \rightarrow i$ .

In tal caso la relazione di conduzione gode della proprietà di simmetria. Nel primo esempio sopra  $i \leftrightarrow j \forall i, j \in E$ . Nel secondo  $1 \leftrightarrow 2$ , ma  $1 \nleftrightarrow 3, 2 \nleftrightarrow 3$ .

**Def 29.1** Data  $(X_n)_{n \geq 0} \sim CM(\delta_i, P)$  definita su  $(\Omega, \mathcal{A}, \mathbf{P})$ ,  $E$  spazio degli stati,  $i \in E$ , si dice stato ricorrente se  $\mathbf{P}_i(T_i < +\infty) = 1$ ;  $i$  si dice transiente se  $\mathbf{P}_i(T_i < +\infty) < 1$ .



Quindi  $i$  è ricorrente se la catena parte da  $i$  e prima o poi ci ritorna quasi certamente (l'istante di primo arrivo - che di fatto è primo ritorno - è finito q. c.); in caso contrario  $i$  è transiente. Ogni stato è quindi ricorrente o transiente, dato che  $\mathbf{P}_i(T_i < +\infty) \leq 1$ .

**Def** Data  $(X_n)_{n \geq 0} \sim CM(v, P)$  definita su  $(\Omega, \mathcal{A}, \mathbf{P})$ ,  $E$  spazio degli stati,  $i \in E$ ,  $i$  si dice stato assorbente se  $p_{ii} = 1$ .

**Teo 29.2 (stati ricorrenti e transienti)**

Hp:  $(X_n)_{n \geq 0} \sim CM(\delta_i, P)$  definita su  $(\Omega, \mathcal{A}, \mathbf{P})$ ,  $E$  spazio degli stati,  $i, j \in E$

Ts: (1)  $i$  è ricorrente  $\iff \mathbf{P}_i(X_n = i \text{ per infiniti } n) = 1$

(2)  $i$  è transiente  $\iff \mathbf{P}_i(X_n = i \text{ per infiniti } n) = 0$

(3)  $j$  è transiente  $\implies \lim_{n \rightarrow +\infty} p_{ij}^{(n)} = 0$

(4)  $i$  è assorbente  $\implies i$  è ricorrente

(5)  $i$  è transiente  $\iff \exists j : i \rightarrow j \text{ e } j \nrightarrow i$

(6)  $i$  è ricorrente e  $i \rightarrow j \implies j$  è ricorrente

(7)  $j$  è transiente e  $i \rightarrow j \implies i$  è transiente

(8)  $i \leftrightarrow j \iff i, j$  sono entrambi ricorrenti o entrambi transienti

(9)  $\exists i \in E : i$  è ricorrente

Per classificare uno stato si suppone che  $X_0$  sia quello stato.

(1),(2) affermano che gli stati transienti caratterizzano il comportamento della catena solo all'inizio; poi l'evoluzione è determinata dagli stati ricorrenti. (3) sottolinea questo aspetto: più passa il tempo, più è improbabile che la catena arrivi in  $j$  se  $j$  è transiente.

Le catene di Markov hanno inoltre la proprietà particolare che  $\mathbf{P}_i(X_n = i \text{ per infiniti } n)$  possa valere solo 0 o 1: questo è dovuto alla legge zero-uno di Kolmogorov, secondo cui ogni evento nella  $\sigma$ -algebra di coda ha probabilità 0 o 1 (p. 72 ProbEss).

(4) è giustificata dal fatto che se  $i$  è assorbente la catena parte da  $i$  e rimane in  $i$  per ogni  $n$ , per cui  $T_i = 1$ .

(5) esprime il fatto che, trovandosi la catena in uno stato transiente, è possibile che ne esca senza poterci tornare.

(8) è una conseguenza di (6), (7). Non vale il viceversa in (6): se  $i \rightarrow j$  e  $j$  è ricorrente, non è vero che allora  $i$  è ricorrente; analogamente in (7).

In (5) e (9) l'implicazione da destra a sinistra vale perché  $|E| < +\infty$ .

1. Se in  $P \exists j : p_{ij} = 0 \forall i \neq j$  e  $\exists k : p_{jk} \neq 0$ , allora  $j$  è uno stato transiente, perché non è possibile arrivare in  $j$  senza partire da  $j$ :  $\exists i : j \rightarrow i$  e  $i \nrightarrow j$ .

2. Considero una catena con la matrice di transizione 
$$\begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & \alpha & 0 & 1-\alpha \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$
. Classifico gli stati. Lo stato 5

è assorbente, quindi è ricorrente (4); 4 è transiente perché conduce a 5, ma 5 non conduce a 4 (5). Poiché 4 comunica con 3, questo implica che anche 3 è transiente (8). 1 e 2 sono entrambi ricorrenti perché non sono transienti (5). Nel fatto che 4 non è ricorrente, ma 5 lo è, si vede un esempio del fatto che non vale il viceversa in (6) né in (7).

**Def 29.3** Data  $(X_n)_{n \geq 0} \sim CM(\delta_i, P)$  definita su  $(\Omega, \mathcal{A}, \mathbf{P})$ ,  $E$  spazio degli stati,  $i \in E$ , si dice periodo dello stato  $i$ , e si indica con  $d(i)$ , 
$$\begin{cases} MCD \{n \geq 1 : p_{ii}^{(n)} > 0\} \\ +\infty \text{ se } p_{ii}^{(n)} = 0 \forall n \geq 1 \end{cases} \in \mathbb{N} \cup \{+\infty\}$$
. Se tutti gli stati hanno lo stesso periodo, la catena si dice aperiodica.

