

Introduzione alla Probabilità

Francesco Caravenna ~ Paolo Dai Pra

Presentazione

L'obiettivo di questo libro è di fornire un'introduzione alla teoria della probabilità e alle sue applicazioni, senza fare ricorso alla teoria della misura, per studenti dei corsi di laurea delle facoltà scientifiche (in particolar modo matematica, fisica e ingegneria). La scelta degli argomenti e l'approccio adottato sono il frutto di alcuni anni di esperienza con il corso di *Probabilità e Statistica*, tenuto per il primo/secondo anno del corso di laurea (triennale) in matematica presso l'Università degli Studi di Padova.

Si è deciso di porre grande enfasi sulla probabilità *discreta*, vale a dire su spazi finiti o numerabili. La prima ragione è che, in questo contesto, sono sufficienti pochissimi strumenti analitici per presentare la teoria in modo completo e rigoroso (bastano sostanzialmente successioni e serie). Questo permette di introdurre il linguaggio e le nozioni basilari di probabilità senza complicazioni tecniche, focalizzando l'attenzione sulle difficoltà sostanziali che gli studenti incontrano nella fase iniziale dello studio di questa disciplina. La seconda ragione è che poche nozioni di probabilità discreta sono sufficienti per discutere problemi e applicazioni estremamente interessanti e non banali, alcuni tuttora oggetto di ricerca matematica. Una selezione di possibili esempi in questa direzione è presentata nel capitolo 2. Riteniamo che la trattazione di uno o più di tali problemi già nella prima parte del corso possa costituire un ottimo elemento formativo.

La trattazione degli spazi di probabilità generali (non discreti) è molto succinta e principalmente focalizzata alla discussione delle variabili casuali assolutamente continue. In questa parte, diversi risultati sono enunciati senza dimostrazione, per ovvie ragioni, ma la presentazione è sempre rigorosa. In particolare, si è scelto di dare sempre definizioni matematicamente precise, sottolineando esplicitamente le questioni tecniche che non possono essere risolte con gli strumenti a disposizione. I prerequisiti sono al livello di un corso di Analisi 1 (limiti, derivate, integrale di Riemann), con l'eccezione del paragrafo 5.5 sulle variabili casuali multidimensionali, per cui è richiesta la conoscenza dell'analisi multivariata (derivate parziali, differenziale, integrale di Riemann in più dimensioni), e che tuttavia può essere omesso.

Indice

1	Spazi di probabilità discreti	3
1.1	Generalità	3
1.2	Proprietà fondamentali	7
1.3	Il calcolo combinatorio	11
1.4	Probabilità condizionata	24
1.5	Indipendenza di eventi	28
2	Esempi rilevanti di probabilità discrete	37
2.1	Permutazioni aleatorie	37
2.2	La passeggiata aleatoria semplice	43
2.3	Il modello di Ising in meccanica statistica	62
2.4	Il modello di Hardy-Weinberg in genetica	70
2.5	Un'applicazione alla finanza: il modello di Black-Scholes discreto	73
3	Variabili casuali discrete	79
3.1	Variabili casuali e loro distribuzioni	79
3.2	Densità discreta	81
3.3	Indipendenza di variabili casuali	84
3.4	Il valor medio	87
3.5	Spazi L^p , momenti, varianza	91
3.6	Funzione generatrice dei momenti	93
3.7	Disuguaglianze	96
3.8	Covarianza e coefficiente di correlazione	98
3.9	Valor medio e indipendenza	100
3.10	Classi notevoli di variabili casuali discrete	102
4	Spazi di probabilità generali (cenni)	111
4.1	σ -algebre e misure di probabilità	111
4.2	Variabili casuali	115
4.3	Valor medio	116
5	Variabili casuali assolutamente continue	119
5.1	Richiami sull'integrale di Riemann	119
5.2	Variabili casuali scalari assolutamente continue	121
5.3	Classi notevoli di variabili casuali assolutamente continue	123
5.4	Calcoli con densità: alcuni esempi	129

5.5	Vettori casuali assolutamente continui	130
6	Teoremi limite classici	131
6.1	La legge dei grandi numeri	131
6.2	Il teorema limite centrale: enunciato e metodo dell'approssimazione normale .	134
6.3	Teorema limite centrale: dimostrazione	138
7	Applicazioni alla statistica matematica	145
7.1	Qualche nozione di Statistica Matematica	145
7.2	Stimatori di Massima Verosimiglianza	149
7.3	Stimatori per campioni Gaussiani	159
7.4	Stima intervallare	159
A	Somme infinite	161
B	Un esempio di misura finitamente additiva ma non σ-additiva	163

Notazioni

Dato un insieme Ω e due suoi sottoinsiemi $A, B \subseteq \Omega$, useremo le notazioni standard

$$\begin{aligned}A \cup B &:= \{\omega \in \Omega : \omega \in A \text{ o } \omega \in B\}, \\A \cap B &:= \{\omega \in \Omega : \omega \in A \text{ e } \omega \in B\}, \\A^c &:= \{\omega \in \Omega : \omega \notin A\}, \\A \setminus B &:= A \cap B^c, \\A \triangle B &:= (A \setminus B) \cup (B \setminus A) = (A \cup B) \setminus (A \cap B),\end{aligned}$$

dove il simbolo “:=” indica una definizione. Le definizioni di unione e intersezione si estendono in modo naturale a una famiglia arbitraria $\{A_i\}_{i \in I}$ di sottoinsiemi di Ω :

$$\begin{aligned}\bigcup_{i \in I} A_i &:= \{\omega \in \Omega : \exists i \in I \text{ tale che } \omega \in A_i\}, \\ \bigcap_{i \in I} A_i &:= \{\omega \in \Omega : \forall i \in I \text{ si ha che } \omega \in A_i\}.\end{aligned}$$

Ricordiamo le leggi di De Morgan:

$$(A \cup B)^c = A^c \cap B^c, \quad (A \cap B)^c = A^c \cup B^c,$$

e più in generale

$$\left(\bigcup_{i \in I} A_i\right)^c = \bigcap_{i \in I} A_i^c, \quad \left(\bigcap_{i \in I} A_i\right)^c = \bigcup_{i \in I} A_i^c.$$

Indicheremo con $\mathbb{N} := \{1, 2, 3, \dots\}$ i numeri naturali, zero escluso; quando vorremo includerlo, useremo la notazione $\mathbb{N}_0 := \{0, 1, 2, \dots\}$. Adotteremo le notazioni standard per i numeri interi, razionali, reali e complessi, indicati rispettivamente con \mathbb{Z} , \mathbb{Q} , \mathbb{R} e \mathbb{C} , e porremo $\mathbb{R}^+ := [0, \infty) = \{x \in \mathbb{R} : x \geq 0\}$ e $\mathbb{Q}^+ := \mathbb{Q} \cap [0, \infty)$.

Diremo che un numero $x \in \mathbb{R}$ è *positivo* se $x \geq 0$ e *strettamente positivo* se $x > 0$; analogamente, diremo che x è *negativo* se $x \leq 0$ e *strettamente negativo* se $x < 0$. Si noti che con queste convenzioni 0 è sia positivo sia negativo. La parte positiva e negativa di un numero $x \in \mathbb{R}$ sono definite rispettivamente mediante $x^+ := \max\{x, 0\}$ e $x^- := -\min\{x, 0\} = \max\{-x, 0\}$. Si noti che $x^+, x^- \geq 0$ e $x = x^+ - x^-$, mentre il valore assoluto di x è dato da $|x| = x^+ + x^-$.

Utilizzeremo gli aggettivi “crescente” e “decescente” in senso debole: una funzione $f : \mathbb{R} \rightarrow \mathbb{R}$ sarà detta *crescente* (risp. *decescente*) se per ogni $x > y$ si ha $f(x) \geq f(y)$ (risp. $f(x) \leq f(y)$). Una funzione costante è dunque sia crescente sia decrescente.

La cardinalità di un insieme A , cioè il numero dei suoi elementi, sarà denotata col simbolo $|A|$; per indicare che un insieme ha cardinalità finita scriveremo $|A| < \infty$. Un insieme A è detto numerabile se è in corrispondenza biunivoca con \mathbb{N} , cioè se esiste una applicazione $f : A \rightarrow \mathbb{N}$ iniettiva e suriettiva. Dati due insiemi A, B , il loro prodotto cartesiano $A \times B$ è definito come l'insieme di tutte le coppie (a, b) con $a \in A$ e $b \in B$. Queste definizioni verranno riprese e approfondite nel paragrafo 1.3.

Somme infinite

Data una successione di numeri reali $(x_n)_{n \in \mathbb{N}}$, è ben noto che la *somma della serie* corrispondente $\sum_{n=1}^{\infty} x_n$ è definita come il limite per $N \rightarrow \infty$ delle somme parziali $s_N := \sum_{n=1}^N x_n$, ammesso naturalmente che tale limite esista. Quando si ha a che fare con una famiglia di numeri reali $\{x_i\}_{i \in I}$ indicizzata da un insieme arbitrario I , per definire la *somma infinita* $\sum_{i \in I} x_i$ conviene procedere in modo differente.

Nel caso speciale di una famiglia a termini positivi ($x_i \geq 0$ per ogni $i \in I$) si pone

$$\sum_{i \in I} x_i := \sup_{A \subseteq I, |A| < \infty} \sum_{j \in A} x_j \in [0, +\infty],$$

dove $\sum_{j \in A} x_j$ è un'ordinaria somma finita, dal momento che $|A| < \infty$. Chiaramente si ha $\sum_{i \in I} x_i \in [0, +\infty]$ e inoltre $\sum_{i \in I} x_i = 0$ se e solo se $x_i = 0 \forall i \in I$. Se $\sum_{i \in I} x_i < +\infty$, la famiglia $\{x_i\}_{i \in I}$ è detta *sommabile*: in questo caso i termini non nulli ($x_i > 0$) sono al più un'infinità numerabile (infatti $|\{i \in I : x_i > \frac{1}{n}\}| < \infty$ per ogni $n \in \mathbb{N}$).

Una famiglia $\{x_i\}_{i \in I}$ generica (non necessariamente a termini positivi) si dice *sommabile* se $\sum_{i \in I} |x_i| < \infty$, cioè se $\sum_{i \in I} x_i^+ < \infty$ e $\sum_{i \in I} x_i^- < \infty$. In questo caso si pone

$$\sum_{i \in I} x_i := \sum_{i \in I} x_i^+ - \sum_{i \in I} x_i^-, \quad (0.1)$$

e si ha $\sum_{i \in I} x_i \in (-\infty, +\infty)$. Più in generale, se si fa l'ipotesi più debole che almeno una delle due somme $\sum_{i \in I} x_i^+, \sum_{i \in I} x_i^-$ sia finita, si può ancora definire $\sum_{i \in I} x_i$ mediante (0.1), tenendo presente che in questo caso $\sum_{i \in I} x_i \in [-\infty, +\infty]$.

Ricordiamo alcune proprietà delle somme infinite. Se $\{x_i\}_{i \in I}$ e $\{y_i\}_{i \in I}$ sono due famiglie sommabili, anche la famiglia $\{x_i + y_i\}_{i \in I}$ lo è e vale che

$$\sum_{i \in I} (x_i + y_i) = \sum_{i \in I} x_i + \sum_{i \in I} y_i.$$

Supponiamo ora di avere una famiglia $\{x_{i,j}\}_{(i,j) \in I \times J}$ i cui elementi sono indicizzati da uno spazio prodotto. Se la famiglia è a termini positivi ($x_{i,j} \geq 0$ per ogni $i \in I, j \in J$), oppure se è sommabile, vale la seguente versione del teorema di Fubini:

$$\sum_{(i,j) \in I \times J} x_{i,j} = \sum_{i \in I} \left(\sum_{j \in J} x_{i,j} \right) = \sum_{j \in J} \left(\sum_{i \in I} x_{i,j} \right).$$

La dimostrazione di queste proprietà non è difficile ma i dettagli sono piuttosto noiosi: il lettore interessato li può trovare nell'appendice A.

Capitolo 1

Spazi di probabilità discreti

1.1 Generalità

Nel corso di questo libro con la dicitura *esperimento aleatorio* indicheremo un'osservazione relativa ad un qualunque fenomeno (fisico, economico, sociale, ...) per il quale il risultato di tale osservazione non sia determinabile con certezza a priori. Il nostro obiettivo è di fornire una descrizione matematica di un esperimento aleatorio, definendo un *modello probabilistico*.

Il primo passo consiste nell'identificare un insieme Ω , detto *spazio campionario*, che contiene tutti gli esiti possibili dell'esperimento.

- Esempio 1.1.** (i) Per il lancio di un dado ordinario a sei facce, lo spazio campionario naturale è $\Omega = \{1, 2, 3, 4, 5, 6\}$
- (ii) Per la rilevazione del numero di accessi giornalieri a un sito web, scelte possibili per lo spazio campionario sono $\Omega = \mathbb{N}_0$ oppure $\Omega = \{0, 1, \dots, 10^{100}\}$.
- (iii) Per la misurazione del tempo di attesa per l'accesso ad uno sportello di un ufficio postale, una scelta naturale di spazio campionario è data da $\Omega = [0, +\infty)$.

Il secondo ingrediente di un modello probabilistico è l'assegnazione di un "grado di fiducia", o *probabilità*, ai sottoinsiemi dello spazio campionario. Con riferimento all'Esempio 1.1, si vuol dare significato ad espressioni quali "probabilità che il numero ottenuto col dado sia pari", o "probabilità che il numero di accessi al sito web sia minore di 100", o "probabilità che il tempo di attesa sia compreso tra 3 e 10 minuti".

Vedremo più avanti in alcuni casi concreti come, sulla base di considerazioni sulla natura dell'esperimento aleatorio in esame, la scelta della probabilità risulti talvolta "naturale". Molto spesso, però, non è così. In ogni caso, la probabilità scelta va sottoposta a verifica sulla base di dati sperimentali, ottenuti da ripetizioni successive dell'esperimento. Tale problema di verifica è uno degli obbiettivi principali della *Statistica*, a cui accenneremo nel capitolo 7.

Comunque essa venga assegnata, una probabilità dovrà soddisfare ad alcune proprietà naturali. Tali proprietà risultano semplici da enunciare nel caso in cui lo spazio campionario Ω sia finito o numerabile. Rimuovendo tale ipotesi, la definizione di probabilità diviene più delicata. Per questa ragione, ci concentreremo innanzitutto sul caso di spazi di probabilità finiti o numerabili, rimandando l'analisi del caso generale al capitolo 4.

Definizione 1.2. Sia Ω un insieme finito o numerabile, e indichiamo con $\mathcal{P}(\Omega)$ la famiglia dei sottoinsiemi di Ω . Una funzione $P : \mathcal{P}(\Omega) \rightarrow [0, 1]$ si dice *probabilità* se soddisfa alle seguenti proprietà:

(P1) $P(\Omega) = 1$.

(P2) (*σ -additività*) Per ogni successione $(A_n)_{n \in \mathbb{N}}$ di sottoinsiemi di Ω a due a due disgiunti, tale cioè che $A_n \cap A_m = \emptyset$ se $n \neq m$, si ha

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

La coppia (Ω, P) è detta *spazio di probabilità discreto*, Ω è chiamato *spazio campionario* e i suoi sottoinsiemi sono detti *eventi*. Diremo che le proprietà (P1) e (P2) costituiscono il *sistema di assiomi* che definisce uno spazio di probabilità discreto.

L'interpretazione di uno spazio di probabilità discreto (Ω, P) è dunque la seguente: l'insieme Ω contiene tutti i possibili esiti di un esperimento aleatorio e, per ogni sottoinsieme $A \subseteq \Omega$, il numero $P(A) \in [0, 1]$ esprime il "grado di fiducia" che si attribuisce all'eventualità che l'esito dell'esperimento sia un elemento di A . Un modo concreto di visualizzare questo "grado di fiducia" consiste nell'immaginare di ripetere l'esperimento aleatorio un numero elevato N di volte: indicando con M il numero di volte in cui l'esito ottenuto appartiene al sottoinsieme A , si ha che $P(A) \approx M/N$. Come vedremo nel capitolo 6, questa *interpretazione frequentista* della probabilità, introdotta in modo informale, riceve una giustificazione rigorosa a posteriori dalla *legge dei grandi numeri*.

Ritornando alla Definizione 1.2, la proprietà (P1) esprime il fatto che l'intero spazio campionario è un evento *certo*, ossia ha probabilità uno. La proprietà (P2) richiede una discussione più accurata. Iniziamo col dedurre due conseguenze degli assiomi (P1) e (P2).

Lemma 1.3. *Sia (Ω, P) uno spazio di probabilità discreto. Allora valgono le seguenti proprietà:*

(i) $P(\emptyset) = 0$.