$d(i) = +\infty$  significa che, una volta passata in  $i$ , la catena non ci torna più.  $d$  esprime le periodicità nell'evoluzione della catena.

**Def 29.4** Data  $(X_n)_{n \geq 0} \sim CM(\delta_i, P)$  definita su  $(\Omega, \mathcal{A}, \mathbf{P})$ ,  $E$  spazio degli stati,  $i \in E$ ,  $i$  si dice aperiodico se  $d(i) = 1$ , altrimenti si dice che ha periodo  $d(i) = d \geq 2$ .

Ogni stato tale che  $p_{ii} > 0$  ha periodo 1, quindi in particolare ogni stato assorbente è aperiodico.

1. Sia  $P = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ . Considero  $i = 1$ :  $p_{11}^{(n)} = 0$  se  $n$  è dispari,  $p_{11}^{(n)} = 1$  se  $n$  è pari, quindi  $d(1) = 2 = d(2)$ .

Infatti  $P^2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = P^n \forall n$  pari, per cui e. g.  $P^n = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \forall n$  dispari (verifica), e si applica 28.2 (2).

2. Sia  $P = \begin{bmatrix} 0 & 1 \\ 1/2 & 1/2 \end{bmatrix}$ . Considero  $i = 2$ :  $d(2) = MCD\{1, 2\} = 1$ .  $p_{11}^{(2)} = 1 \cdot \frac{1}{2} > 0$ , ma anche  $p_{11}^{(n)} > 0 \forall n > 2$ , quindi  $d(1) = MCD\{2, 3, 4, \dots\} = 1$ .

**Teo 29.5 (equivalenza tra comunicanza e medesimo periodo)**

Hp:  $(X_n)_{n \geq 0} \sim CM(v, P)$  definita su  $(\Omega, \mathcal{A}, \mathbf{P})$ ,  $E$  spazio degli stati,  $i, j \in E$

$$\text{Ts: } i \leftrightarrow j \iff d(i) = d(j)$$

**Def 29.6** Data  $(X_n)_{n \geq 0} \sim CM(v, P)$  definita su  $(\Omega, \mathcal{A}, \mathbf{P})$ ,  $E$  spazio degli stati,  $C \subseteq E$  si dice classe chiusa se  $i \in C, i \rightarrow j \implies j \in C$ .

Una classe è quindi chiusa se "non se ne può uscire".

$$1. \text{ Data la } P = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & \alpha & 0 & 1 - \alpha \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \text{ di prima, } \{1, 2\}, \{5\}, \{3, 4, 5\}, \{1, 2, 5\} \text{ sono classi chiuse.}$$

**Def 29.7** Data  $(X_n)_{n \geq 0} \sim CM(v, P)$  definita su  $(\Omega, \mathcal{A}, \mathbf{P})$ ,  $E$  spazio degli stati,  $C \subseteq E$  si dice classe irriducibile se è chiusa e  $\forall i, j \in C \ i \leftrightarrow j$ .

E' quindi irriducibile una classe chiusa in cui tutti gli stati comunicano. Poiché due stati comunicanti sono entrambi ricorrenti o entrambi transienti, tutti gli stati di una classe chiusa hanno il medesimo carattere: si vedrà che sono tutti ricorrenti.

1. Nell'esempio precedente, le classi  $\{1, 2\}, \{5\}$  sono anche irriducibili.

**Def 29.7** Data  $(X_n)_{n \geq 0} \sim CM(v, P)$  definita su  $(\Omega, \mathcal{A}, \mathbf{P})$ ,  $E$  spazio degli stati,  $(X_n)_{n \geq 0}$ , o equivalentemente  $P$ , si dice irriducibile se  $E$  è una classe irriducibile.

1. Nell'esempio precedente, se  $\mathbf{P}(X_0 \in \{1, 2\}) = 1$ , allora  $\mathbf{P}(X_n \in \{1, 2\}) = 1 \ \forall n$ , perché  $\{1, 2\}$  è una classe chiusa. In tal caso tutte le informazioni sull'evoluzione di una catena sono codificate da  $P' = P|_C = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ , che peraltro è irriducibile: la catena può essere studiata su  $E' = \{1, 2\}$ .

Se  $\mathbf{P}(X_0 \in \{3, 4, 5\}) = 1$ , allora  $\mathbf{P}(X_n \in \{3, 4, 5\}) = 1 \ \forall n$ , perché  $\{3, 4, 5\}$  è una classe chiusa. In tal caso

tutte le informazioni sull'evoluzione di una catena sono codificate da  $P' = P|_C = \begin{bmatrix} 0 & 1 & 0 \\ \alpha & 0 & 1 - \alpha \\ 0 & 0 & 1 \end{bmatrix}$ , che non è irriducibile, dato che la classe non lo è.

**Teo 29.9 (classi irriducibili)**

Hp:  $(X_n)_{n \geq 0} \sim CM(v, P)$  definita su  $(\Omega, \mathcal{A}, \mathbf{P})$ ,  $E$  spazio degli

stati,  $E_T$  è insieme degli stati transienti

Ts: (1) se  $C \subseteq E$  è una classe irriducibile, tutti gli stati sono

ricorrenti *con lo stesso periodo*

(2) se  $E$  è una classe irriducibile, tutti gli stati di  $E$  sono ricorrenti

(3)  $\exists C_1, \dots, C_k$  classi disgiunte e irriducibili di  $E$  :

$E = E_T \cup C_1 \cup \dots \cup C_k$  e la decomposizione è unica;

se  $\exists l = 1, \dots, k : X_0 \in C_l$  q. c.,  $X_n \in C_l$  q. c.  $\forall n \geq 1$ ;

se  $X_0 \in E_T$  q. c.,  $\exists \bar{n} \geq 1, \exists k : X_n \in C_k \forall n \geq \bar{n}$

$C_1 \cup \dots \cup C_k$  è l'insieme degli stati ricorrenti, indicato con  $E_R$ , ed è una classe chiusa perché unione di classi chiuse.

Con lo stesso periodo la G l'ha detto solo nel riassuntino.

**Dim** (1) Dato  $i \in C$ ,  $i$  è transiente  $\iff \exists j : i \rightarrow j$  e  $j \nrightarrow i$ . Questo è assurdo per definizione di classe irriducibile: tutti gli stati comunicano, per cui  $i$  è ricorrente  $\forall i \in E$ . Il fatto che abbiano lo stesso periodo segue da 29.5.

(2) Caso particolare di 1. ■

**Def** Data  $(X_n)_{n \geq 0} \sim CM(v, P)$  definita su  $(\Omega, \mathcal{A}, \mathbf{P})$ ,  $E$  spazio degli stati,  $C \subseteq E$  classe chiusa, si dice istante di primo passaggio nella classe (o tempo di assorbimento nella classe) e si indica con  $S_C$ ,  $\left\{ \begin{array}{l} \min \{n \geq 0 : X_n \in C\} \\ +\infty \text{ se } X_n \notin C \forall n \geq 0 \end{array} \right.$  ;  
si dice istante di primo arrivo nella classe, e si indica con  $T_C$ ,  $\left\{ \begin{array}{l} \min \{n \geq 1 : X_n \in C\} \\ +\infty \text{ se } X_n \notin C \forall n \geq 1 \end{array} \right.$ . Si dice tempo medio di assorbimento in  $C$ , e si indica con  $\tau_i^C$ ,  $\mathbf{E}_i(T_C)$ .

$S_C \leq T_C$  sempre. Se  $X_0 \in C$ ,  $S_C = 0$  e  $T_C = 1$  perché  $C$  è chiusa; se  $X_0 \notin C$ ,  $S_C = T_C$ . Si noti inoltre che vale  $\mathbf{P}_i(T_C < +\infty) = \mathbf{P}_i(S_C < +\infty)$ , perché se  $X_0 \in C$   $S_C, T_C$  sono entrambe certamente finite e si ha  $1 = 1$ , mentre se  $X_0 \notin C$   $S_C = T_C$  e quindi  $(T_C < +\infty) = (S_C < +\infty)$ .

1. Considero il gioco di Albus e Bellatrix, che hanno  $a, b$  galeoni rispettivamente:  $X_n$  è il numero di galeoni di

Albus,  $P = \begin{bmatrix} 1 & p & 0 & \dots & 0 \\ 1-p & 0 & p & \dots & 0 \\ 0 & 1-p & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1-p & 1 \end{bmatrix}$ . 0 e  $a+b$  sono stati assorbenti e quindi ricorrenti e aperiodici ( $\{0\}, \{a+b\}$  sono classi irriducibili); tutti gli altri stati sono transienti (???) di periodo 2 (direi invece: 1 e  $a+b-1$  sono transienti, gli altri sono ricorrenti).

Se il grafo è della catena  $Y_n$  con  $Y_n$  capitale di B e mi interessa la rovina di B, è naturale considerare la classe  $C = \{0\}$  e voler calcolare la probabilità che B vada in rovina  $\mathbf{P}_b(T_C < +\infty)$  e il numero medio di passi affinché ciò accada  $\mathbf{E}_i(T_C)$ . Se voglio sapere quando finisce il gioco, considero la classe  $C' = \{0, a+b\}$  (che è chiusa, ma non irriducibile) e calcolo  $T_{C'}$ .

**Def** Data  $(X_n)_{n \geq 0} \sim CM(v, P)$  definita su  $(\Omega, \mathcal{A}, \mathbf{P})$ ,  $i \in E$  spazio degli stati,  $C \subseteq E$  classe chiusa, si dice probabilità di assorbimento nella classe  $C$  partendo da  $i$ , e si indica con  $\lambda_i^C, \mathbf{P}_i(T_C < +\infty)$ .

$\lambda_i^C = \mathbf{P}_i(S_C < +\infty)$ . Ricordo che  $E = E_T \cup C_1 \cup \dots \cup C_k$ . Se  $i \in C$ ,  $\lambda_i^C = 1$  perché  $T_C = 0$ ; se  $i \in \tilde{C}$  classe chiusa disgiunta da  $C$  (quindi in particolare se  $\tilde{C}$  è irriducibile),  $T_C = +\infty$  e  $\lambda_i^C = 0$ , perché partendo da una classe chiusa non è possibile uscirne. Il valore  $\lambda_i^C$  è quindi interessante quando  $i \in E_T \setminus C =: D$ , che è una classe non irriducibile. Si noti che una classe chiusa può non essere irriducibile e contenere stati transienti.