(ii) *Se A_1, A_2, \dots, A_k sono eventi a due a due disgiunti (con $k \geq 2$), allora*

$$P(A_1 \cup A_2 \cup \dots \cup A_k) = \sum_{j=1}^k P(A_j). \quad (1.1)$$

Dimostrazione. (i) Sia $x = P(\emptyset) \in [0, 1]$ e si definisca $A_n = \emptyset$ per ogni $n \in \mathbb{N}$. Chiaramente $(A_n)_{n \in \mathbb{N}}$ è una successione di sottoinsiemi disgiunti di Ω . Allora, per l'assioma (P2) e il fatto che $\bigcup_{n=1}^{+\infty} A_n = \emptyset$, si ha

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

Tale identità è possibile se e solo se $x = 0$.

(ii) Prolunghiamo la famiglia di eventi disgiunti A_1, A_2, \dots, A_k ad una successione infinita di eventi a due a due disgiunti, ponendo $A_n = \emptyset$ per $n > k$. Allora, per l'assioma (P2)

$$P(A_1 \cup A_2 \cup \dots \cup A_k) = P\left(\bigcup_{j=1}^{+\infty} A_j\right) = \sum_{j=1}^{+\infty} P(A_j) = \sum_{j=1}^k P(A_j). \quad \square$$

Osservazione 1.4. Riscriviamo l'equazione (1.1) nel caso speciale $k = 2$:

$$P(A \cup B) = P(A) + P(B), \quad \forall A, B \subseteq \Omega, \quad A \cap B = \emptyset. \quad (1.2)$$

È interessante notare che la relazione (1.1) nel caso generale ($k \geq 2$) segue da (1.2) attraverso una semplice dimostrazione per induzione (esercizio!).

La proprietà (1.1) (o equivalentemente (1.2)), detta anche *additività finita*, è una condizione “naturale”, che corrisponde ad un'idea intuitiva di probabilità (si pensi all'interpretazione frequentista). È pertanto significativo domandarsi se le coppie di assiomi $\{(P1), (P2)\}$ e $\{(P1), (1.1)\}$ siano equivalenti, cioè se da ciascuna coppia si possa dedurre l'altra. La risposta è affermativa nel caso in cui Ω sia un insieme finito, dal momento che non esistono successioni infinite di eventi disgiunti e non vuoti ($\mathcal{P}(\Omega)$ ha un numero finito di elementi). Se invece Ω è infinito, gli assiomi $\{(P1), (P2)\}$ sono strettamente più forti di $\{(P1), (1.1)\}$, cioè esistono funzioni $P : \Omega \rightarrow [0, 1]$ che soddisfano (1.1) ma non (P1). Un esempio di una tale funzione per $\Omega = \mathbb{N}$, costruito usando l'assioma della scelta, è descritto nell'Appendice B (la cui lettura può essere omessa, essendo piuttosto sofisticati gli argomenti usati).

Dunque, la σ additività non è una conseguenza dell'additività finita. Benché la teoria della probabilità finitamente additiva sia sviluppata in una parte della letteratura matematica, motivata da diverse applicazioni, in questo testo considereremo soltanto la teoria della probabilità σ -additiva, che si adatta assai bene alla maggior parte delle applicazioni e che viene adottata nella grande maggioranza della letteratura. Le ragioni per cui l'assioma di σ -additività è rilevante rispetto al più debole (1.1) sono diverse, in parte non comprensibili in questa fase iniziale della presentazione della teoria. Tuttavia, una implicazione rilevante della σ -additività è già descritta nel prossimo paragrafo.

1.1.1 Probabilità e densità

Data una probabilità P su uno spazio finito o numerabile Ω , possiamo associare a P una funzione p , detta *densità*, definita su Ω a valori in $[0, 1]$, mediante la relazione

$$p(\omega) = P(\{\omega\}). \quad (1.3)$$

Si noti che p è definita su Ω , mentre P è definita su $\mathcal{P}(\Omega)$. Usando l'assioma (P2), è facile verificare che

$$P(A) = \sum_{\omega \in A} p(\omega), \quad \forall A \subseteq \Omega. \quad (1.4)$$

In particolare, prendendo $A = \Omega$ si ottiene

$$\sum_{\omega \in \Omega} p(\omega) = 1. \quad (1.5)$$

È anche possibile percorrere il cammino inverso. Supponiamo cioè che sia assegnata una funzione $p : \Omega \rightarrow [0, 1]$ per cui valga la relazione (1.5). Possiamo allora *definire* una funzione $P : \mathcal{P}(\Omega) \rightarrow [0, 1]$ mediante (1.4) ed è facile (anche se piuttosto noioso) mostrare che tale P è effettivamente una probabilità, cioè gli assiomi (P1) e (P2) sono soddisfatti. C'è dunque una corrispondenza biunivoca tra l'insieme delle probabilità P su Ω e l'insieme delle funzioni $p : \Omega \rightarrow [0, 1]$ che soddisfano (1.5): tale corrispondenza è data dalle relazioni (1.3) e (1.4).

Questo argomento mostra che in uno spazio di probabilità discreto la probabilità è determinata dal suo valore sui *singoletti*, ossia sugli eventi costituiti da un solo elemento di Ω . Per gli spazi di probabilità più generali che vedremo più avanti, quest'ultima affermazione in generale è falsa.

1.1.2 Esempi

Concludiamo questo paragrafo con alcuni esempi di spazi di probabilità discreti.

Esempio 1.5. Sia Ω un insieme *finito*. Per $A \subseteq \Omega$, definiamo

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

ove $|\cdot|$ indica il numero di elementi di un insieme. Si vede facilmente che P è una probabilità, la cui densità, grazie alla definizione (1.4), è data da $p(\omega) = 1/|\Omega|$. Lo spazio (Ω, P) così definito si dice *spazio di probabilità uniforme*. Esso è il modello probabilistico adeguato a descrivere gli esperimenti aleatori in cui tutti gli esiti si possono ritenere equiprobabili. Ad esempio: il lancio di un dado regolare, l'estrazione di un numero dalla ruota del lotto, la successione delle carte in un mazzo accuratamente mescolato. . .

Esempio 1.6. Sia Ω un insieme *finito*, e sia $H : \Omega \rightarrow \mathbb{R}$ una funzione arbitraria. Fissato un parametro reale $\beta \geq 0$, definiamo

$$p(\omega) := \frac{1}{Z(\beta)} e^{-\beta H(\omega)},$$

dove

$$Z(\beta) := \sum_{\omega \in \Omega} e^{-\beta H(\omega)}.$$

Si noti che la relazione (1.5) è verificata, di conseguenza la funzione p determina, mediante (1.4), una probabilità che, per evidenziare la dipendenza da β , indicheremo con P_β . Prendendo a prestito la terminologia della meccanica statistica, la probabilità P_β viene detta *misura di Gibbs* relativa alla *funzione Hamiltoniana* (o *energia*) H e alla *temperatura inversa* β . L'interpretazione è la seguente: gli elementi $\omega \in \Omega$ rappresentano gli stati di un sistema fisico, a cui è associata una energia $H(\omega)$; quando il sistema è in equilibrio termico alla temperatura assoluta T , ponendo $\beta = \frac{1}{k_B T}$ (dove k_B è la costante di Boltzmann) si ha che la probabilità di osservare il sistema in uno stato ω è data da $p(\omega) = P_\beta(\{\omega\})$.

Si noti che, nel caso $\beta = 0$ (temperatura infinita), $p(\cdot)$ non dipende da ω , pertanto P_0 non è altro che la probabilità uniforme su Ω . Consideriamo invece il limite $\beta \rightarrow +\infty$ di temperatura zero (assoluto). Indichiamo con $m := \min\{H(\omega) : \omega \in \Omega\}$ il minimo assoluto della Hamiltoniana, e introduciamo l'insieme (non vuoto)

$$A := \{\omega \in \Omega : H(\omega) = m\},$$

costituito dagli elementi di Ω con energia minima. Mostriamo ora che

$$\lim_{\beta \rightarrow +\infty} P_\beta(A) = 1. \quad (1.6)$$

In altre parole, nel limite $\beta \rightarrow +\infty$, P_β “si concentra” sugli elementi di minima energia. Per dimostrare (1.6) è sufficiente (perché?) mostrare che, per ogni $\omega \notin A$,

$$\lim_{\beta \rightarrow +\infty} P_\beta(\{\omega\}) = 0.$$

Si noti che

$$P_\beta(\{\omega\}) = \frac{1}{Z(\beta)} e^{-\beta H(\omega)},$$

e che

$$Z(\beta) \geq e^{-\beta m}.$$

Pertanto

$$P_\beta(\{\omega\}) \leq \frac{e^{-\beta H(\omega)}}{e^{-\beta m}} = e^{-\beta[H(\omega)-m]}. \quad (1.7)$$

Essendo $\omega \notin A$, si ha $H(\omega) > m$, e (1.6) segue immediatamente da (1.7).

Esempio 1.7. Sia $\Omega = \mathbb{N}_0 = \{0, 1, 2, \dots\}$ e poniamo

$$p(n) := e^{-\lambda} \frac{\lambda^n}{n!},$$

dove $\lambda > 0$ è un parametro reale fissato. Ricordando che $e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!}$, $\forall x \in \mathbb{R}$, la relazione (1.5) è verificata e dunque è possibile definire una probabilità P tramite (1.4). Come vedremo in seguito, tale probabilità è particolarmente utile nella descrizione delle *file di attesa*.

1.2 Proprietà fondamentali

Iniziamo con l'espone alcune conseguenze quasi immediate degli assiomi (P1) e (P2). Qui e nel seguito, indichiamo con (Ω, P) un generico spazio di probabilità discreto.

Proposizione 1.8. *Siano $A, B \subseteq \Omega$. Allora valgono le seguenti proprietà:*

(i)

$$P(A^c) = 1 - P(A).$$

(ii) *Se $A \subseteq B$ allora*

$$P(B \setminus A) = P(B) - P(A).$$

In particolare

$$P(A) \leq P(B).$$

(iii)

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

In particolare

$$P(A \cup B) \leq P(A) + P(B).$$

Dimostrazione. (i) Per la proprietà di additività si ha

$$1 = P(\Omega) = P(A \cup A^c) = P(A) + P(A^c),$$

da cui la conclusione è immediata.

(ii) Basta osservare che, di nuovo per l'additività,

$$P(B) = P[A \cup (B \setminus A)] = P(A) + P(B \setminus A).$$

(iii) Scriviamo

$$A \cup B = [A \setminus (A \cap B)] \cup [B \setminus (A \cap B)] \cup (A \cap B).$$

I tre eventi nella precedente unione sono disgiunti. Dunque, usando l'additività e la relazione (ii), si ottiene

$$\begin{aligned} P(A \cup B) &= P[A \setminus (A \cap B)] + P[B \setminus (A \cap B)] + P(A \cap B) \\ &= P(A) - P(A \cap B) + P(B) - P(A \cap B) + P(A \cap B) \\ &= P(A) + P(B) - P(A \cap B). \end{aligned} \quad \square$$

L'identità della parte (iii) della Proposizione 1.8 può essere generalizzata all'unione di più di due eventi. Ad esempio, supponiamo di voler calcolare $P(A \cup B \cup C)$ per tre eventi A, B, C . Usando due volte l'identità appena citata

$$\begin{aligned} P(A \cup B \cup C) &= P((A \cup B) \cup C) = P(A \cup B) + P(C) - P((A \cup B) \cap C) \\ &= P(A) + P(B) - P(A \cap B) + P(C) - P((A \cap C) \cup (B \cap C)) \\ &= P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C). \end{aligned}$$

Non è difficile, a questo punto, "indovinare" la formula generale per l'unione di un numero finito arbitrario di eventi. Il seguente risultato è chiamato *formula di inclusione-esclusione*.

Proposizione 1.9. *Si considerino n eventi A_1, A_2, \dots, A_n di Ω . Allora*

$$P(A_1 \cup A_2 \cup \dots \cup A_n) = \sum_{k=1}^n \sum_{\substack{J \subseteq \{1, 2, \dots, n\} \\ \text{tali che } |J|=k}} (-1)^{k+1} P\left(\bigcap_{i \in J} A_i\right). \quad (1.8)$$

Dimostrazione. Dimostriamo per induzione su n che la relazione (1.8) è vera per ogni n -pla di eventi A_1, A_2, \dots, A_n . Per $n = 1$ la formula (1.8) si riduce a $P(A_1) = P(A_1)$, e dunque non c'è nulla da dimostrare. Supponiamo allora che l'asserto sia vero per ogni $k \leq n$, e mostriamo che è vero per $n + 1$. Siano $A_1, A_2, \dots, A_n, A_{n+1}$ eventi. Usando il fatto che, per ipotesi induttiva, (1.8) vale per $n = 2$ otteniamo

$$\begin{aligned} &P(A_1 \cup A_2 \cup \dots \cup A_n \cup A_{n+1}) \\ &= P(A_1 \cup A_2 \cup \dots \cup A_n) + P(A_{n+1}) - P((A_1 \cup A_2 \cup \dots \cup A_n) \cap A_{n+1}) \quad (1.9) \\ &= P(A_1 \cup A_2 \cup \dots \cup A_n) + P(A_{n+1}) - P(B_1 \cup B_2 \cup \dots \cup B_n), \end{aligned}$$

dove abbiamo posto per comodità $B_i = A_i \cap A_{n+1}$, per $i = 1, 2, \dots, n$. Usando nuovamente l'ipotesi induttiva, questa volta per n eventi, otteniamo

$$\begin{aligned} P(A_1 \cup A_2 \cup \dots \cup A_n) &= \sum_{k=1}^n \sum_{\substack{J \subseteq \{1,2,\dots,n\} \\ \text{tale che } |J|=k}} (-1)^{k+1} P\left(\bigcap_{i \in J} A_i\right) \\ &= \sum_{k=1}^n \sum_{\substack{J \subseteq \{1,2,\dots,n+1\} \\ \text{tale che } |J|=k \text{ e } n+1 \notin J}} (-1)^{k+1} P\left(\bigcap_{i \in J} A_i\right), \end{aligned} \quad (1.10)$$

e analogamente

$$\begin{aligned} P(B_1 \cup B_2 \cup \dots \cup B_n) &= \sum_{k=1}^n \sum_{\substack{J \subseteq \{1,2,\dots,n\} \\ \text{tale che } |J|=k}} (-1)^{k+1} P\left(\bigcap_{i \in J} B_i\right) \\ &= \sum_{k=1}^n \sum_{\substack{J \subseteq \{1,2,\dots,n\} \\ \text{tale che } |J|=k}} (-1)^{k+1} P\left(A_{n+1} \cap \left(\bigcap_{i \in J} A_i\right)\right) \\ &= - \sum_{k=2}^{n+1} \sum_{\substack{J \subseteq \{1,2,\dots,n+1\} \\ \text{tale che } |J|=k \text{ e } n+1 \in J}} (-1)^{k+1} P\left(\bigcap_{i \in J} A_i\right). \end{aligned} \quad (1.11)$$

Sostituendo (1.10) e (1.11) nell'ultimo membro di (1.9), si ottiene

$$P(A_1 \cup A_2 \cup \dots \cup A_n \cup A_{n+1}) = \sum_{k=1}^{n+1} \sum_{\substack{J \subseteq \{1,2,\dots,n+1\} \\ \text{tale che } |J|=k}} (-1)^{k+1} P\left(\bigcap_{i \in J} A_i\right),$$

che è quanto si voleva dimostrare. \square

Va notato come le dimostrazioni dei risultati delle Proposizioni 1.8 e 1.9 usino solo l'additività finita e non la σ -additività, che invece gioca un ruolo fondamentale nella seguente.