**Prop 29.10 (la probabilità di assorbimento risolve un'equazione lineare)**

Hp:  $(X_n)_{n \geq 0} \sim CM(v, P)$  definita su  $(\Omega, \mathcal{A}, \mathbf{P})$ ,  $E$  spazio

degli stati,  $C \subseteq E$  classe chiusa,  $D = E_T \setminus C$ ,  $i \in D$

$$\text{Ts: } \lambda_i^C \text{ è tale che } \lambda_i^C = \sum_{k \in C} p_{ik} + \sum_{j \in D} p_{ij} \lambda_j^C$$

Il significato dell'equazione è controllare cosa accade al passo immediatamente successivo a  $i$ : la probabilità di essere assorbiti in  $C$  partendo da  $i$  è la probabilità di essere assorbiti in  $C$  al passo immediatamente successivo più la probabilità di andare - al passo immediatamente successivo - in uno stato  $j$  in  $E_T \setminus C$  (non si considerano tutte le altre classi irriducibili, perché in tal caso  $\lambda_j^C = 0$ ) ed essere assorbiti in  $C$  in un tempo successivo. Se si scrive l'equazione della tesi  $\forall i \in E$ , si ottiene un sistema lineare quadrato di dimensione  $|E| \times |E|$  nelle incognite  $\lambda_1^C, \dots, \lambda_{|E|}^C$ .

**Dim** E' noto che  $\lambda_i^C = \mathbf{P}_i(T_C < +\infty) = \sum_{k \in E} \mathbf{P}_i(T_C < +\infty | X_1 = k) \mathbf{P}(X_1 = k)$  per le probabilità totali, con  $\Omega$  disintegrato con i possibili valori di  $X_1$ . Quanto scritto coincide con  $\sum_{k \in C} \mathbf{P}_i(T_C < +\infty | X_1 = k) \mathbf{P}_i(X_1 = k) + \sum_{k \in E_T \setminus C} \mathbf{P}_i(T_C < +\infty | X_1 = k) \mathbf{P}_i(X_1 = k)$ . Non si somma con  $k$  nelle classi di stati ricorrenti, perché in tal caso

$\mathbf{P}_i(T_C < +\infty | X_1 = k) = 0$ , dato che  $X_0$  è uno stato transitorio non in  $C$  e  $X_1$  è uno stato ricorrente facente parte di una classe chiusa: la catena non uscirà mai da quella classe e quindi non arriverà in  $C$ . Vale  $\mathbf{P}_i(T_C < +\infty | X_1 = k) = \mathbf{P}(T_C < +\infty | X_0 = i, X_1 = k)$  (per quanto dimostrato a suo tempo sulle probabilità condizionate): ciò è uguale a  $\mathbf{P}(S_C < +\infty | X_0 = k) = \mathbf{P}(T_C < +\infty | X_0 = k) = \lambda_k^C$ , per la proprietà di Markov (non ben giustificato). Si ottiene quindi  $\sum_{k \in C} \lambda_k^C \mathbf{P}_i(X_1 = k) + \sum_{k \in E_T \setminus C} \lambda_k^C \mathbf{P}_i(X_1 = k)$ : ma se  $k \in C$ ,  $\lambda_k^C = 1$ , quindi si ha  $\sum_{k \in C} p_{ik} + \sum_{k \in E_T \setminus C} \lambda_k^C p_{ik}$ . ■

1. Si considera ancora il gioco di A e B, con  $Y_n$  che descrive il capitale di Bellatrix;  $C = \{0\}$  è la classe che descrive lo stato rovina di B. La probabilità che B vada in rovina, sapendo che  $X_0 = b$ , è  $\mathbf{P}_b(T_C < +\infty) = \lambda_b^C$ .

Per trovarlo scrivo il sistema 
$$\begin{cases} \lambda_1^C = p + (1-p)\lambda_2^C \\ \dots \\ \lambda_{a+b-1}^C = p\lambda_{a+b-2}^C \end{cases}$$
. Vale ovviamente  $\lambda_0^C = 1$ ,  $\lambda_{a+b}^C = 0$ . Se  $p = q$  vale  $\lambda_i^C = \frac{a+b-i}{a+b} \forall i = 1, \dots, a+b-1$ , quindi in particolare  $\lambda_b^C = \frac{a}{a+b}$ , che è la percentuale di capitale iniziale posseduta da A.

Se  $p \neq q$  con  $q > p$ , posto  $\alpha := \frac{p}{1-p}$

Si ricorda che  $\tau_i^C = \mathbf{E}_i(T_C)$ . Se  $i \in C = E_R$  insieme degli stati ricorrenti,  $T_C < +\infty$  q. c. e  $\mathbf{E}_i(T_C) < +\infty$ ; se  $i$  è ricorrente e  $i \notin C$ ,  $\mathbf{E}_i(T_C) = +\infty$ . Se  $i$  è transiente,  $\tau_i^C = 1$ .

**Prop 29.11 (il tempo medio di assorbimento risolve un'equazione lineare)**

Hp:  $(X_n)_{n \geq 0} \sim CM(v, P)$  definita su  $(\Omega, \mathcal{A}, \mathbf{P})$ ,

$E$  spazio degli stati,  $i \in E_T$

Ts: (1)  $\tau_i^{E_R} < +\infty$

(2)  $\tau_i^{E_R}$  è tale che  $\tau_i^{E_R} = 1 + \sum_{j \in E^T} p_{ij} \tau_j^{E_R}$

Il senso di (1) è che partendo da uno stato transiente il numero medio di passi per finire in uno stato ricorrente è finito. (2) è analoga a 29.10: o si finisce subito in uno stato ricorrente o ci si finisce partendo da uno stato transiente  $j$ .

**Def 30.1** Data  $(X_n)_{n \geq 0} \sim CM(v, P)$  definita su  $(\Omega, \mathcal{A}, \mathbf{P})$ , dato  $E$  spazio degli stati finito, una misura di probabilità  $\pi$  su  $(E, 2^E)$  si dice invariante per  $P$  se  $\pi P = \pi$ .

Si ricorda che, essendo  $E$  finito, ogni densità discreta su  $E$  può essere scritta come un vettore con  $|E|$  componenti: la definizione significa quindi che  $\pi$  è un autovettore sinistro per  $P$ , relativo all'autovalore 1 (che  $P$  ammette, dato

che la somma degli elementi su ogni riga è 1). Se  $\pi$  è invariante per  $P$  e  $v = \pi$ ,  $vP^n = \pi P^n = \pi \forall n$ , cioè se la catena ha come distribuzione iniziale una distribuzione invariante, ogni  $X_n$  ha  $\pi$  come legge:  $X_n \sim \pi \forall n \geq 0$ .

1. Se  $P$  è inoltre bistocastica, la distribuzione uniforme  $\pi = (\frac{1}{E} | \dots | \frac{1}{E})$  è invariante per  $P$ . Infatti  $(\frac{1}{E} | \dots | \frac{1}{E}) P = (\frac{1}{E} \sum_{i \in E} p_{ij} | \dots | \frac{1}{E} \sum_{i \in E} p_{ij}) = (\frac{1}{E} | \dots | \frac{1}{E})$ .
2. Se le  $X_n$  sono iid, per cui  $X_n \sim CM \left( v, \begin{bmatrix} v(1) & \dots & v(|E|) \\ \dots & \dots & \dots \\ v(1) & \dots & v(|E|) \end{bmatrix} \right)$ ,  $v$  è invariante per  $P$ .

### Teo 30.2 (Markov-Kakutani)

Hp:  $|E| < +\infty$ ,  $P \in M_{\mathbb{R}}(|E|, |E|)$  matrice stocastica

Ts:  $\exists \pi \in \mathbb{R}^{|E|} : \pi P = \pi$ , con  $\pi$  densità di probabilità

La tesi è che ogni matrice stocastica ammette almeno un autovettore sinistro relativo all'autovalore 1; in generale non vale l'unicità. In particolare, se non è unica, ce ne sono infinite: se esistono  $\pi_1, \pi_2$  invarianti per  $\pi$ , allora  $\forall \alpha \in (0, 1)$   $\alpha\pi_1 + (1 - \alpha)\pi_2$  è invariante per  $\pi$ , per la proprietà distributiva del prodotto matriciale; la combinazione dev'essere convessa affinché  $\alpha\pi_1 + (1 - \alpha)\pi_2$  sia una distribuzione di probabilità.

Il teorema ha qualche analogia con il teorema delle contrazioni: si afferma l'esistenza di un punto fisso per l'applicazione lineare  $\mathbf{x}P$ .

**Dim\*** (riempi buchi) Sia  $N = |E|$  e  $v$  una densità di probabilità qualsiasi su  $(E, 2^E)$ . Definisco per ricorrenza la successione di vettori  $v_n = \frac{1}{n} \sum_{k=0}^{n-1} vP^k \in \mathbb{R}^N$ . Ogni  $v_n$  è una densità di probabilità su  $(E, 2^E)$ , perché  $\frac{1}{n} \sum_{k=0}^{n-1} vP^k = v \sum_{k=0}^{n-1} \frac{1}{n} P^k$ , il secondo fattore è una matrice stocastica  $S$  per quanto visto all'inizio, quindi la somma degli elementi di  $v_n$  è  $\langle v, C_1(S) + \dots + C_N(S) \rangle = 1$ . L'immagine della successione è contenuta in  $S = \left\{ \mathbf{w} \in \mathbb{R}^N : w_i \geq 0 \forall i = 1, \dots, N \text{ e } \sum_{j=1}^N w_j = 1 \right\}$ , che è chiuso e limitato e quindi compatto. Allora, per definizione di compatto,  $v_n$  ammette una sottosuccessione convergente a un elemento di  $S$ , cioè  $\exists (n_k)_{k \geq 1}, \exists \pi \in \mathbb{R}^N : \lim_{k \rightarrow +\infty} v_{n_k} = \pi$ ; quindi, essendo  $P$  un'applicazione lineare e dunque continua,  $\pi - \pi P = \lim_{k \rightarrow +\infty} v_{n_k} - (\lim_{k \rightarrow +\infty} v_{n_k}) P = \lim_{k \rightarrow +\infty} v_{n_k} - (\lim_{k \rightarrow +\infty} v_{n_k} P) = \lim_{k \rightarrow +\infty} (v_{n_k} - v_{n_k} P)$ . Questa è una somma telescopica, perché  $v_{n_k} - v_{n_k} P = \frac{1}{n_k} \sum_{k=0}^{n_k-1} vP^k - \frac{1}{n_k} \sum_{k=0}^{n_k-1} vP^k P = \frac{1}{n_k} \left( \sum_{k=0}^{n_k-1} vP^k - \frac{1}{n_k} \sum_{k=0}^{n_k-1} vP^{k+1} \right) = \frac{1}{n_k} (v - vP^{n_k})$ .  
 $\lim_{k \rightarrow +\infty} \frac{1}{n_k} (v - vP^{n_k}) = 0$  perché  $P^{n_k}$  è un'applicazione limitata per la condizione di stocasticità: allora  $\pi = \pi P$ . ■