Proposizione 1.10. *Sia $P : \mathcal{P}(\Omega) \rightarrow [0, 1]$ una funzione che soddisfa (P1) e l'additività in (1.1). Allora le seguenti proprietà sono equivalenti:*

- (a) P è σ -additiva.
(b) Se $(A_n)_{n \in \mathbb{N}}$ è una successione crescente di eventi, cioè $A_n \subseteq A_{n+1}$ per ogni $n \in \mathbb{N}$, allora

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = \lim_{n \rightarrow +\infty} P(A_n).$$

- (c) Se $(A_n)_{n \in \mathbb{N}}$ è una successione decrescente di eventi, cioè $A_{n+1} \subseteq A_n$ per ogni $n \in \mathbb{N}$, allora

$$P\left(\bigcap_{n=1}^{\infty} A_n\right) = \lim_{n \rightarrow +\infty} P(A_n).$$

Dimostrazione. (a) \Rightarrow (b). Per una data successione crescente $(A_n)_{n \in \mathbb{N}}$ di eventi, definiamo un'altra successione $(B_n)_{n \in \mathbb{N}}$ tramite $B_1 := A_1$, e $B_n := A_n \setminus A_{n-1}$ per $n \geq 2$. Per costruzione, gli eventi B_n sono a due a due disgiunti e, per ogni $n \in \mathbb{N}$,

$$\bigcup_{k=1}^n B_k = A_n.$$

Inoltre

$$\bigcup_{n=1}^{\infty} B_n = \bigcup_{n=1}^{\infty} A_n.$$

Allora, per la σ -additività,

$$\begin{aligned} P\left(\bigcup_{n=1}^{\infty} A_n\right) &= P\left(\bigcup_{n=1}^{\infty} B_n\right) = \sum_{n=1}^{\infty} P(B_n) = \lim_{n \rightarrow +\infty} \sum_{k=1}^n P(B_k) \\ &= \lim_{n \rightarrow +\infty} P\left(\bigcup_{k=1}^n B_k\right) = \lim_{n \rightarrow +\infty} P(A_n). \end{aligned}$$

(b) \Rightarrow (a). Sia (A_n) una successione di eventi a due a due disgiunti. Notando che la successione $(B_n)_{n \in \mathbb{N}}$ definita da $B_n := \bigcup_{k=1}^n A_k$ è crescente e usando l'additività finita e la (b), si ha:

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = P\left(\bigcup_{n=1}^{\infty} B_n\right) = \lim_{n \rightarrow +\infty} P(B_n) = \lim_{n \rightarrow +\infty} \sum_{k=1}^n P(A_k) = \sum_{n=1}^{+\infty} P(A_n).$$

(b) \Rightarrow (c). Sia $(A_n)_{n \in \mathbb{N}}$ una successione decrescente di eventi. Posto $B_n := A_n^c$, $(B_n)_{n \in \mathbb{N}}$ è una successione crescente di eventi. Allora, usando (b), si ha

$$\begin{aligned} P\left(\bigcap_{n=1}^{\infty} A_n\right) &= P\left(\left(\bigcup_{n=1}^{\infty} B_n\right)^c\right) = 1 - P\left(\bigcup_{n=1}^{\infty} B_n\right) \\ &= 1 - \lim_{n \rightarrow +\infty} P(B_n) = \lim_{n \rightarrow +\infty} P(A_n). \end{aligned}$$

(c) \Rightarrow (b). Del tutto simile all'implicazione precedente. Si lasciano i dettagli al lettore. \square

Le proprietà in (b) e (c) nella Proposizione 1.10 vengono dette rispettivamente *continuità dal basso* e *continuità dall'alto* (della probabilità). Un utile corollario è il seguente.

Corollario 1.11 (Subadditività). *Sia $(A_n)_{n \in \mathbb{N}}$ una successione di eventi. Allora*

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

Dimostrazione. Sia $B_n := \bigcup_{k=1}^n A_k$. Evidentemente $(B_n)_{n \in \mathbb{N}}$ è una successione crescente di eventi. Inoltre $\bigcup_{n \in \mathbb{N}} B_n = \bigcup_{n \in \mathbb{N}} A_n$. Per la parte (iii) della Proposizione 1.8, sappiamo che $P(A_1 \cup A_2) \leq P(A_1) + P(A_2)$. Con una facile dimostrazione per induzione, la precedente disuguaglianza si estende a:

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

Ma allora, usando anche la Proposizione 1.10, si ha

$$\begin{aligned}
 P\left(\bigcup_{n=1}^{\infty} A_n\right) &= P\left(\bigcup_{n=1}^{\infty} B_n\right) = \lim_{n \rightarrow +\infty} P(B_n) = \lim_{n \rightarrow +\infty} P\left(\bigcup_{k=1}^n A_k\right) \\
 &\leq \lim_{n \rightarrow +\infty} \sum_{k=1}^n P(A_k) = \sum_{n=1}^{+\infty} P(A_n). \quad \square
 \end{aligned}$$

1.3 Il calcolo combinatorio

Ricordiamo dall'Esempio 1.5 che uno spazio di probabilità discreto (Ω, P) si dice *uniforme* se Ω è un insieme finito e si ha $P(A) = \frac{|A|}{|\Omega|}$, per ogni $A \subseteq \Omega$. Pertanto, il calcolo della probabilità di un evento in uno spazio uniforme si riduce a contarne il numero di elementi. I problemi di conteggio, anche in insiemi abbastanza semplici, sono tipicamente non banali e vanno affrontati con attenzione. Lo strumento matematico fondamentale in questo contesto è il *calcolo combinatorio*, che ora descriviamo.

1.3.1 Principi basilari

Dati due insiemi A, B , si dice che A è *in corrispondenza biunivoca* con B se esiste un'applicazione $f : A \rightarrow B$ biunivoca, cioè iniettiva e suriettiva. Chiaramente A è in corrispondenza biunivoca con B se e soltanto se B è in corrispondenza biunivoca con A : si scrive talvolta “ A e B sono in corrispondenza biunivoca”, che rende palese la simmetria della relazione (si tratta in effetti di una relazione di equivalenza). Dato $n \in \mathbb{N}$, si dice che un insieme A ha *cardinalità* n e si scrive $|A| = n$ se A è in corrispondenza biunivoca con l'insieme $\{1, 2, \dots, n\}$. Si noti che la proprietà “ A ha cardinalità n ” è la formalizzazione matematica dell'affermazione intuitiva “ A ha n elementi”. *In questo paragrafo considereremo solo insiemi finiti*, cioè insiemi che hanno cardinalità n per un opportuno $n \in \mathbb{N}$.

Per determinare la cardinalità di un insieme, la strategia tipica consiste nel ricondurre il calcolo all'applicazione combinata (talvolta non banale) di alcuni principi o osservazioni basilari. Una prima osservazione, elementare ma molto utile, è che se un insieme A è in corrispondenza biunivoca con un insieme B , allora $|A| = |B|$. Un'altra osservazione, anch'essa molto intuitiva, è la seguente: se A, B sono due sottoinsiemi (di uno stesso spazio) *disgiunti*, cioè tali che $A \cap B = \emptyset$, allora $|A \cup B| = |A| + |B|$. Più in generale, se A_1, \dots, A_k sono sottoinsiemi a due a due disgiunti, tali cioè che $A_i \cap A_j = \emptyset$ per $i \neq j$, allora $|\bigcup_{i=1}^k A_i| = \sum_{i=1}^k |A_i|$. La dimostrazione di queste osservazioni è semplice ed è lasciata per esercizio.

Un principio leggermente meno elementare riguarda la cardinalità degli *insiemi prodotto*. Ricordiamo che, dati due insiemi A, B , il loro prodotto cartesiano $A \times B$ è definito come l'insieme delle coppie ordinate (a, b) , con $a \in A$ e $b \in B$. Vale allora la relazione $|A \times B| = |A||B|$. Il modo più semplice per convincersi della validità di questa formula consiste nel disporre gli elementi di $A \times B$ in una tabella rettangolare, dopo aver numerato gli elementi dei due insiemi. Più precisamente, se $A = \{a_1, a_2, \dots, a_m\}$, $B = \{b_1, b_2, \dots, b_k\}$, possiamo

elencare gli elementi dell'insieme $A \times B$ nel modo seguente:

$$\begin{pmatrix} (a_1, b_1) & (a_1, b_2) & \cdots & (a_1, b_{k-1}) & (a_1, b_k) \\ (a_2, b_1) & (a_2, b_2) & \cdots & (a_2, b_{k-1}) & (a_2, b_k) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ (a_m, b_1) & (a_m, b_2) & \cdots & (a_m, b_{k-1}) & (a_m, b_k) \end{pmatrix},$$

da cui è chiaro che $|A \times B| = m \cdot k = |A| \cdot |B|$.

Diamo ora una dimostrazione più formale. Per $x \in A$ indichiamo con $\{x\} \times B$ il sottoinsieme di $A \times B$ costituito dagli elementi che hanno x come prima componente, cioè $\{x\} \times B := \{(x, b) : b \in B\}$. Possiamo quindi scrivere $A \times B = \cup_{x \in A} (\{x\} \times B)$, e si noti che questa unione è disgiunta, poiché $(\{x_1\} \times B) \cap (\{x_2\} \times B) = \emptyset$ se $x_1 \neq x_2$. Per l'osservazione enunciata sopra si ha dunque $|A \times B| = \sum_{x \in A} |\{x\} \times B|$. Si noti ora che l'insieme $\{x\} \times B$ è in corrispondenza biunivoca con B , per qualunque $x \in A$: la corrispondenza è data semplicemente da $(x, b) \mapsto b$. Di conseguenza $|\{x\} \times B| = |B|$ e si ottiene la formula $|A \times B| = \sum_{x \in A} |B| = |A| |B|$.

Per induzione si estende facilmente la formula al caso di più di due fattori: più precisamente, se A_1, \dots, A_k sono insiemi finiti, l'insieme prodotto $A_1 \times \cdots \times A_k$, definito come l'insieme delle k -uple (a_1, \dots, a_k) , con $a_i \in A_i$, ha cardinalità data dalla formula $|A_1 \times \cdots \times A_k| = |A_1| \cdots |A_k| = \prod_{i=1}^k |A_i|$. Un'ulteriore estensione di questa formula, elementare ma non banale, conduce a quello che è noto come il principio fondamentale del calcolo combinatorio. Prima di vedere di che cosa si tratta, discutiamo qualche applicazione delle formule appena viste.

1.3.2 Disposizioni con ripetizione

Dato un insieme $A = \{a_1, \dots, a_n\}$ di cardinalità $n \in \mathbb{N}$ e dato $k \in \mathbb{N}$, le funzioni definite su $\{1, \dots, k\}$ a valori in A sono dette *disposizioni con ripetizione di k elementi estratti da A* . È facile vedere che le disposizioni con ripetizione sono in corrispondenza biunivoca naturale con gli elementi dell'insieme $A^k := A \times \cdots \times A$ (k volte): la corrispondenza è quella che a $(x_1, \dots, x_k) \in A^k$ associa la funzione $f : \{1, \dots, k\} \rightarrow A$ definita da $f(i) := x_i$. La formula sulla cardinalità degli insiemi prodotto dà $|A^k| = |A|^k = n^k$: ci sono dunque n^k possibili disposizioni con ripetizione di k elementi estratti da un insieme di n elementi.

Una disposizione con ripetizione può dunque essere vista come una sequenza *ordinata* (x_1, \dots, x_k) di elementi $x_i \in A$, *non necessariamente distinti*: si può cioè avere $x_i = x_j$ per $i \neq j$. Sottolineiamo che l'ordine in cui compaiono gli elementi è importante: per esempio, (a_1, a_2) e (a_2, a_1) sono due disposizioni differenti.

Esempio 1.12. (1) I compleanni di un gruppo ordinato di 4 persone costituiscono una disposizione con ripetizione di 4 elementi estratti dall'insieme dei giorni dell'anno, che ha cardinalità 366 (contando il 29 febbraio). Sono dunque possibili $366^4 \approx 1.8 \cdot 10^{10}$ sequenze distinte di compleanni.

(2) Per compilare una colonna di una schedina del Totocalcio occorre scegliere, per ciascuna delle 13 partite in esame, tra la vittoria della squadra di casa (1), il pareggio (x) o la vittoria della squadra in trasferta (2). Una colonna compilata è dunque una disposizione con ripetizione di 13 elementi estratti dall'insieme $\{1, x, 2\}$ e di conseguenza ci sono $3^{13} \approx 1.6 \cdot 10^6$ modi possibili di compilare una colonna.

(3) Le possibili "parole" (anche prive di significato) costituite da 10 lettere dell'alfabeto inglese coincidono con le disposizioni con ripetizione di 10 elementi estratti da un insieme

che ne contengono 26: il loro numero è dunque pari a $26^{10} \approx 1.4 \cdot 10^{14}$. Le parole che effettivamente hanno un significato (per esempio nella lingua inglese) sono naturalmente molte meno: anche includendo i termini tecnici, il numero totale di parole di qualunque lunghezza della lingua inglese non supera il milione. Di conseguenza, la probabilità che digitando una sequenza di dieci lettere a caso si ottenga una parola di senso compiuto è certamente minore di $10^6 / (1.4 \cdot 10^{14}) < 10^{-8}$.

Osservazione 1.13. Se $B = \{b_1, \dots, b_k\}$ è un insieme di cardinalità $k \in \mathbb{N}$, si indica con A^B l'insieme di tutte le funzioni da B in A . L'insieme A^B è in corrispondenza biunivoca con A^k : una corrispondenza è per esempio quella che a $(x_1, \dots, x_k) \in A^k$ associa la funzione $f \in A^B$ definita da $f(b_i) := x_i$. Come conseguenza della formula sulla cardinalità degli insiemi prodotto, otteniamo dunque che $|A^B| = |A|^k = n^k$, cioè $|A^B| = |A|^{|B|}$.

1.3.3 Il principio fondamentale del calcolo combinatorio

Un esempio molto ricorrente nelle applicazioni è quello in cui gli elementi di un insieme possano essere determinati attraverso scelte successive. Per esempio, sia E l'insieme delle funzioni iniettive da $\{1, \dots, k\}$ in A (si noti che necessariamente $k \leq n$). Possiamo determinare ogni funzione $f \in E$ scegliendo innanzitutto la prima componente $f(1)$ come un elemento qualunque di A , quindi scegliendo la seconda componente $f(2)$ come un elemento qualunque di $A \setminus \{f(1)\}$, e così via. Abbiamo n esiti possibili per la scelta di $f(1)$, $(n-1)$ per la scelta di $f(2)$, \dots , $(n-k+1)$ per la scelta di $f(k)$. Per analogia con gli insiemi prodotto, dovrebbe essere intuitivamente chiaro che $|E| = n \cdot (n-1) \cdots (n-k+1)$. Si noti che l'insieme dei valori ammissibili per $f(i)$ dipende dagli esiti delle scelte precedenti, tuttavia il numero di valori ammissibili è sempre lo stesso, pari a $n-i+1$. Generalizzando questo esempio, giungiamo al *principio fondamentale del calcolo combinatorio*, che possiamo formulare come segue.

Teorema 1.14 (Principio fondamentale del calcolo combinatorio). *Supponiamo che gli elementi di un insieme E possano essere determinati mediante k scelte successive, in cui ogni scelta abbia un numero fissato di esiti possibili: la prima scelta ha n_1 esiti possibili, la seconda scelta ne ha n_2 , \dots , la k -esima scelta ne ha n_k , dove $n_1, \dots, n_k \in \mathbb{N}$. Supponiamo inoltre che sequenze distinte di esiti determinino elementi distinti di E . Allora $|E| = n_1 \cdot n_2 \cdots n_k$.*

Così enunciato, questo principio può apparire un po' vago (per esempio, il concetto di "scelta" non è stato definito precisamente). Una riformulazione matematicamente precisa del Teorema 1.14, con la relativa dimostrazione, è data dal Teorema 1.16 più in basso, che comporta tuttavia notazioni abbastanza pesanti e risulta di poco aiuto per l'applicazione del principio a casi concreti. Nella pratica, si fa tipicamente riferimento all'enunciato del Teorema 1.14. L'idea cruciale è che gli elementi dell'insieme E possono essere messi *in corrispondenza biunivoca* con le sequenze di esiti delle scelte, che hanno una struttura di spazio prodotto, da cui segue la formula per la cardinalità. La condizione che sequenze distinte di esiti determinino elementi distinti di E serve proprio a garantire che la corrispondenza sia biunivoca: la mancata verifica di questa condizione è la principale fonte di errori nell'applicazione del principio. Qualche esempio chiarirà la situazione.

Esempio 1.15. (a) Un mazzo di carte da poker è costituito da 52 carte, identificate dal *seme* (cuori \heartsuit , quadri \diamondsuit , fiori \clubsuit , picche \spadesuit) e dal *tipo* (un numero da 1 a 10 oppure J, Q, K). Indichiamo con E l'insieme delle carte di numero pari (figure escluse) e di colore

rosso (cioè di cuori o di quadri). Ogni elemento di E può essere determinato attraverso due scelte successive: la scelta del seme, che ha 2 esiti possibili (cuori e quadri), e la scelta del tipo, che ne ha 5 (cioè 2, 4, 6, 8, 10). Segue dunque che $|E| = 2 \cdot 5 = 10$.

- (b) Dato un mazzo di carte da poker, si chiama *full* un sottoinsieme di 5 carte costituito dall'unione di un *tris* (un sottoinsieme di 3 carte dello stesso tipo) e di una *coppia* (un sottoinsieme di 2 carte dello stesso tipo). Indichiamo con E l'insieme dei possibili full. Sottolineiamo che gli elementi di E sono *sottoinsiemi* di 5 carte, non disposizioni: in particolare, le carte non sono ordinate.