**Prop 30. 1 (relazione tra probabilità invarianti e proprietà di  $E$ )**

Hp:  $(X_n)_{n \geq 0} \sim CM(v, P)$  definita su  $(\Omega, \mathcal{A}, \mathbf{P})$ ,  $E$  spazio degli stati,

$\pi$  è una distribuzione di probabilità invariante per  $P$ ,  $j \in E$

Ts: (1) se  $j$  è transiente,  $\pi_j = 0$

(2) se  $E$  è irriducibile,  $\exists !$  distribuzione invariante per  $P$

(3) se  $E$  è irriducibile,  $\pi_i > 0 \forall i \in E$

In (1)  $\pi_j$  è la componente  $j$ -esima di  $\pi$ . Acquisirà senso quando si vedrà che  $\pi$  descrive il comportamento sul lungo periodo della catena, per cui non c'è probabilità sugli stati transienti. (3) significa che  $\pi$  è positiva su tutti gli stati, che sono tutti ricorrenti.

**Dim** (1) Se  $j$  è transiente,  $\lim_{n \rightarrow +\infty} p_{ij}^{(n)} = 0$ . Ma poiché  $\pi$  è invariante per  $P$ ,  $\pi_j = (\pi P^n)_j = \sum_{i \in E} \pi_i p_{ij}^n$ : tutti gli addendi della somma finita tendono a 0 per  $n \rightarrow +\infty$ , quindi  $\lim_{n \rightarrow +\infty} \sum_{i \in E} \pi_i p_{ij}^{(n)} = 0 = \pi_j$ .

(3) Essendo  $\pi$  una distribuzione di probabilità,  $\exists \bar{k} : \pi_{\bar{k}} > 0$ . Poiché  $E$  è irriducibile,  $\forall j \in E \bar{k} \leftrightarrow j$  e quindi  $\exists n : p_{\bar{k}j}^{(n)} > 0$ . Ma  $\pi_j = (\pi P^n)_j = \sum_{i \in E} \pi_i p_{ij}^{(n)} > \pi_{\bar{k}} p_{\bar{k}j}^{(n)} > 0$ , da cui la tesi. ■

Se invece  $E$  non è irriducibile, esistono infinite distribuzioni invarianti, che potrebbero valere 0 anche su stati ricorrenti.

Sono solo gli stati ricorrenti a determinare la distribuzione invariante.

**Def** Data  $(X_n)_{n \geq 0} \sim CM(v, P)$  definita su  $(\Omega, \mathcal{A}, \mathbf{P})$ ,  $E$  spazio degli stati finito,  $\pi$  densità di probabilità su  $(E, 2^E)$ ,  $\pi$  si dice reversibile se  $\pi_i p_{ij} = \pi_j p_{ji} \forall i, j \in E$ .

Significa che gli indici possono essere scambiati. L'equazione nella definizione è detta equazione di bilancio dettagliato. Se  $P$  è simmetrica, è necessario e sufficiente che  $\pi$  abbia tutte le componenti uguali per essere reversibile: infatti  $P$  è bistocastica.

**Prop 30.5 (una probabilità reversibile è invariante)**

Hp:  $(X_n)_{n \geq 0} \sim CM(v, P)$  definita su  $(\Omega, \mathcal{A}, \mathbf{P})$ ,  $E$  spazio degli stati,

$\pi$  è una distribuzione di probabilità reversibile

Ts:  $\pi$  è invariante per  $P$

**Dim\*** Voglio mostrare che  $(\pi P)_j = \sum_{i \in E} \pi_i p_{ij} = \pi_j$ . Ma poiché  $\pi$  è reversibile,  $\sum_{i \in E} \pi_i p_{ij} = \sum_{i \in E} \pi_j p_{ji} = \pi_j \cdot 1 = \pi_j$  perché si sta sommando su una riga. ■



1. Considero il modello di Ehrenfest con  $N = 2$  particelle; se  $X_n$  descrive il numero di particelle nel primo

contenitore,  $(X_n)_{n \geq 0} \sim CM(v, P)$  con  $P = \begin{bmatrix} 0 & 1 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 1 & 0 \end{bmatrix}$ .  $E$  è quindi irriducibile e tutti gli stati sono

ricorrenti con periodo  $d = 2$ . Come calcolo l'unica distribuzione invariante  $\pi$ ?  $\pi = (\pi_0, \pi_1, \pi_2)$  soddisfa

$$\begin{cases} \frac{1}{2}\pi_1 = \pi_0 \\ \pi_0 + \pi_2 = \pi_1 \\ \frac{1}{2}\pi_1 = \pi_2 \end{cases} \text{ ed è tale che } \pi_k \geq 0 \forall k \text{ e } \pi_0 + \pi_1 + \pi_2 = 1. \text{ Risolvendo il sistema lineare si ottiene}$$

$\pi = (\frac{1}{4}, \frac{1}{2}, \frac{1}{4})$ :  $\pi_0 = \frac{1}{2^N}$  e in generale  $\pi = bin(N, \frac{1}{2})$ .  $\pi$  è irreversibile e quindi invariante.