Gli elementi di E possono essere determinati univocamente attraverso 4 scelte successive: 1) il tipo del tris; 2) il tipo della coppia; 3) i semi delle carte che compaiono nel tris; 4) i semi delle carte che compaiono nella coppia. Per la prima scelta ci sono 13 esiti possibili, per la seconda scelta, qualunque sia l'esito della prima scelta, ci sono 12 esiti possibili (chiaramente i due tipi devono essere differenti, perché non esistono cinque carte dello stesso tipo). Per la terza scelta, occorre scegliere tre semi nell'insieme {cuori, quadri, fiori, picche}: per enumerazione diretta, è facile vedere che ci sono 4 esiti possibili; analogamente, per la quarta scelta occorre scegliere due semi e per questo ci sono 6 esiti possibili (ritorneremo nell'Esempio 1.18 sul modo di contare i sottoinsiemi). Applicando il Teorema 1.14 si ottiene dunque che $|E| = 13 \cdot 12 \cdot 4 \cdot 6 = 3\,744$.

- (c) Dato un mazzo di carte da poker, indichiamo con E l'insieme delle *doppie coppie*, cioè i sottoinsiemi di 5 carte costituiti dall'unione di due *coppie* di tipi diversi, più una quinta carta di tipo diverso dai tipi delle due coppie.

Per determinare $|E|$ si potrebbe essere tentati di procedere analogamente al caso dei full, attraverso sei scelte successive: 1) il tipo della prima coppia; 2) il tipo della seconda coppia; 3) il tipo della “quinta carta”; 4) i semi delle carte che compaiono nella prima coppia; 5) i semi delle carte che compaiono nella seconda coppia; 6) il seme della “quinta carta”. Ci sono 13 esiti possibili per la prima scelta, 12 per la seconda scelta, 11 per la terza, 6 per la quarta, 6 per la quinta, 4 per la sesta: si otterrebbe dunque $|E| = 13 \cdot 12 \cdot 11 \cdot 6^2 \cdot 4 = 247\,104$. Tuttavia questo risultato è errato.

La ragione è che le scelte 1) e 2) sono ambigue, dal momento che non esiste una “prima” e una “seconda” coppia. In effetti, sequenze distinte di esiti delle sei scelte sopra elencate non conducono a elementi distinti di E : ciascun elemento di E , cioè ciascuna doppia coppia, viene infatti selezionata *esattamente due volte*. Per esempio, la doppia coppia $\{5\heartsuit, 5\diamondsuit, 6\heartsuit, 6\clubsuit, 7\spadesuit\}$ viene determinata sia con l'esito “5” della scelta 1) e l'esito “6” della scelta 2), sia viceversa. Per tale ragione, il risultato corretto è $|E| = 247\,104/2 = 123\,552$, cioè la metà di quanto ottenuto in precedenza.

Un modo alternativo di ottenere il risultato corretto è di riunire le scelte 1) e 2) nell'unica scelta 1bis) “i tipi delle due coppie”, che ha $13 \cdot 12/2 = 78$ esiti possibili (anche su questo torneremo nell'Esempio 1.18). Le scelte 1bis), 3), 4), 5) e 6) permettono di applicare correttamente il Teorema 1.14, ottenendo $|E| = 78 \cdot 11 \cdot 6^2 \cdot 4 = 123\,552$.

Passiamo ora a riformulare il principio fondamentale del calcolo combinatorio in modo matematicamente più preciso. Occorre innanzitutto esprimere astrattamente il concetto di “scelta”, in modo che possa essere applicato a insiemi arbitrari, che non abbiano necessariamente una struttura di spazio prodotto. Dato un insieme E , definiamo una *scelta su E* come una *partizione* di E , vale a dire una famiglia di sottoinsiemi $\{E_1, \dots, E_m\}$ tali che $E = \bigcup_{i=1}^m E_i$ e $E_i \cap E_j = \emptyset$ per $i \neq j$. Il “numero di esiti della scelta” è per definizione il

numero m di elementi della partizione. Intuitivamente, l'indice i numera gli "esiti" della scelta mentre l'insieme E_i corrisponde agli elementi di E compatibili con l'esito i della scelta. Per esempio, riconsideriamo l'insieme E delle funzioni iniettive da $\{1, \dots, k\}$ in $A = \{a_1, \dots, a_n\}$: la scelta della prima componente corrisponde alla partizione $\{E_1, \dots, E_n\}$ definita da $E_i := \{f \in E : f(1) = a_i\}$ e ha dunque n esiti possibili.

Estendiamo ora la definizione: fissato un insieme E , due scelte successive su E sono il dato di una partizione $\{E_1, \dots, E_m\}$ di E (che rappresenta la prima scelta) e, per ogni elemento E_i di tale partizione, una partizione $\{E_{i,1}, \dots, E_{i,k_i}\}$ di E_i (che rappresenta la seconda scelta). Si noti che la seconda scelta non è "una scelta su E ", come definita sopra, ma piuttosto una scelta su E_i , per ogni $i = 1, \dots, m$. In particolare, il numero k_i di esiti della seconda scelta può in generale dipendere dall'esito i della prima scelta. Nel caso in cui ciò accada, cioè se $k_i = k$ per ogni $i = 1, \dots, m$, diremo che la seconda scelta ha un numero fissato k di esiti possibili. Ritornando all'esempio delle funzioni iniettive da $\{1, \dots, k\}$ in $A = \{a_1, \dots, a_n\}$, la scelta delle prime due componenti è un esempio di due scelte successive su E : infatti, per ogni elemento $E_i = \{f \in E : f(1) = a_i\}$ della prima scelta, la seconda scelta è data dalla partizione $\{E_{i,j}\}_{j \in \{1, \dots, n\} \setminus \{i\}}$ definita da $E_{i,j} := \{f \in E : f(1) = a_i, f(2) = a_j\}$. In particolare, la seconda scelta ha un numero fissato $n - 1$ di esiti possibili.[†]

Il passaggio da due a k scelte successive è solo notazionalmente più complicato. Per definizione, dato un insieme E , k scelte successive su E sono il dato di una famiglia di partizioni, definite nel modo seguente:

- la prima scelta è una partizione $\{E_1, \dots, E_{n_1}\}$ di E ;
- per ogni $2 \leq j \leq k$ e per ogni elemento $E_{i_1, \dots, i_{j-1}}$ della $(j - 1)$ -esima scelta, la j -esima scelta è una partizione $\{E_{i_1, \dots, i_{j-1}, \ell}\}_{\ell \in \{1, \dots, n_j^*\}}$ di $E_{i_1, \dots, i_{j-1}}$, dove il numero n_j^* di elementi della partizione (cioè il numero di esiti della j -esima scelta) può in generale dipendere dagli esiti delle scelte precedenti: $n_j^* = n_j^*(i_1, \dots, i_{j-1})$.

Nel caso in cui $n_j^*(i_1, \dots, i_{j-1}) = n_j$ non dipenda da i_1, \dots, i_{j-1} , diremo che la j -esima scelta ha un numero fissato n_j di esiti possibili.

Alla luce delle notazioni appena introdotte, possiamo riformulare il Teorema 1.14 nel modo seguente.

Teorema 1.16. *Siano definite k scelte successive su un insieme E , tali che ogni scelta abbia un numero fissato di esiti possibili: la prima scelta ha n_1 esiti possibili, la seconda scelta ne ha n_2, \dots , la k -esima scelta ne ha n_k , dove $n_1, \dots, n_k \in \mathbb{N}$. Supponiamo che gli elementi di E siano determinati univocamente dalle k scelte, cioè $|E_{i_1, \dots, i_k}| = 1$ per ogni scelta di i_1, \dots, i_k . Allora la cardinalità di E è pari a $n_1 \cdot n_2 \cdot \dots \cdot n_k$.*

Dimostrazione. Per definizione di scelte successive, $E = \bigcup_{i=1}^{n_1} E_i$; a sua volta $E_i = \bigcup_{j=1}^{n_2} E_{i,j}$, eccetera: di conseguenza vale la relazione

$$E = \bigcup_{i_1=1}^{n_1} \dots \bigcup_{i_k=1}^{n_k} E_{i_1, \dots, i_k}. \quad (1.12)$$

Mostriamo che questa unione è disgiunta, cioè $E_{i_1, \dots, i_k} \cap E_{i'_1, \dots, i'_k} = \emptyset$ se $(i_1, \dots, i_k) \neq (i'_1, \dots, i'_k)$. Se $(i_1, \dots, i_k) \neq (i'_1, \dots, i'_k)$ significa che $i_j \neq i'_j$ per qualche $1 \leq j \leq k$: prendendo il più piccolo di tali valori di j , possiamo supporre che $i_1 = i'_1, \dots, i_{j-1} = i'_{j-1}$ mentre $i_j \neq i'_j$. Per definizione di scelte successive, $E_{i_1, \dots, i_k} \subseteq E_{i_1, \dots, i_{j-1}, i_j}$ e analogamente $E_{i'_1, \dots, i'_k} \subseteq E_{i'_1, \dots, i'_{j-1}, i'_j} = E_{i_1, \dots, i_{j-1}, i'_j}$, per cui basta mostrare che $E_{i_1, \dots, i_{j-1}, i_j} \cap E_{i_1, \dots, i_{j-1}, i'_j} = \emptyset$. Ma per ipotesi gli insiemi $\{E_{i_1, \dots, i_{j-1}, \ell}\}_{\ell \in \{1, \dots, n_j\}}$ formano una partizione di $E_{i_1, \dots, i_{j-1}}$, in particolare sono a due a due disgiunti: quindi $E_{i_1, \dots, i_{j-1}, i_j} \cap E_{i_1, \dots, i_{j-1}, i'_j} = \emptyset$ poiché $i_j \neq i'_j$. Essendo l'unione in (1.12) disgiunta e ricordando che per ipotesi $|E_{i_1, \dots, i_k}| = 1$, si ottiene

$$|E| = \sum_{i_1=1}^{n_1} \dots \sum_{i_k=1}^{n_k} |E_{i_1, \dots, i_k}| = n_1 \cdot n_2 \cdot \dots \cdot n_k,$$

che è la relazione voluta. □

Ritornando per un'ultima volta all'insieme E delle funzioni iniettive da $\{1, \dots, k\}$ in $A = \{a_1, \dots, a_n\}$, è facile verificare che valgono le ipotesi del Teorema 1.16: le scelte di $f(1), f(2), \dots, f(k)$ costituiscono k scelte successive su E e gli elementi di E sono univocamente determinati da queste scelte. Dato che la scelta di $f(i)$ ha $n - i + 1$ esiti possibili, segue dal Teorema 1.14 che vale la formula $|E| = n(n - 1) \cdot \dots \cdot (n - k + 1)$ ottenuta in precedenza.

[†]Si noti che è risultato conveniente indicizzare gli esiti della seconda scelta usando l'insieme $\{1, \dots, n\} \setminus \{i\}$ piuttosto che $\{1, \dots, n - 1\}$.

1.3.4 Disposizioni semplici e permutazioni

Dato un insieme $A = \{a_1, \dots, a_n\}$ di cardinalità $n \in \mathbb{N}$ e dato $k \in \mathbb{N}$, abbiamo visto che le funzioni da $\{1, \dots, k\}$ a valori in A sono dette disposizioni con ripetizione di k elementi estratti da A , e hanno cardinalità n^k . Se $k \leq n$, le funzioni *iniettive* da $\{1, \dots, k\}$ in A sono dette *disposizioni semplici (o senza ripetizione) di k elementi estratti da A* . Abbiamo già incontrato questo insieme di funzioni nello scorso paragrafo e abbiamo visto che la sua cardinalità è data dalla formula $n(n-1) \cdots (n-k+1)$.

Nel caso speciale in cui $k = n$, le disposizioni semplici di n elementi estratti da A sono dette *permutazioni di A* . Si osservi che una permutazione $f : A \rightarrow A$ può essere vista come una elencazione ordinata di tutti gli elementi di A , cioè $(f(1), f(2), \dots, f(n))$. È interessante notare che l'insieme delle permutazioni di un insieme fissato A costituisce un *gruppo* rispetto alla composizione di applicazioni, che è non commutativo per $n \geq 3$. A meno di corrispondenze biunivoche, non costa niente considerare il caso “speciale” $A = \{1, \dots, n\}$: in questo caso, il gruppo delle permutazioni è indicato con S_n . Munito della probabilità uniforme, lo spazio S_n ha proprietà interessanti e per certi versi sorprendenti, alcune delle quali verranno discusse nel paragrafo 2.1.

Introduciamo il simbolo $n!$, detto “ n fattoriale”, definito da

$$n! := n(n-1) \cdots 1 = \prod_{i=1}^n i \quad \text{per } n \in \mathbb{N}, \quad 0! := 1. \quad (1.13)$$

Per quanto abbiamo visto, la cardinalità del gruppo delle permutazioni di un insieme di n elementi è pari a $n!$ (in particolare $|S_n| = n!$). Analogamente, possiamo riscrivere la cardinalità delle disposizioni semplici di k elementi estratti da un insieme che ne contiene n come $n!/(n-k)!$.

Esempio 1.17. Supponiamo di mischiare un mazzo di carte da poker. La sequenza ordinata delle carte che ne risulta è una permutazione delle carte del mazzo. Il numero delle possibili sequenze ottenute in questo modo è dunque pari $52! \approx 8 \cdot 10^{67}$.

1.3.5 Combinazioni

Sia $A = \{a_1, \dots, a_n\}$ un insieme di cardinalità $n \in \mathbb{N}$ e sia $k \in \mathbb{N}_0$ con $0 \leq k \leq n$. I sottoinsiemi di A di cardinalità k sono detti *combinazioni di k elementi estratti da A* . Se una disposizione semplice corrisponde a una sequenza ordinata, una combinazione può essere vista come una collezione *non ordinata* di elementi. Indichiamo con $\mathcal{C}_{n,k}$ l'insieme delle combinazioni e con $\mathcal{D}_{n,k}$ l'insieme delle disposizioni semplici di k elementi estratti da A .

Per $k = 0$ si ha $\mathcal{C}_{n,0} = \{\emptyset\}$ e dunque $|\mathcal{C}_{n,0}| = 1$. Per determinare $|\mathcal{C}_{n,k}|$ per $k \in \{1, \dots, n\}$, ricordiamo che ci sono $|\mathcal{D}_{n,k}| = n!/(n-k)!$ disposizioni (cioè sequenze ordinate) di k elementi distinti estratti da A . Dato che nelle combinazioni l'ordine degli elementi non conta, dobbiamo identificare le disposizioni che danno origine alla stessa combinazione, cioè che selezionano lo stesso sottoinsieme di A : dato che ci sono $k!$ riordinamenti possibili (cioè permutazioni) di k elementi fissati, si ottiene $|\mathcal{C}_{n,k}| = |\mathcal{D}_{n,k}|/k! = \binom{n}{k}$, dove abbiamo introdotto il *coefficiente binomiale*, definito da

$$\binom{n}{k} := \frac{n!}{k!(n-k)!}, \quad \text{per } n \in \mathbb{N}_0, \quad k \in \{0, \dots, n\}.$$

Si noti che la formula $|\mathcal{C}_{n,k}| = \binom{n}{k}$ vale anche per $k = 0$.

Questo argomento può essere formalizzato in modo più preciso (e più tecnico). Cominciamo con un piccolo risultato preparatorio. Siano D, E due insiemi finiti e sia $g : D \rightarrow E$ un'applicazione suriettiva. Per ogni $y \in E$, introduciamo il sottoinsieme $g^{-1}(y) := \{x \in D : g(x) = y\}$ costituito dagli elementi di D che vengono mandati da g in y . Supponiamo che valga la seguente proprietà: esiste $k \in \mathbb{N}$ tale che, per ogni $y \in E$, si ha $|g^{-1}(y)| = k$ (cioè per ogni $y \in E$ esistono esattamente k elementi $x \in D$ che vengono mandati da g in y). Segue allora che $|D| = k|E|$. La dimostrazione è semplice: possiamo sempre scrivere $E = \bigcup_{y \in D} g^{-1}(y)$ e inoltre l'unione è disgiunta (esercizio). Quindi $|E| = \sum_{y \in D} |g^{-1}(y)| = \sum_{y \in D} k = k|D|$.

Fissiamo ora $k \in \{1, \dots, n\}$ e definiamo una applicazione $g : \mathcal{D}_{n,k} \rightarrow \mathcal{C}_{n,k}$ nel modo seguente: data $f \in \mathcal{D}_{n,k}$, definiamo $g(f) := \text{Im}(f)$, dove $\text{Im}(f)$ indica l'immagine di f (ricordiamo che f è una funzione iniettiva da $\{1, \dots, k\}$ in A). È immediato verificare che g è ben definita, cioè effettivamente $g(f) \in \mathcal{C}_{n,k}$ per ogni $f \in \mathcal{D}_{n,k}$, e che g è suriettiva. Se mostriamo che $|g^{-1}(B)| = k!$, per ogni $B \in \mathcal{C}_{n,k}$, si ottiene $|\mathcal{D}_{n,k}| = k!|\mathcal{C}_{n,k}|$ e quindi la formula $|\mathcal{C}_{n,k}| = \binom{n}{k}$ è dimostrata.