## 16.2 Comportamento asintotico di catene

1. Dato  $(\Omega, \mathcal{A}, \mathbf{P})$  spazio di probabilità, se  $(X_n)_{n \geq 1}$  è una successione di variabili aleatorie, con  $X_n : \Omega \rightarrow E$   $\forall n$  e  $|E| < +\infty$ , la relazioni tra i vari tipi di convergenza diventano più semplici. La convergenza quasi certa implica la convergenza in  $L^p$  perché, se  $X_n \rightarrow X$  q. c. e le  $X_n, X$  stanno tutte in  $L^p$ ,  $|X_n - X|^p$  è una v. a. semplice e  $\mathbf{E}(|X_n - X|^p) = \sum_{i \in E} x_i \mathbf{P}(|X_n - X|^p = x_i)$ .  $\lim_{n \rightarrow +\infty} \mathbf{P}(|X_n - X|^p) = 0$  perché  $|X_n - X|^p \rightarrow^{\mathbf{P}} 0$ , quindi  $X_n \rightarrow^{L^p} X$ .

Data  $(X_n)_{n \geq 0} \sim CM(v, P)$ , se  $v = \pi$  invariante per  $P$ ,  $P^{X_n} = \pi \forall n$ , quindi  $X_n \rightarrow X \sim \pi$  (la successione delle leggi è una successione costante).

1. Sia  $P = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$  con  $v = \begin{pmatrix} \alpha \\ 1 - \alpha \end{pmatrix}$ ,  $\alpha \in (0, 1)$ . Se  $n$  è pari,  $P^n = Id$ , altrimenti  $P^n = P$ : quindi

$$v^{(n)} = vP^n = \begin{cases} v \text{ se } n \text{ è pari} \\ vP \text{ se } n \text{ è dispari} \end{cases} = \begin{cases} (\alpha, 1 - \alpha) \text{ se } n \text{ è pari} \\ (1 - \alpha, \alpha) \text{ se } n \text{ è dispari} \end{cases} : \text{quindi la successione delle leggi non}$$

converge, a meno che  $\alpha = \frac{1}{2}$ , nel qual caso converge a  $\begin{pmatrix} 1/2 \\ 1/2 \end{pmatrix}$ .

### Teo 30.6

Hp:  $(X_n)_{n \geq 0} \sim CM(v, P)$  definita su  $(\Omega, \mathcal{A}, \mathbf{P})$ ,  $E$  spazio degli stati

irriducibile,  $\pi$  è l'unica distribuzione invariante per  $P$

$$\text{Ts: (1) } \pi_i = \frac{1}{\mathbf{E}_i(T_i)} \forall i \in E$$

$$(2) \frac{1}{n+1} \sum_{k=0}^n h(X_k) \rightarrow \sum_{i \in E} h(i) \pi_i \text{ q. c. } \forall h : E \rightarrow \mathbb{R} \text{ misurabile}$$

Il senso di (1) è che  $\pi_i$  rappresenta una frequenza.  $\sum_{k=0}^n h(X_k)$  è una successione di variabili aleatorie detta somma ergodica. Essendo  $E$  finito,  $h : E \rightarrow \mathbb{R}$  è sempre continua (perché  $E$  non ha punti di accumulazione) e limitata. (2) è un generalizzazione della legge dei grandi numeri: afferma che la media campionaria delle  $h(X_k)$  converge q. c. al valore atteso di  $h(X_k)$ , calcolato - secondo la regola del valore atteso, e sfruttando il fatto che  $E$  è finito, per cui  $h(X_k)$  è semplice - nello spazio  $(E, 2^E, \pi)$  come  $\mathbf{E}_\pi(h)$ : il significato è che  $\pi$  è, asintoticamente, la legge di  $h(X_k)$ .

Ci sono due casi particolari di applicazione del teorema. Se infatti  $h(i) = i \forall i \in E$ , si ottiene il

### **Teorema ergodico**

Hp:  $(X_n)_{n \geq 0} \sim CM(v, P)$  definita su  $(\Omega, \mathcal{A}, \mathbf{P})$ ,  $E$  spazio degli stati

irriducibile,  $\pi$  è l'unica distribuzione invariante per  $P$

$$\text{Ts: } \frac{1}{n+1} \sum_{k=0}^n X_k \rightarrow \sum_{i \in E} i \pi_i \text{ q. c.}$$

La tesi afferma quindi che la successione delle medie campionarie converge quasi certamente alla media  $\sum_{i \in E} i \pi_i$ : una probabilità invariante descrive il comportamento in media di una catena di Markov per  $n \rightarrow +\infty$ . Infatti, se in particolare le  $X_n$  sono iid con distribuzione iniziale  $v$ , la tesi afferma che  $\frac{1}{n+1} \sum_{k=0}^n X_k \rightarrow \sum_{i \in E} i v_i$ , che è proprio la LGN.