Indichiamo con S_B l'insieme delle permutazioni di B , cioè le applicazioni $\pi : B \rightarrow B$ biunivoche, e fissiamo un elemento arbitrario $f_0 \in g^{-1}(B)$. È molto facile convincersi che, per ogni $\pi \in S_B$, si ha $\pi \circ f_0 \in g^{-1}(B)$: infatti l'applicazione $\pi \circ f_0$ è iniettiva, perché lo sono sia f_0 sia π , e $\text{Im}(\pi \circ f_0) = \text{Im}(f_0) = B$, perché π è una permutazione di B . Risulta dunque ben posta l'applicazione $H : S_B \rightarrow g^{-1}(B)$ definita da $H(\pi) := \pi \circ f_0$. Supponiamo che $H(\pi_1) = H(\pi_2)$: per ogni $b \in B$, se $i \in \{1, \dots, k\}$ è tale che $f_0(i) = b$ (tale i esiste perché $\text{Im}(f_0) = B$), otteniamo $(\pi_1 \circ f_0)(i) = (\pi_2 \circ f_0)(i)$, cioè $\pi_1(b) = \pi_2(b)$; dato che $b \in B$ è arbitrario, segue che $\pi_1 = \pi_2$, dunque l'applicazione H è iniettiva. Se ora consideriamo un arbitrario $f \in g^{-1}(B)$, è facile costruire $\pi \in S_B$ tale che $\pi \circ f_0 = f$, cioè $H(\pi) = f$, quindi l'applicazione H è suriettiva. Avendo mostrato che H è biunivoca, segue che gli insiemi S_B e $g^{-1}(B)$ sono in corrispondenza biunivoca e dunque $|g^{-1}(B)| = |S_B| = k!$ se $B \in \mathcal{C}_{n,k}$, che è quanto restava da dimostrare.

Elenchiamo alcune semplici proprietà dei coefficienti binomiali:

$$\binom{n}{0} = \binom{n}{n} = 1, \quad \binom{n}{k} = \binom{n}{n-k}, \quad \forall n \in \mathbb{N}_0, \forall k \in \{0, \dots, n\}.$$

Vale inoltre la relazione

$$\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k}, \quad \forall n \in \mathbb{N}, \forall k \in \{0, \dots, n\}, \quad (1.14)$$

come si verifica facilmente.[†] Ricordiamo infine la formula nota come *binomio di Newton*:

$$(a+b)^n = \sum_{k=0}^n \binom{n}{k} a^k b^{n-k}, \quad \forall n \in \mathbb{N}_0, \forall a, b \in \mathbb{R}, \quad (1.15)$$

che si dimostra per induzione usando (1.14).

Esempio 1.18. Ritornando brevemente all'Esempio 1.15, il numero di modi di scegliere 3 “semi” tra i quattro possibili $\{\heartsuit, \diamondsuit, \clubsuit, \spadesuit\}$ è pari al numero di combinazioni di 3 elementi estratti da un insieme che ne contiene 4 ed è dunque dato da $\binom{4}{3} = 4$ (come avevamo concluso per enumerazione diretta). Analogamente, il numero di modi di scegliere 2 semi è pari a $\binom{4}{2} = 6$ e il numero di modi di scegliere due “tipi” tra i 13 possibili è pari a $\binom{13}{2} = 78$.

Una “mano” a Poker è un sottoinsieme di 5 carte distinte estratte da un mazzo che ne contiene 52. Il numero di possibili mani è dato dunque da $\binom{52}{5} = 2\,598\,960$. Ricordando l'Esempio 1.15, le probabilità di fare *full* oppure *doppia coppia* valgono rispettivamente $3\,744/2\,598\,960 \approx 0.14\%$ e $123\,552/2\,598\,960 \approx 4,8\%$.

[†]Esiste anche una dimostrazione combinatoria della relazione (1.14). Le combinazioni di k elementi estratte da $\{1, \dots, n\}$, che sono $\binom{n}{k}$, possono essere infatti divise in due sottoinsiemi: quelle che contengono 1 e quelle che non lo contengono; le prime sono in corrispondenza biunivoca con le combinazioni di $(k-1)$ elementi estratti da $\{2, \dots, n\}$, e sono dunque $\binom{n-1}{k-1}$, mentre le seconde sono in corrispondenza biunivoca con le combinazioni di k elementi estratti da $\{2, \dots, n\}$, che sono $\binom{n-1}{k}$.

Esempio 1.19. Si consideri un'urna contenente N palline, di cui m rosse e $N - m$ verdi (con $m \leq N$). Supponiamo di eseguire n estrazioni successive, secondo uno dei seguenti due schemi di estrazione:

- *Estrazioni con reimmissione.* Dopo ogni estrazione, la pallina estratta viene reinserita nell'urna.
- *Estrazioni senza reimmissione.* Le palline estratte non vengono reinserite. In questo caso dev'essere $n \leq N$.

Calcoliamo, per ciascuno dei due schemi, la probabilità che esattamente k delle n palline estratte siano rosse.

Caso di estrazioni con reimmissione. Supponiamo di numerare le palline da 1 a N e, per fissare le idee, assumiamo che le palline rosse siano quelle numerate da 1 a m . L'esito di n estrazioni successive può essere interpretato come una disposizione con ripetizione di n elementi presi dall'insieme $\{1, 2, \dots, N\}$. Sia dunque Ω l'insieme di tali disposizioni, e P la probabilità uniforme su Ω . Sappiamo che $|\Omega| = N^n$. Denotiamo infine con A l'insieme delle disposizioni contenenti esattamente k palline rosse. Si tratta di calcolare

$$P(A) = \frac{|A|}{|\Omega|}.$$

Per determinare $|A|$, utilizziamo il principio fondamentale. Un elemento di A è determinato dalle seguenti scelte successive.

- Si scelgono le k posizioni (su n possibili) in cui mettere le palline rosse: per questa scelta ci sono $\binom{n}{k}$ esiti possibili.
- Si dispongono k palline rosse (prese dalle m presenti nell'urna) nelle posizioni prescelte: ci sono m^k tali disposizioni.
- Si dispongono $(n - k)$ palline verdi (prese dalle $(N - m)$ presenti nell'urna) nelle rimanenti posizioni: ci sono $(N - m)^{n-k}$ tali disposizioni.

Si ottiene pertanto

$$|A| = \binom{n}{k} m^k (N - m)^{n-k},$$

da cui segue facilmente che

$$P(A) = \binom{n}{k} \left(\frac{m}{N}\right)^k \left(1 - \frac{m}{N}\right)^{n-k}.$$

Questo risultato verrà reinterpretato più avanti, nell'Esempio 1.37.

Caso di estrazioni senza reimmissione. Enumeriamo le palline come nel caso precedente. Un naturale spazio campionario, in cui la probabilità uniforme esprime la casualità dell'estrazione, è quello delle *disposizioni senza ripetizione*. Poichè, tuttavia, l'evento "il numero di palline rosse estratte è k " non dipende dall'ordine di estrazione, è forse ancora più naturale scegliere come spazio campionario l'insieme delle *combinazioni*. Sia dunque Ω l'insieme dei sottoinsiemi di n elementi dell'insieme $\{1, 2, \dots, N\}$, e P la probabilità uniforme su di esso. L'evento di cui vogliamo calcolare la probabilità è

$$A = \{\omega \in \Omega : |\omega \cap \{1, 2, \dots, m\}| = k\} = \{\omega \in \Omega : |\omega \cap \{m + 1, \dots, N\}| = n - k\}.$$

Chiaramente $A = \emptyset$ se $k > m$ oppure se $(n-k) > (N-m)$. Se invece $k \leq m$ e $(n-k) \leq (N-m)$, ogni elemento di A è determinato da due scelte successive: occorre scegliere k elementi da $\{1, 2, \dots, m\}$ e $(n-k)$ da $\{m+1, \dots, N\}$. Di conseguenza possiamo scrivere

$$|A| = \binom{m}{k} \binom{N-m}{n-k},$$

dove usiamo la convenzione secondo cui $\binom{i}{j} = 0$ se $j < 0$ o $j > i$. Ricordando che $|\Omega| = \binom{N}{n}$, possiamo dunque concludere che

$$P(A) = \frac{\binom{m}{k} \binom{N-m}{n-k}}{\binom{N}{n}}.$$

1.3.6 Coefficienti multinomiali*

Dato un insieme A , ricordiamo che una partizione di A è una famiglia $\{A_1, \dots, A_r\}$ di sottoinsiemi di A tali che $A_i \cap A_j = \emptyset$, per ogni $i \neq j$, e $\bigcup_{i=1}^r A_i = A$. Sottolineiamo che l'ordine dei sottoinsiemi che compaiono nella partizione è importante: $\{A_1, A_2, A_3\}$ e $\{A_2, A_1, A_3\}$ sono due partizioni distinte. Supponiamo ora che $|A| = n \in \mathbb{N}_0$ e siano assegnati $r \in \mathbb{N}$ e $k_1, \dots, k_r \in \mathbb{N}_0$ tali che $k_1 + \dots + k_r = n$: ci proponiamo di calcolare il numero di partizioni $\{A_1, \dots, A_r\}$ di A tali che $|A_1| = k_1, \dots, |A_r| = k_r$.

Il caso $r = 1$ è banale: esiste infatti un'unica partizione di A in un solo sottoinsieme, che è naturalmente A stesso. Il caso $r = 2$ è strettamente connesso con le combinazioni: infatti una partizione in due sottoinsiemi $\{A_1, A_2\}$ è identificata da A_1 , poiché $A_2 = A_1^c$; dovendo essere $|A_1| = k_1$, il numero di tali partizioni coincide con il numero di combinazioni di k_1 elementi estratti da A , pari a $\binom{n}{k_1} = \frac{n!}{k_1!(n-k_1)!} = \frac{n!}{k_1!k_2!}$. A questo punto non è difficile indovinare la formula nel caso generale: il numero di partizioni $\{A_1, \dots, A_r\}$ di A tali che $|A_1| = k_1, \dots, |A_r| = k_r$ (dove $k_1 + \dots + k_r = n$) è dato da

$$\frac{n!}{k_1! k_2! \dots k_r!}. \quad (1.16)$$

La dimostrazione è una applicazione del principio fondamentale del calcolo combinatorio. Una partizione $\{A_1, \dots, A_r\}$ può essere infatti determinata mediante le seguenti scelte successive:

- si sceglie A_1 come sottoinsieme di A di cardinalità k_1 , in $\binom{n}{k_1}$ modi possibili;
- si sceglie A_2 come sottoinsieme di $A \setminus A_1$ di cardinalità k_2 , in $\binom{n-k_1}{k_2}$ modi possibili;
- si procede analogamente fino alla scelta di A_{r-1} , per cui ci sono $\binom{n-(k_1+k_2+\dots+k_{r-2})}{k_{r-1}}$ possibilità;
- una volta scelto A_{r-1} , l'ultimo elemento della partizione è automaticamente determinato dalla relazione $A_r = A \setminus (A_1 \cup \dots \cup A_{r-1})$.

Applicando il Teorema 1.14, il numero di partizioni cercato è dato dunque da

$$\binom{n}{k_1} \cdot \binom{n-k_1}{k_2} \cdot \binom{n-(k_1+k_2)}{k_3} \dots \binom{n-(k_1+k_2+\dots+k_{r-2})}{k_{r-1}},$$

che non è altro che una riscrittura di (1.16), come si verifica dopo qualche semplificazione.

La quantità in (1.16), per $r \in \mathbb{N}$ e $n, k_1, \dots, k_r \in \mathbb{N}_0$ con $k_1 + \dots + k_r = n$, è detta *coefficiente multinomiale*: può essere infatti vista come un'estensione del coefficiente binomiale, a cui si riduce per $r = 2$. I coefficienti multinomiali ricorrono frequentemente nei cosiddetti problemi di occupazione. Un caso molto importante è descritto nell'esempio seguente.

Esempio 1.20. Supponiamo di disporre casualmente n palline numerate in r urne, dove $n \in \mathbb{N}_0$, $r \in \mathbb{N}$ (ogni urna può contenere un numero qualunque di palline). Chiediamoci qual è la probabilità che la prima urna contenga k_1 palline, la seconda ne contenga k_2 , ..., la r -esima ne contenga k_r , dove naturalmente $k_1, \dots, k_r \in \mathbb{N}_0$ sono tali che $k_1 + \dots + k_r = n$.

Lo spazio di probabilità naturale per questo esperimento aleatorio è l'insieme Ω delle disposizioni con ripetizione di n elementi presi dall'insieme $\{1, \dots, r\}$ (che rappresenta le urne), vale a dire

$$\Omega_{n,r} := \{f : \{1, \dots, n\} \rightarrow \{1, \dots, r\}\}, \quad (1.17)$$

munito della probabilità P uniforme, visto che la disposizione delle palline nelle urne avviene in modo completamente casuale. L'evento che ci interessa (che la prima urna contenga k_1 palline, la seconda ne contenga k_2 , ..., la r -esima ne contenga k_r) è rappresentato da

$$\mathcal{O}_{k_1, \dots, k_r} := \{f \in \Omega_{n,r} : |f^{-1}(1)| = k_1, \dots, |f^{-1}(r)| = k_r\}.$$

Essendo lo spazio di probabilità uniforme, dobbiamo calcolare la cardinalità di $\mathcal{O}_{k_1, \dots, k_r}$, cioè il numero di disposizioni di palline nelle urne tali che la prima urna ne contenga k_1 , la seconda ne contenga k_2 , ..., la r -esima ne contenga k_r . Un momento di riflessione mostra che la risposta è data proprio dal coefficiente multinomiale (1.16).[†] Di conseguenza, per ogni scelta di $k_1, \dots, k_r \in \mathbb{N}_0$ con $k_1 + \dots + k_r = n$, si ha

$$P(\mathcal{O}_{k_1, \dots, k_r}) = \frac{|\mathcal{O}_{k_1, \dots, k_r}|}{|\Omega_{n,r}|} = \frac{1}{r^n} \frac{n!}{k_1! \dots k_r!}, \quad (1.18)$$

poiché sappiamo che $|\Omega_{n,r}| = r^n$.

1.3.7 Le statistiche di Maxwell-Boltzmann, Bose-Einstein e Fermi-Dirac*

Nell'esempio 1.20 abbiamo parlato di n palline disposte in r urne, ma lo stesso schema si può applicare a contesti completamente diversi, come ad esempio:

- (1) gli esiti del lancio di n dadi regolari (o di n lanci di uno stesso dado regolare) a r facce;
- (2) la disposizione di n passeggeri su un treno composto da r vagoni;
- (3) la distribuzione di n incidenti in un certo periodo di r giorni;
- (4) le configurazioni di un sistema di n particelle, ciascuna delle quali può assumere r stati distinti.

[†]Infatti, indicando con $A := \{1, \dots, n\}$ l'insieme delle palline e con $A_i \subseteq A$ le palline contenute nell'urna i -esima, le disposizioni cercate sono in corrispondenza biunivoca con le partizioni $\{A_1, \dots, A_r\}$ di A tali che $|A_1| = k_1, \dots, |A_r| = k_r$.

In tutti questi casi, spesso non si è interessati a conoscere *quali*, ma soltanto *quante* “palline” ci siano in ciascuna “urna”. Se indichiamo con k_i il numero di “palline” nell’“urna” i -esima, ci si concentra cioè sui numeri $k_1, \dots, k_r \in \mathbb{N}_0$, che sono detti *numeri di occupazione*. Per descrivere esperimenti aleatori che coinvolgono tali numeri, risulta dunque naturale introdurre lo spazio campionario

$$\widehat{\Omega}_{n,r} := \{(k_1, \dots, k_r) \in (\mathbb{N}_0)^r : k_1 + \dots + k_r = n\}. \quad (1.19)$$

Per quanto visto nell’esempio 1.20, una probabilità P_{MB} “naturale” sullo spazio $\widehat{\Omega}_{n,r}$ è quella corrispondente alla densità p_{MB} definita da

$$p_{MB}((k_1, \dots, k_r)) = P_{MB}(\{(k_1, \dots, k_r)\}) := \frac{1}{r^n} \frac{n!}{k_1! \dots k_r!}, \quad (1.20)$$

in accordo con (1.18). Per ragioni storiche legate alla meccanica statistica, la probabilità P_{MB} è detta *statistica di Maxwell-Boltzmann* ed è efficacemente utilizzata in svariate situazioni, quali ad esempio quelle descritte nei punti (1), (2), (3) in alto. L’applicazione della probabilità P_{MB} alla situazione (4) è invece delicata e merita una discussione più approfondita.

Supponiamo che un sistema fisico sia composto da n particelle, dove ogni particella può assumere r stati distinti. Facciamo inoltre l’ipotesi che, qualunque siano gli stati delle n particelle, l’energia totale del sistema sia sempre la stessa. Allora, quando il sistema è in equilibrio termico, secondo i principi della meccanica statistica *tutte le configurazioni del sistema sono equiprobabili*.

Il punto fondamentale è decidere che cosa si debba intendere con *configurazione del sistema*. A priori sembra naturale identificare la configurazione del sistema con la n -upla degli stati assunti dalle particelle che lo compongono. In questa ottica, affermare che tutte le configurazioni sono equiprobabili significa considerare la probabilità uniforme sullo spazio $\Omega_{n,r}$ definito in (1.17), cioè sulle disposizioni con ripetizione di n elementi estratti dall’insieme $\{1, \dots, r\}$ (che rappresenta i possibili stati). Di conseguenza, se siamo interessati ai numeri di occupazione, cioè a quante (e non quali) particelle assumano un certo stato, la probabilità “giusta” sullo spazio $\widehat{\Omega}_{n,r}$ dei numeri di occupazione, definito in (1.19), sembrerebbe proprio essere la statistica di Maxwell-Boltzmann P_{MB} , introdotta in (1.20).

Tuttavia, se le n particelle che compongono il sistema sono *indistinguibili*, la statistica di Maxwell-Boltzmann conduce a predizioni errate. La ragione è che, in base ai principi della meccanica quantistica, la scelta dello spazio $\Omega_{n,r}$ definito in (1.17) per descrivere le configurazioni del sistema è fondamentalmente sbagliata: infatti, essendo le particelle indistinguibili, non esiste alcun esperimento che permetta di distinguere due elementi di $\Omega_{n,r}$ con gli stessi numeri di occupazione. Di conseguenza, le configurazioni del sistema devono essere *identificate* con le r -uple dei numeri di occupazione, cioè con gli elementi di $\widehat{\Omega}_{n,r}$: in altri termini, non ha fisicamente senso chiedersi *quali*, ma soltanto *quante* particelle assumano un certo stato. L’affermazione che tutte le configurazioni sono equiprobabili conduce dunque a munire $\widehat{\Omega}_{n,r}$ della *probabilità uniforme*. Tale probabilità, che indicheremo con P_{BE} , è detta *statistica di Bose-Einstein* ed è dunque definita dalla densità

$$p_{BE}((k_1, \dots, k_r)) = P_{BE}(\{(k_1, \dots, k_r)\}) := \frac{1}{|\widehat{\Omega}_{n,r}|}. \quad (1.21)$$

Gli esperimenti hanno effettivamente mostrato che la statistica di Bose-Einstein si applica con successo a sistemi composti da particelle (o atomi) indistinguibili a spin intero, che

vengono dette *bosoni* (un esempio è costituito dai fotoni). Per sistemi composti da particelle a spin semi-intero, dette *fermioni* (ne sono un esempio protoni, neutroni, elettroni, neutrini, ...), il principio di esclusione di Pauli impone l'ulteriore restrizione che *due particelle non possono assumere lo stesso stato*. Questo significa che i numeri di occupazione possono assumere soltanto i valori 0 e 1: in particolare, si deve avere $n \leq r$. Di conseguenza, le configurazioni del sistema sono descritte dall'insieme ristretto

$$\widehat{\Omega}'_{n,r} := \{(k_1, \dots, k_r) \in \{0, 1\}^r : k_1 + \dots + k_r = n\}. \quad (1.22)$$

Dovendo essere tutte le configurazioni equiprobabili, la probabilità “corretta” su $\widehat{\Omega}'_{n,r}$ è quella uniforme, detta *statistica di Fermi-Dirac* e indicata con P_{FD} :

$$p_{FD}((k_1, \dots, k_r)) = P_{FD}(\{(k_1, \dots, k_r)\}) := \frac{1}{|\widehat{\Omega}'_{n,r}|}. \quad (1.23)$$

Anche in questo caso, gli esperimenti hanno mostrato che la probabilità P_{FD} descrive correttamente il comportamento di sistemi composti da fermioni.

Osserviamo che i valori di $|\widehat{\Omega}_{n,r}|$ e $|\widehat{\Omega}'_{n,r}|$, che compaiono nelle definizioni (1.21) e (1.23), possono essere esplicitati:

$$|\widehat{\Omega}_{n,r}| = \binom{n+r-1}{n}, \quad |\widehat{\Omega}'_{n,r}| = \binom{r}{n}. \quad (1.24)$$

La seconda relazione segue immediatamente dal fatto che $\widehat{\Omega}'_{n,r}$ è in corrispondenza biunivoca naturale con l'insieme delle combinazioni di n elementi estratti da $\{1, \dots, r\}$: la corrispondenza è quella che a $(k_1, \dots, k_r) \in \widehat{\Omega}'_{n,r}$ associa il sottoinsieme degli indici $i \in \{1, \dots, r\}$ per cui $k_i = 1$ (si ricordi che $k_i \in \{0, 1\}$). Per dimostrare la prima relazione in (1.24), procediamo in modo analogo, mostrando che $\widehat{\Omega}_{n,r}$ è in corrispondenza biunivoca con l'insieme delle combinazioni di $(r-1)$ elementi estratti da $\{1, \dots, n+r-1\}$, che sappiamo avere cardinalità pari a $\binom{n+r-1}{r-1} = \binom{n+r-1}{n}$. Un sottoinsieme di $(r-1)$ elementi di $\{1, \dots, n+r-1\}$ può essere indicato con $\{t_1, \dots, t_{r-1}\}$, con $1 \leq t_1 < t_2 < \dots < t_{r-1} \leq n+r-1$. Se ora definiamo

$$k_1 := t_1 - 1, \quad k_i := t_i - t_{i-1} - 1, \quad \text{per } 1 < i < r, \quad k_r := (n+r-1) - t_{r-1},$$

si ha per costruzione $k_i \in \mathbb{N}_0$ e $k_1 + \dots + k_r = n$, cioè $(k_1, \dots, k_r) \in \widehat{\Omega}_{n,r}$. È facile vedere che la corrispondenza ora descritta è biunivoca: la corrispondenza inversa è quella che a (k_1, \dots, k_r) associa il sottoinsieme $\{t_1, \dots, t_{r-1}\}$ di $\{1, \dots, n+r-1\}$ definito da

$$t_1 := k_1 + 1, \quad t_i := t_{i-1} + k_i + 1, \quad \text{per } 2 \leq i \leq r-1.$$

Una rappresentazione grafica di questa corrispondenza si ottiene disegnando i numeri di occupazione mediante palline separate da sbarrette: per esempio, per $n = 7$ e $r = 4$, all'elemento $(1, 4, 0, 2) \in \widehat{\Omega}_{7,4}$ corrisponde la stringa “o|oooo|oo” composta da $n+r-1 = 10$ simboli. Le sbarrette identificano allora un sottoinsieme di $\{1, \dots, 10\}$ di cardinalità $r-1 = 4$.

Le statistiche di Bose-Einstein e Fermi-Dirac, definite dalle relazioni (1.21) e (1.23), sono dette *statistiche quantistiche*, mentre la statistica di Maxwell-Boltzmann, definita in (1.20), è talvolta indicata come *statistica classica*.

Esempio 1.21. L'analisi che ha portato all'introduzione delle statistiche di Bose-Einstein (1.21) e Fermi-Dirac (1.23) è basata sull'ipotesi che l'energia totale del sistema sia indipendente dagli stati delle particelle che lo compongono. Nel caso più realistico in cui l'energia del sistema non sia costante, si procede in modo del tutto analogo a quanto visto nell'Esempio 1.6: se Ω è lo spazio delle configurazioni del sistema, $H : \Omega \rightarrow \mathbb{R}$ è la funzione energia e $\beta \in [0, \infty)$ è l'inverso della temperatura assoluta, quando il sistema è in equilibrio termico alla temperatura β la probabilità di osservare il sistema in una configurazione $\omega \in \Omega$ è data da

$$P_\beta(\{\omega\}) := \frac{e^{-\beta H(\omega)}}{Z(\beta)}, \quad \text{dove} \quad Z(\beta) := \sum_{\omega \in \Omega} e^{-\beta H(\omega)}.$$

Il punto cruciale qui è la scelta dello spazio delle configurazioni Ω . Se il sistema è composto da n particelle *indistinguibili*, ciascuna delle quali può assumere r stati, si dovrà scegliere $\Omega = \widehat{\Omega}_{n,r}$ come in (1.19) se le particelle sono bosoni e $\Omega = \widehat{\Omega}'_{n,r}$ come in (1.22) se le particelle sono fermioni. Se invece le particelle sono distinguibili (per esempio se sono diverse tra loro) si sceglierà $\Omega = \Omega_{n,r}$ come in (1.17).

Come abbiamo già osservato nell'Esempio 1.6, per $\beta = 0$ (limite di temperatura infinita) la probabilità $P_\beta(\{\omega\})$ non dipende più da ω , dunque P_0 non è altro che la probabilità uniforme su Ω . Ritroviamo in questo modo le statistiche di Bose-Einstein ($P_0 = P_{BE}$ se $\Omega = \widehat{\Omega}_{n,r}$, cf. (1.21) e di Fermi-Dirac ($P_0 = P_{FD}$ se $\Omega = \widehat{\Omega}'_{n,r}$, cf. (1.23)). Sempre dall'Esempio 1.6 sappiamo che per $\beta \rightarrow \infty$ (limite di temperatura nulla) la probabilità P_β si concentra sull'insieme dei minimi assoluti della funzione H . Vedremo più avanti (cf. Esempio ??) che in questo regime un sistema di bosoni indistinguibili dà origine a un fenomeno molto interessante, noto come *condensazione di Bose-Einstein*, in cui gli effetti quantistici diventano visibili su grande scala.

Esempio 1.22. Si consideri un sistema di n particelle non interagenti, ciascuna delle quali può assumere r stati differenti, tutti con la stessa energia. Qual è la probabilità $p_{n,r}$ che gli stati assunti dalle particelle siano tutti diversi? Per ipotesi l'energia totale del sistema è indipendente dagli stati delle singole particelle, per cui si può applicare l'analisi svolta in questo paragrafo. Osserviamo inoltre che se $n > r$ si ha ovviamente $p_{n,r} = 0$, per cui supporremo d'ora in avanti che $n \leq r$.

In termini di numeri di occupazione (k_1, \dots, k_r) , dire che gli stati assunti dalle particelle siano tutti diversi significa imporre $k_i \in \{0, 1\}$ per ogni $1 \leq i \leq r$. In altri termini, $p_{n,r}$ è la probabilità dell'insieme $\widehat{\Omega}'_{n,r}$ definito in (1.22). Il valore di $p_{n,r}$ dipende naturalmente dal tipo di particelle considerate. Se le particelle sono fermioni indistinguibili, sappiamo che $\widehat{\Omega}'_{n,r}$ è l'intero spazio di configurazioni ammissibili, per cui ovviamente $p_{n,r}^{FD} = 1$, in accordo col principio di esclusione di Pauli. Se invece le n particelle sono bosoni indistinguibili, occorre usare la statistica di Bose-Einstein definita in (1.21), da cui si ottiene

$$p_{n,r}^{BE} = P_{BE}(\widehat{\Omega}'_{n,r}) = \frac{|\widehat{\Omega}'_{n,r}|}{|\widehat{\Omega}_{n,r}|} = \frac{\binom{r}{n}}{\binom{n+r-1}{n}} = \prod_{i=0}^{n-1} \frac{r-i}{r+i} = \prod_{i=0}^{n-1} \frac{1-\frac{i}{r}}{1+\frac{i}{r}}, \quad (1.25)$$

avendo usato le relazioni in (1.24). Infine, se le particelle sono distinguibili, usando la statistica di Maxwell-Boltzmann (1.20) si ottiene

$$p_{n,r}^{MB} = P_{MB}(\widehat{\Omega}'_{n,r}) = \sum_{(k_1, \dots, k_r) \in \widehat{\Omega}'_{n,r}} P_{MB}((k_1, \dots, k_r)) = \frac{n!}{r^n} |\widehat{\Omega}'_{n,r}| = \prod_{i=0}^{n-1} \left(1 - \frac{i}{r}\right), \quad (1.26)$$

dove si è usata la seconda relazione in (1.24) e il fatto che $p_{MB}((k_1, \dots, k_r)) = n!/r^n$ per ogni $(k_1, \dots, k_r) \in \widehat{\Omega}'_{n,r}$, poiché $k_i \in \{0, 1\}$ per ogni $1 \leq i \leq r$. Dalle relazioni (1.25) e (1.26) segue in particolare che $p_{n,r}^{BE} < p_{n,r}^{MB}$ per ogni $n \geq 1$.

Per avere un'idea più precisa dei risultati ottenuti, supponiamo che r sia grande e che $n = o(r^{2/3})$. Più precisamente, fissiamo un'arbitraria successione positiva $(\epsilon_r)_{r \in \mathbb{N}}$ tale che $\lim_{r \rightarrow \infty} \epsilon_r = 0$ e studiamo il comportamento asintotico di $p_{n,r}^{MB}$ e $p_{n,r}^{BE}$ nel limite $r \rightarrow \infty$, assumendo che $n \leq \epsilon_r r^{2/3}$. Dato che $(1+x) = \exp(x + O(x^2))$ per $x \rightarrow 0$, da (1.26) si ottiene

$$p_{n,r}^{MB} = \exp\left(-\sum_{i=0}^{n-1} \frac{i}{r} + \sum_{i=0}^{n-1} O\left(\frac{i^2}{r^2}\right)\right) = \exp\left(-\frac{1}{2} \frac{n^2}{r} + o(1)\right), \quad (1.27)$$

dove abbiamo usato il fatto che $\sum_{i=0}^{n-1} i = \frac{1}{2}n(n-1)$ e $\sum_{i=0}^{n-1} i^2 = O(n^3)$. Analogamente, da (1.25) si ottiene:

$$p_{n,r}^{BE} = \exp\left(-\frac{n^2}{r} + o(1)\right). \quad (1.28)$$

Le due relazioni (1.27) e (1.28) da un lato ci dicono che, sia per la statistica di Maxwell-Boltzmann sia per quella di Bose-Einstein, il regime “critico” per la probabilità $p_{n,r}$ è $n \approx \sqrt{r}$: in altri termini, $p_{n,r} \approx 1$ se $n \ll \sqrt{r}$ mentre $p_{n,r} \approx 0$ se $n \gg \sqrt{r}$. Dall'altro lato, se già sapevamo che $p_{n,r}^{BE} < p_{n,r}^{MB}$ per ogni $n \geq 1$, le relazioni (1.27) e (1.28) mostrano che nel regime $n \sim c\sqrt{r}$, con $c \in (0, \infty)$, la differenza tra $p_{n,r}^{MB}$ e $p_{n,r}^{BE}$ resta positiva anche nel limite di r grande: si ha in effetti $p_{n,r}^{BE} \rightarrow e^{-c}$ e $p_{n,r}^{MB} \rightarrow e^{-c/2}$.

Ricordiamo infine che la statistica di Maxwell-Boltzmann, adottata per descrivere un sistema di “particelle distinguibili”, si può applicare anche a sistemi “macroscopici” come quelli descritti nei punti (1), (2) e (3) all'inizio del paragrafo. In particolare, $p_{n,r}^{MB}$ coincide con la probabilità che, se n passeggeri salgono su un treno composto da r vagoni e si dispongono a caso, nessun vagone contenga più di un passeggero.

1.4 Probabilità condizionata

Nello studio di un modello probabilistico, risulta interessante studiare l'influenza che l'occorrere di un dato evento B ha sulla probabilità di occorrenza di un altro evento A .

Esempio 1.23. Nelle estrazioni per una ruota del Lotto, vengono estratte “a caso” 5 palline da un'urna contenente palline numerate da 1 a 90. Supponiamo di giocare due numeri su quella ruota, e precisamente l'1 e il 3. Una persona presente all'estrazione, mi avvisa che dei 5 numeri estratti 3 sono dispari. Qual è la probabilità di fare “ambo” sulla base di questa informazione? E qual è la probabilità in assenza di tale informazione?