Se invece  $h(X_k) = I_{\{j\}}(X_k)$ , cioè si ottiene

Hp:  $(X_n)_{n \geq 0} \sim CM(v, P)$  definita su  $(\Omega, \mathcal{A}, \mathbf{P})$ ,  $E$  spazio degli stati

irriducibile,  $\pi$  è l'unica distribuzione invariante per  $P$

$$\text{Ts: } \frac{1}{n+1} \sum_{k=0}^n I_{\{j\}}(X_k) \rightarrow \pi_j \text{ q. c.}$$

$\frac{1}{n+1} \sum_{k=0}^n I_{\{j\}} X_k$ , al variare di  $n$ , descrive la frequenza relativa di elementi della catena che passano in  $j$ , e può quindi essere vista come approssimazione di  $\mathbf{P}(X_n = j)$ , la frazione di tempo che la catena passa in  $i$ .

1. Nel modello di Ehrenfest, la frazione di tempo in cui il contenitore è privo di particelle è  $\frac{1}{2^N} = \pi_0$ .

Nell'esempio sopra si è visto che anche per una catena a due stati non valeva alcun tipo di convergenza per la

catena: la ragione è che il periodo di ogni stato non è 1.

Hp:  $(X_n)_{n \geq 0} \sim CM(v, P)$  definita su  $(\Omega, \mathcal{A}, \mathbf{P})$ ,  $E$  spazio degli stati irriducibile,

$P$  aperiodica,  $\pi$  è l'unica distribuzione invariante per  $P$

$$\text{Ts: (3) } \lim_{n \rightarrow +\infty} p_{ij}^{(n)} = \pi_j \quad \forall i, j \in E \text{ se } v = \delta_i$$

$$(4) \quad \lim_{n \rightarrow +\infty} v^{(n)} = \pi \quad \forall v$$

In (3), poiché  $p_{ij}^{(n)} = \mathbf{P}_i(X_n = j)$ , si sta affermando che la legge asintotica della catena, sapendo che  $X_0 = i$ , è descritta da  $\pi$ . Il limite in (4) è inteso componente per componente. (4) è detta convergenza all'equilibrio: la legge della catena si avvicina a  $\pi$ , legge asintotica della catena, con il passare del tempo, cioè per  $n$  grande  $\mathbf{P}(X_n = j) \simeq \pi_j$ : è equivalente a dire che  $X_n \rightarrow \pi$ . Questa è un'informazione più forte di quella fornita dal teorema ergodico (come il TCL dà più informazioni della LGN).

Si può usare il teorema ergodico anche per calcolare il numero reale  $\int_{\mathbb{R}} h(x) d\mu(x) = c$ , introducendo  $(X_n)_{n \geq 0} \sim CM(v, P)$ , con  $E = \mathbb{R}$ , tale che  $\mu$  sia una distribuzione invariante per  $P$ . Allora  $\frac{1}{n+1} \sum_{k=0}^n h(X_k) \rightarrow \int_{\mathbb{R}} h(x) d\mu(x) = c$ , cioè il valore dell'integrale può essere stimato simulando l'evoluzione della catena e calcolando il lato sinistro.

1. Presi gli stati  $a, b \in E$  e  $P = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ ,  $P$  è bistocastica, quindi la distribuzione uniforme  $\pi = \begin{pmatrix} 1/2 \\ 1/2 \end{pmatrix}$  è

invariante per  $P$ , che è anche irriducibile. Allora, per il teorema ergodico,  $\frac{1}{n+1} \sum_{k=0}^n I_{\{a\}} X_k \rightarrow \frac{1}{2}$  q. c. Se  $v = \delta_a$ ,  $v^{(n)} = \begin{cases} \delta_a & \text{se } n \text{ è pari} \\ \delta_b & \text{se } n \text{ è dispari} \end{cases}$  per quanto già visto su  $b$ : in questo caso non c'è convergenza all'equilibrio, cioè non vale (4)  $\lim_{n \rightarrow +\infty} v^{(n)} = \pi = \frac{1}{2}$ , per cui  $P$  non è aperiodica (infatti ogni stato ha periodo 2)

(analogamente, per il modello di Ehrenfest, non si può applicare 4). Vale però che  $\mathbf{E}_a(T_a) = \frac{1}{\pi_1} = 2 = \mathbf{E}_b(T_b)$ .

Ma se introduco la sottocatena  $Y_n = X_{2n}$ , allora  $(Y_n)_{n \geq 0} \sim CM(v, P)$  con  $v^{(n)} = \delta_a$ .

2. Presi gli stati  $1, 2 \in E$ ,  $P = \begin{bmatrix} 0 & 1 \\ \alpha & 1-\alpha \end{bmatrix}$ , si è già calcolato che  $\mathbf{E}_a(T_a) = \frac{1}{\pi_a} = \frac{1+\alpha}{\alpha}$ ;  $\pi$  è unica perché  $E$  è irriducibile.  $\pi_a + \pi_b = 1$ , quindi si deduce che  $\pi_b = \frac{1}{\alpha+1}$  e  $\mathbf{E}_b(T_b) = \alpha + 1$ .