È chiaro che la soluzione di tale problema richiede che si definisca il significato di calcolare una probabilità *sulla base* di una data informazione. Prima di proporre una definizione formale, cerchiamo una soluzione “ragionevole”. Lo spazio campionario in questione è $\Omega =$ “insieme di tutte le cinquine di numeri tra 1 e 90”. Assumendo l'equità dell'estrazione, scegliamo come probabilità P quella uniforme. Due eventi compaiono nell'enunciato del problema: $A =$ “i cinque numeri estratti contengono l'1 e il 3”, $B =$ “dei cinque numeri estratti 3 sono dispari”. In assenza dell'informazione sull'occorrenza di B , scriveremmo semplicemente

$$P(A) = \frac{|A|}{|\Omega|}.$$

Poichè la scelta di un elemento di A corrisponde alla scelta di tre numeri diversi da 1 e 3, si ha che $|A| = \binom{88}{3}$, e quindi

$$P(A) = \frac{\binom{88}{3}}{\binom{90}{5}} = \frac{20}{8010} \simeq 0.0025.$$

Assumere l'occorrenza di B significa escludere la possibilità che la cinquina estratta non sia in B . Inoltre, anche sapendo che la cinquina estratta è in B , non vi è alcun motivo per rimuovere l'ipotesi di equiprobabilità degli elementi di B . Dunque, la procedura "naturale" consiste nel rimpiazzare lo spazio campionario Ω con B , e calcolare le probabilità dei sottoinsiemi di B secondo la probabilità uniforme su B . Poichè A non è un sottoinsieme di B , si tratterà di calcolare la probabilità di $A \cap B$ secondo la probabilità uniforme su B . Concludiamo allora che l'oggetto più ragionevole per esprimere la probabilità di A *condizionata* all'occorrenza di B è

$$\frac{|A \cap B|}{|B|}.$$

Come esercizio, calcoliamo tale probabilità. Gli elementi di $A \cap B$ sono costituiti dalle cinque contenenti 1, 3, un altro numero dispari diversi da 1 e 3, e due numeri pari. Dunque

$$|A \cap B| = 43 \binom{45}{2}.$$

Inoltre, poichè gli elementi di B contengono 3 numeri dispari e 2 pari,

$$|B| = \binom{45}{3} \binom{45}{2}.$$

Infine

$$\frac{|A \cap B|}{|B|} = \frac{43}{\binom{45}{3}} = \frac{6}{1980} \simeq 0.003,$$

che è maggiore della probabilità in assenza di informazioni.

L'esempio appena trattato assieme all'osservazione che, se P è la probabilità uniforme

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

motiva la definizione che segue.

Definizione 1.24. Sia (Ω, P) uno spazio di probabilità discreto, A e B due eventi per cui $P(B) > 0$. La *probabilità di A condizionata a B* si denota con $P(A|B)$ ed è definita da

$$P(A|B) = \frac{P(A \cap B)}{P(B)}.$$

Alcune proprietà formali della probabilità condizionata sono sintetizzate nella seguente proposizione.

Proposizione 1.25. Sia B un evento fissato, con $P(B) > 0$, e consideriamo la funzione

$$\begin{aligned} \mathcal{P}(\Omega) &\longrightarrow [0, 1] \\ A &\longrightarrow P(A|B). \end{aligned}$$

Tale funzione è una probabilità su Ω .

La dimostrazione, che consiste nella verifica della validità degli assiomi (P1) e (P2), è lasciata per esercizio. Vale la pena sottolineare che, fissato un evento A , la funzione $B \mapsto P(A|B)$ non è una probabilità.

La seguente proposizione fornisce una caratterizzazione della probabilità condizionata che ne motiva ulteriormente la definizione.

Proposizione 1.26. *Sia B un evento fissato, con $P(B) > 0$. Allora $P(\cdot|B)$ è l'unica probabilità Q su Ω con le seguenti proprietà:*

(1) $Q(B) = 1$;

(2) per ogni coppia di eventi E, F con $P(F) > 0$ si ha $\frac{Q(E)}{Q(F)} = \frac{P(E)}{P(F)}$.

Dimostrazione. È immediato verificare che $P(\cdot|B)$ soddisfa le proprietà elencate. Viceversa, sia Q una probabilità che soddisfa le proprietà 1, 2 e sia A un evento arbitrario. Applicando la proprietà 2 agli eventi $E = A \cap B$ e $F = B$, visto che $Q(B) = 1$ per la proprietà 1 si ottiene la relazione

$$Q(A \cap B) = \frac{Q(A \cap B)}{Q(B)} = \frac{P(A \cap B)}{P(B)} =: P(A|B). \quad (1.29)$$

Osserviamo ora che possiamo scrivere

$$A = A \cap (B \cup B^c) = (A \cap B) \cup (A \cap B^c).$$

Dato che $A \cap B \subseteq B$ e $A \cap B^c \subseteq B^c$, gli eventi $A \cap B$ e $A \cap B^c$ sono disgiunti, quindi $Q(A) = Q(A \cap B) + Q(A \cap B^c)$. Per la proprietà 1 si ha $Q(B) = 1$, quindi $Q(A \cap B^c) \leq Q(B^c) = 0$ e di conseguenza $Q(A) = Q(A \cap B)$. Ricordando l'equazione (1.29), abbiamo dimostrato che $Q(A) = P(A|B)$, cioè Q coincide con la probabilità condizionata a B . \square

In molte situazioni, la nozione di probabilità condizionata è utile nella costruzione stessa di un modello probabilistico: talvolta è "naturale" assegnare il valore di alcune probabilità condizionate, e da esse dedurre il valore di probabilità non condizionate.

Esempio 1.27. Due urne contengono, rispettivamente, 3 palline rosse e 1 verde e 1 pallina rossa e 1 verde. Si sceglie, con ugual probabilità, una delle due urne e poi, dall'urna scelta, si estrae una pallina. Qual è la probabilità di estrarre una pallina rossa?

Denotiamo con a e b le due urne. Come spazio campionario, si può scegliere l'insieme costituito dalle coppie $(a, r), (a, v), (b, r), (b, v)$, dove la prima componente indica l'urna scelta e la seconda il colore della pallina estratta. L'evento $A = \{(a, r), (a, v)\}$ corrisponde a "l'urna scelta è la a ", l'evento $R = \{(a, r), (b, r)\}$ corrisponde a "la pallina estratta è rossa". Dev'essere senz'altro $P(A) = 1/2$, visto che le urne vengono scelte con uguale probabilità. Inoltre, supponendo di aver scelto l'urna a , la probabilità di estrarre una pallina rossa è $3/4$. Perciò porremo $P(R|A) = 3/4$. Analogamente $P(R|A^c) = 1/2$. Il procedimento per dedurre $P(R)$ dai dati a disposizione è indicato dal risultato che segue.

Proposizione 1.28. *Sia $(B_n)_{n=1}^N$ una sequenza di eventi finita ($N < +\infty$) o infinita ($N = +\infty$) tali che*

(a) Per ogni n

$$P(B_n) > 0.$$

(b) Gli eventi sono a due a due disgiunti, cioè

$$B_n \cap B_m = \emptyset$$

se $n \neq m$.

(c)

$$\bigcup_{n=1}^N B_n = \Omega.$$

Allora, per ogni evento A ,

$$P(A) = \sum_{n=1}^N P(A|B_n)P(B_n).$$

Tale identità prende il nome di formula delle probabilità totali.

Dimostrazione. Si osservi che

$$A = \bigcup_{n=1}^N (A \cap B_n),$$

e gli eventi di quest'ultima unione sono disgiunti. Usando l'additività di P e la definizione di probabilità condizionata, si ha

$$P(A) = \sum_{n=1}^N P(A \cap B_n) = \sum_{n=1}^N P(A|B_n)P(B_n). \quad \square$$

Supponiamo ora che A e B siano due eventi tali che $P(A) > 0$, $P(B) > 0$, sicché entrambe le probabilità condizionate $P(A|B)$ e $P(B|A)$ sono definite. È pressoché immediato verificare la seguente relazione.

Teorema 1.29. (*Formula di Bayes*) Se $P(A) > 0$ e $P(B) > 0$, allora

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)}. \quad (1.30)$$

Dimostrazione. La formula di Bayes (1.30) è equivalente a

$$P(B|A)P(A) = P(A|B)P(B),$$

che è vera in quanto entrambi i membri sono uguali a $P(A \cap B)$. □

Nell'ipotesi che $0 < P(B) < 1$, usando la formula delle probabilità totali, la formula di Bayes può essere riscritta nella forma

$$P(B|A) = \frac{P(A|B)P(B)}{P(A|B)P(B) + P(A|B^c)P(B^c)}. \quad (1.31)$$

Analogamente, se $(B_n)_{n=1}^N$ è una sequenza di eventi soddisfacenti alle ipotesi della Proposizione 1.28, si ha

$$P(B_n|A) = \frac{P(A|B_n)P(B_n)}{P(A)} = \frac{P(A|B_n)P(B_n)}{\sum_{k=1}^N P(A|B_k)P(B_k)}. \quad (1.32)$$

Le versioni (1.31) e (1.32) della formula di Bayes sono quelle che più spesso capita di usare negli esercizi.

La formula di Bayes, a dispetto della sua semplicità, è una delle formule fondamentali della Probabilità, ed è all'origine di un'intera area della Statistica, la *Statistica Bayesiana*. La rilevanza della formula di Bayes nelle applicazioni, si può già apprezzare in applicazioni semplici, come quella che segue.

Esempio 1.30. Per determinare la presenza di un certo virus viene elaborato un test clinico avente la seguente efficacia: se il virus è presente allora il test risulta positivo il 99% dei casi; se il virus è assente il test risulta positivo il 2% dei casi. E' noto che 2 persone su 10.000 hanno il virus. Supponiamo che un individuo scelto a caso risulti positivo al test. Con quale sicurezza possiamo affermare che sia malato?

Come accade sovente negli esercizi in cui si applica la formula di Bayes, non è rilevante descrivere nel dettaglio lo spazio campionario. Si considerino gli eventi, così descritti in modo informale: $A =$ "l'individuo è malato"; $B =$ "il test è risultato positivo". I dati del problema sono:

$$\begin{aligned} P(A) &= 0.0002 \\ P(B|A) &= 0.99 \\ P(B|A^c) &= 0.02. \end{aligned} \tag{1.33}$$

Calcoliamo $P(A|B)$. Utilizzando la formula di Bayes e la formula delle probabilità totali, si ha

$$P(A|B) = P(B|A) \frac{P(A)}{P(B)} = P(B|A) \frac{P(A)}{P(B|A)P(A) + P(B|A^c)P(A^c)} \simeq 0.01$$

che è estremamente bassa. Quindi, anche se un individuo risulta positivo al test, è molto improbabile che sia malato. Questo test dunque darà una grande percentuale di falsi positivi.

E se avessimo voluto specificare per bene lo spazio campionario? Si sarebbe potuto procedere così. Definiamo

$$\Omega = \{(m, p), (m, n), (s, p), (s, n)\} = \{m, s\} \times \{p, n\}$$

dove m e s indicano la presenza (m) o l'assenza del virus, p e n il risultato del test: $p =$ positivo, $n =$ negativo. Qual è la probabilità P su Ω ? Per individuare P dobbiamo usare i dati del problema. Si noti che gli eventi A e B definiti sopra, corrispondono ai seguenti sottoinsiemi di Ω :

$$\begin{aligned} A &= \{(m, p), (m, n)\} \\ B &= \{(m, p), (s, p)\}. \end{aligned}$$

Usando i dati in (1.33), è effettivamente possibile calcolare la probabilità di tutti i sottoinsiemi di Ω (provarci!), da cui si può dedurre il valore di $P(A|B)$. Tuttavia, per rispondere al quesito posto, questi dettagli sono poco rilevanti.

1.5 Indipendenza di eventi

Si è visto come la probabilità condizionata $P(A|B)$ rappresenti la probabilità dell'evento A sotto la condizione del verificarsi dell'evento B . E' possibile che tale condizione non modifichi la probabilità di A , ossia

$$P(A|B) = P(A). \tag{1.34}$$

Usando la definizione di probabilità condizionata, si vede che l'identità (1.34) equivale a:

$$P(A \cap B) = P(A)P(B). \quad (1.35)$$

L'identità in (1.35), rispetto a quella in (1.34) ha il vantaggio di essere esplicitamente simmetrica in A e B , e di essere definita (e banalmente vera) anche quando $P(B) = 0$. Essa viene dunque scelta per caratterizzare la nozione di *indipendenza*.

Definizione 1.31. In uno spazio di probabilità discreto (Ω, P) , due eventi A e B si dicono *indipendenti* se

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

Esempio 1.32. Da due mazzi di carte da Poker si estraggono due carte, una per mazzo. Lo spazio campionario naturale è l'insieme delle coppie (i, j) nel prodotto cartesiano $\Omega = X \times X$, dove X è l'insieme delle carte di un mazzo. Possiamo assumere che la scelta sia "casuale", cioè descritta dalla probabilità P uniforme su Ω . Consideriamo due eventi A e B , di cui l'evento A dipende solo dall'estrazione dal primo mazzo, l'evento B solo dall'estrazione dal secondo mazzo. In altre parole, se $F, G \subseteq X$, A e B sono della forma:

$$\begin{aligned} A &= \{(i, j) \in \Omega : i \in F\} \\ B &= \{(i, j) \in \Omega : j \in G\}. \end{aligned}$$

Si noti che $|\Omega| = 52^2$, $|A| = 52|F|$, $|B| = 52|G|$, $|A \cap B| = |F||G|$. Ne segue facilmente che

$$P(A \cap B) = \frac{|A \cap B|}{|\Omega|} = \frac{|F||G|}{52^2} = \frac{|A|}{|\Omega|} \frac{|B|}{|\Omega|} = P(A)P(B).$$

Dunque A e B sono indipendenti. Notare che gli eventi A e B si riferiscono a due *ripetizioni* dello stesso esperimento aleatorio. L'indipendenza esprime il fatto che l'esito di un esperimento non "influenza" l'esito dell'altro esperimento. Questo contesto di *prove indipendenti ripetute*, rilevante in molti aspetti della Probabilità e della Statistica, è quello in cui la nozione di indipendenza appare in modo naturale.

L'esempio 1.32 si può facilmente generalizzare al caso di 3 o più mazzi di carte. In questo caso, P è la probabilità uniforme su $\Omega = X^n$, $F_1, F_2, \dots, F_n \subseteq X$ e, per $i = 1, 2, \dots, n$, $A_i = \{\omega = (\omega_1, \omega_2, \dots, \omega_n) \in X^n : \omega_i \in F_i\}$. Una semplice generalizzazione dell'argomento visto sopra, mostra che

$$P(A_1 \cap A_2 \cap \dots \cap A_n) = \prod_{i=1}^n P(A_i). \quad (1.36)$$

Naturalmente se avessimo considerato solo alcuni degli A_i , per esempio la coppia A_i, A_j con $i \neq j$, lo stesso facile calcolo sulle cardinalità avrebbe mostrato che

$$P(A_i \cap A_j) = P(A_i)P(A_j).$$

Tutto ciò suggerisce che, nell'estendere la nozione di indipendenza a tre o più eventi, dobbiamo richiedere la proprietà "moltiplicativa" (1.36), ma anche il fatto che se una famiglia di eventi è costituita da eventi indipendenti, anche ogni sua sottofamiglia è costituita da eventi indipendenti. Queste due richieste, non sono implicate l'una dall'altra, come mostrano i seguenti esempi.

Esempio 1.33. Sia $\Omega = \{1, 2, 3, 4\}$, $P =$ Probabilità uniforme, $A = \{1, 2\}$, $B = \{2, 3\}$, $C = \{1, 3\}$. Si vede immediatamente che le coppie (A, B) , (B, C) e (A, C) sono formate da eventi indipendenti. Tuttavia:

$$P(A \cap B \cap C) = P(\emptyset) = 0 \neq P(A)P(B)P(C) = 1/8.$$

Esempio 1.34. Sia $\Omega = \{1, 2, \dots, 6\}^2$, con la probabilità uniforme. Siano

$$\begin{aligned} A &= \{(i, j) : j = 1, 2, \text{ o } 5\} \\ B &= \{(i, j) : j = 4, 5, \text{ o } 6\} \\ C &= \{(i, j) : i + j = 9\}. \end{aligned}$$

Si ha

$$\begin{aligned} P(A \cap B) &= \frac{1}{6} \neq \frac{1}{4} = P(A)P(B) \\ P(A \cap C) &= \frac{1}{36} \neq \frac{1}{18} = P(A)P(C) \\ P(B \cap C) &= \frac{1}{12} \neq \frac{1}{18} = P(B)P(C) \end{aligned}$$

ma

$$P(A \cap B \cap C) = \frac{1}{36} = P(A)P(B)P(C).$$

Definizione 1.35. Sia I un qualunque insieme di indici, e sia $\{A_i : i \in I\}$ una famiglia di eventi in uno spazio di probabilità discreto (Ω, P) . Diremo che tali eventi sono *indipendenti* se per ogni sottoinsieme finito J di I , si ha

$$P\left(\bigcap_{j \in J} A_j\right) = \prod_{j \in J} P(A_j).$$

La proposizione che segue afferma che se in una famiglia di eventi indipendenti si rimpiazzano alcuni eventi con i loro complementari, si ottiene ancora una famiglia di eventi indipendenti.

Proposizione 1.36. Sia $\{A_i : i \in I\}$ una famiglia di eventi indipendenti, $I' \subseteq I$, e definiamo

$$B_i = \begin{cases} A_i^c & \text{se } i \in I' \\ A_i & \text{se } i \in I \setminus I'. \end{cases}$$

Allora $\{B_i : i \in I\}$ è una famiglia di eventi indipendenti.

Dimostrazione. Sia $J \subset I$ finito, e sia $J' = J \cap I'$. Dobbiamo mostrare che

$$P\left(\bigcap_{j \in J} B_j\right) = \prod_{j \in J} P(B_j). \quad (1.37)$$

Possiamo supporre che

$$J = \{j_1, j_2, \dots, j_m\}, \quad J' = \{j_1, \dots, j_k\},$$

dove $k \leq m$. Se $k = 0$, la (1.37) segue immediatamente dall'indipendenza di $\{A_i : i \in I\}$. Supponiamo $k = 1$. Usando l'indipendenza di $\{A_i : i \in I\}$, si ha

$$\begin{aligned}
 P(B_{j_1} \cap B_{j_2} \cap \cdots \cap B_{j_k}) &= P(A_{j_1}^c \cap A_{j_2} \cap \cdots \cap A_{j_k}) \\
 &= P([A_{j_2} \cap \cdots \cap A_{j_k}] \setminus [A_{j_1} \cap A_{j_2} \cap \cdots \cap A_{j_k}]) \\
 &= P(A_{j_2}) \cdots P(A_{j_k}) - P(A_{j_1})P(A_{j_2}) \cdots P(A_{j_k}) \\
 &= [1 - P(A_{j_1})]P(A_{j_2}) \cdots P(A_{j_k}) \\
 &= P(A_{j_1}^c)P(A_{j_2}) \cdots P(A_{j_k}) \\
 &= P(B_{j_1})P(B_{j_2}) \cdots P(B_{j_k}).
 \end{aligned}$$

A questo punto si procede per induzione su k , per trattare tutti i casi $0 \leq k \leq m$. Si lasciano i semplici dettagli al lettore. \square

Nel prossimo esempio vedremo un calcolo classico basato sull'indipendenza.

Esempio 1.37. Si eseguono N prove ripetute di un gioco in cui la probabilità di vincere è $p \in [0, 1]$. Si assuma che i risultati di prove distinte siano indipendenti.

- a. Qual è la probabilità di vincere $n \leq N$ volte sugli N tentativi?
 - b. Qual è la probabilità di vincere per la prima volta all' n -esimo tentativo?
- Come spazio campionario possiamo scegliere

$$\Omega = \{x = (x_1, x_2, \dots, x_N) : x_i \in \{0, 1\}\} = \{0, 1\}^N,$$

dove $x_i = 1$ significa che l' i -esima prova del gioco è stata vinta. Definiamo per $t \in \{0, 1\}$ l'evento

$$A_i(t) = \{x \in \Omega : x_i = t\}.$$

L'indipendenza delle prove effettuate si traduce nel fatto che gli eventi $A_1(t_1), A_2(t_2), \dots, A_N(t_N)$ sono indipendenti, per ogni scelta di t_1, \dots, t_N . Inoltre è chiaro che per ogni $y \in \Omega$

$$\{y\} = \bigcap_{i=1}^N A_i(y_i).$$

Per l'indipendenza, allora,

$$P(\{y\}) = \prod_{i=1}^N P(A_i(y_i)).$$

Inoltre, poichè per ogni tentativo la probabilità di vincere è p ,

$$P(A_i(t)) = \begin{cases} p & \text{se } t = 1 \\ 1 - p & \text{se } t = 0 \end{cases}$$

o, equivalentemente, $P(A_i(t)) = p^t(1 - p)^{1-t}$. Ne segue allora che

$$P(\{y\}) = \prod_{i=1}^N p^{y_i}(1 - p)^{1-y_i} = p^{\sum_{i=1}^N y_i} (1 - p)^{N - \sum_{i=1}^N y_i}.$$

Gli eventi di cui vogliamo calcolare la probabilità sono

$$A = \{y \in \Omega : \sum_{i=1}^N y_i = n\}$$

e

$$B = A_1(0) \cap \dots \cap A_{n-1}(0) \cap A_n(1).$$

Da quanto visto sopra,

$$\begin{aligned} P(A) &= \sum_{y \in A} P(\{y\}) \\ &= \sum_{y \in A} p^n (1-p)^{N-n} \\ &= |A| p^n (1-p)^{N-n} \\ &= \binom{N}{n} p^n (1-p)^{N-n}, \end{aligned}$$

dove abbiamo usato il fatto che scegliere un elemento di A equivale a scegliere gli n tentativi vincenti sugli N a disposizione. Si noti che l'espressione ottenuta per la probabilità di A è analoga a quella trovata nell'esempio 1.19 per lo schema di estrazioni con reimmissione, con $p := \frac{m}{N}$. Infatti lo schema di estrazioni con reimmissione è un esempio di prove ripetute e indipendenti.

Più facile è calcolare la probabilità di B :

$$P(B) = P(A_1(0)) \cdot \dots \cdot P(A_{n-1}(0)) P(A_n(1)) = (1-p)^{n-1} p.$$

I due esempi che seguono, liberamente tratti da casi giudiziari reali, mostrano quanto la non comprensione della nozione di condizionamento possa condurre a conclusioni errate.

Esempio 1.38. Le indagini relative ad un omicidio hanno condotto alle seguenti conclusioni:

- il colpevole possiede un determinato set di caratteristiche (individuate da un testimone, per es. capelli rossi, zoppicante, automobile verde, ecc.) che lo rendono piuttosto raro: si stima che una frazione $p \ll 1$ di popolazione possenga tali caratteristiche;
- il colpevole risiede in una città di n abitanti; inoltre $np \simeq 0.05$.

Una ricerca su un database di individui schedati ha identificato un (unico) individuo che possiede i due requisiti sopra citati. Per affermarne la colpevolezza, l'accusa argomenta come segue.

- (1) La probabilità che nella città vi siano almeno due individui con le caratteristiche individuate dal testimone è circa $(np)^2 = 0.0025$.
- (2) Pertanto con probabilità $1 - (np)^2 = 0.9975$ quello trovato è l'unico individuo con tali caratteristiche: la sua colpevolezza è dunque accertata con probabilità 0.9975.

Questo argomento è sbagliato, in particolare la conclusione. Vediamo perché. Una ragionevole assunzione è che ogni individuo possenga la caratteristica individuata con probabilità p ,

indipendentemente dagli altri. Pertanto, per quanto visto nell'Esempio 1.37, se A_k denota l'evento "in città ci sono k individui con quella caratteristica, si ha

$$P(A_k) = \binom{n}{k} p^k (1-p)^{n-k}.$$

In particolare

$$P(A_0) = (1-p)^n \quad P(A_1) = np(1-p)^{n-1}.$$

Quindi la probabilità che vi siano almeno due individui con i requisiti richiesti è

$$P[(A_0 \cup A_1)^c] = 1 - (1-p)^n - np(1-p)^{n-1} \simeq \frac{n(n-1)}{2} p^2 \simeq \frac{1}{2} (np)^2,$$

dove la prima approssimazione si ottiene con uno sviluppo di Taylor al secondo ordine della funzione $f(p) = 1 - (1-p)^n - np(1-p)^{n-1}$ attorno a $p = 0$. Il punto 1. dell'argomento dell'accusa va dunque corretto di un fattore $1/2$. Ma non è certo questo l'errore più rilevante, anche perché la correzione rende l'argomento ancor più stringente. L'accusa infatti non ha tenuto conto del fatto che un individuo con le caratteristiche individuate è *già stato trovato*. La quantità probante non "è dunque la "probabilità che vi siano almeno due individui con i requisiti richiesti", bensì la "probabilità che ve ne siano almeno due *condizionata* alla conoscenza che la ricerca ha identificato un individuo che li possiede. Consideriamo quindi gli eventi $B =$ "vi sono almeno due individui con i requisiti richiesti" e $C =$ "la ricerca ha identificato esattamente un individuo che li possiede". Se k è il numero di residenti della città inseriti nel database, e assumiamo che la frequenza delle caratteristiche in questione tra gli schedati sia la stessa che nel resto della popolazione. Allora

$$P(C) = kp(1-p)^{k-1}.$$

Inoltre, $B \cap C = B' \cap C$ dove $B' =$ "tra gli individui *non* schedati almeno uno ha le caratteristiche richieste". Inoltre gli eventi B' e C , riferendosi a gruppi distinti di individui, sono indipendenti. Pertanto

$$P(B \cap C) = P(B' \cap C) = P(B')P(C) = [1 - (1-p)^{n-k}] P(C),$$

da cui

$$P(B|C) = 1 - (1-p)^{n-k} \simeq (n-k)p \simeq np = 0.05$$

almeno nel caso, verosimile, che $k \ll n$. Dunque, sulla base delle conoscenze acquisite, la probabilità che l'individuo trovato sia l'unico con le caratteristiche date, e quindi che sia colpevole, è

$$P(B^c|C) = 0.95,$$

probabilmente non sufficiente a fugare "ogni ragionevole dubbio".

Per apprezzare la sottigliezza della questione, consideriamo il seguente quesito. Immaginiamo di avere l'informazione del testimone, ma di non aver condotto alcuna ricerca su database. Qual è la probabilità che il colpevole sia l'unico individuo con le caratteristiche richieste? In questo caso, l'unica informazione disponibile è che "esiste almeno un individuo con le caratteristiche richieste", che corrisponde all'evento A_0^c . Osservando che $B = (A_0 \cup A_1)^c$,

$$P(C|A_0^c) = P[(A_0 \cup A_1)^c | A_0^c] = \frac{P[(A_0 \cup A_1)^c]}{P(A_0^c)}.$$

Usando la precedente stima per $P[(A_0 \cup A_1)^c]$ e l'analoga stima

$$P(A_0^c) = 1 - (1 - p)^n \simeq np,$$

otteniamo

$$P[(A_0 \cup A_1)^c | A_0^c] \simeq \frac{1}{2} np$$

che differisce di un fattore $\frac{1}{2}$ dal risultato trovato prima!

Esempio 1.39. Una donna venne assassinata, il marito era il principale sospettato. Nel corso delle indagini si scoprì che il marito aveva più volte picchiato la moglie. L'accusa affermò che questo rappresentasse un importante indizio per la colpevolezza. La difesa ribatté che, secondo i dati forniti dalla Polizia di Stato, tra gli uomini che picchiano le loro mogli, solo 1 su 10000 finisce poi per assassinarla. Pertanto tale dato contribuisce solo in modo molto marginale alla tesi di colpevolezza. In primo grado il giudice accolse la tesi della difesa.

In secondo grado l'accusa (ci piace pensare con l'aiuto di un matematico) rilevò il seguente errore nell'argomento della difesa. La frazione $\frac{1}{10000}$ fornisce una stima della probabilità che una donna venga ammazzata dal marito *condizionata* al fatto che il marito la picchiasse. Ma noi non solo sappiamo che il marito la picchiava, ma anche che la donna è stata effettivamente assassinata. Quindi la probabilità che utilizza l'intera informazione disponibile è: probabilità che "una donna venga ammazzata dal marito" *condizionata* a "la donna veniva picchiata dal marito ed è stata ammazzata".

Per formalizzare il problema, consideriamo una popolazione numerosa e sufficientemente omogenea di donne sposate, e consideriamo i seguenti eventi, relativi ad un individuo casualmente scelto in questa popolazione:

- A = "la donna viene assassinata"
- B = "la donna è stata picchiata dal marito"
- C = "la donna viene assassinata dal marito".

Possiamo assumere che $P(C|B) = \frac{1}{10000}$. Inoltre i dati della Polizia indicano che, nella totalità della popolazione, circa una donna su 100000 viene assassinata, cioè possiamo assumere $P(A) = \frac{1}{100000}$. La probabilità che desideriamo calcolare è $P(C|A \cap B)$. Usando il fatto che $C \subseteq A$ otteniamo

$$P(C|A \cap B) = \frac{P(A \cap B \cap C)}{P(A \cap B)} = \frac{P(B \cap C)}{P(A \cap B)} = \frac{P(C|B)}{P(A|B)}.$$

I dati a disposizione non ci consentono di calcolare $P(A|B)$. Si noti che la conoscenza che una donna viene picchiata dal marito, rende assai più probabile il suo assassinio da parte del marito. È quanto meno plausibile assumere che tutte le donne, vengano o meno picchiate dal marito, hanno grosso modo la stessa probabilità di essere assassinate da *una persona diversa dal marito*, cioè

$$P(A \setminus C|B) = P(A \setminus C);$$

in altre parole, $A \setminus C$ e B sono indipendenti. Abbiamo allora

$$\begin{aligned} P(C|A \cap B) &= \frac{P(C|B)}{P(A|B)} = \frac{P(C|B)}{P(C|B) + P(A \setminus C|B)} = \frac{P(C|B)}{P(C|B) + P(A \setminus C)} \\ &\geq \frac{P(C|B)}{P(C|B) + P(A)} = \frac{\frac{1}{10000}}{\frac{1}{10000} + \frac{1}{100000}} = \frac{10}{11}. \end{aligned}$$

Quindi, la sola informazione che avesse picchiato la moglie, rende la probabilità che l'imputato sia colpevole almeno $\frac{10}{11}$! L'imputato è stato poi condannato...

Capitolo 2

Esempi rilevanti di probabilità discrete

In questo capitolo vedremo all'opera le nozioni di probabilità viste nel capitolo 1, applicate ad alcuni esempi rilevanti e non banali. Con l'eccezione di un'osservazione non essenziale nel sottoparagrafo 2.1.1, la nozione di indipendenza (paragrafo 1.5) non è necessaria per i paragrafi 2.1, 2.2 e 2.3, mentre viene usata nel paragrafo 2.4. Infine, il contenuto del paragrafo 2.5, incluso in questo capitolo per omogeneità di argomento, è di carattere più avanzato e richiede il concetto di variabile casuale, sviluppato nel capitolo 3.

2.1 Permutazioni aleatorie

Come già visto nel capitolo 1, denotiamo con S_n l'insieme delle funzioni biettive dall'insieme $\{1, 2, \dots, n\}$ in sé. S_n è un gruppo non commutativo (per $n \geq 3$) se dotato dell'operazione di composizione. In tutti gli esempi che vedremo, assumeremo che la probabilità P su S_n sia quella uniforme. Naturalmente anche P dipende da n , ma non c'è ragione di appesantire le notazioni chiamandola, ad esempio, P_n . Lo spazio di probabilità (S_n, P) è un buon modello per l'esperimento aleatorio che consiste nel *mescolare* accuratamente n oggetti, e quindi osservare l'ordinamento ottenuto. In questo paragrafo esaminiamo alcune proprietà interessanti dello spazio (S_n, P) , prendendo spunto da alcuni problemi.

2.1.1 Cicli

Problema 2.1. Un gruppo di n amici affitta una casa per una vacanza. Dopo alcuni giorni tutti convengono che sia il caso di fare delle pulizie, ma si stenta a trovare dei volontari. Laura, che è volenterosa e bizzarra, avanza la seguente proposta. Ognuno scrive il proprio nome su una carta. Quindi le n carte vengono accuratamente mescolate e distribuite. Laura allora leggerà ad alta voce il nome sulla sua carta. Quindi la persona il cui nome è stato letto leggerà a sua volta il nome sulla sua carta; si prosegue così finché non viene letto il nome di Laura. A questo punto, le persone il cui nome è stato chiamato formeranno la squadra per le pulizie.

- (i) Qual è la probabilità che Laura si trovi a dover fare le pulizie da sola?
- (ii) Qual è la probabilità che tutti debbano fare le pulizie?

(iii) Più in generale, qual è la probabilità che la squadra delle pulizie sia composta da m persone?

Soluzione. È conveniente riformulare il problema con un linguaggio più formale. Etichettiamo gli n amici con i numeri $1, 2, \dots, n$, assumendo che il numero 1 corrisponda a Laura. L'esito del mescolamento delle n carte può essere descritto da un elemento $\sigma \in S_n$: la carta in mano alla persona i ha il nome della persona $\sigma(i)$.

La squadra per le pulizie si ottiene applicando *ripetutamente* σ a 1:

$$\sigma(1), \sigma \circ \sigma(1) =: \sigma^2(1), \dots, \sigma^{k-1}(1), 1,$$

dove $k \geq 1$ è il più piccolo numero intero tale che $\sigma^k(1) = 1$. La sequenza

$$(1, \sigma(1), \sigma^2(1), \dots, \sigma^{k-1}(1))$$

viene detta *ciclo* di *lunghezza* k . La costruzione fatta a partire dall'elemento 1 può essere ripetuta a partire da un elemento arbitrario. È chiaro che ogni elemento $i \in \{1, 2, \dots, n\}$ appartiene ad uno ed un solo ciclo: in altre parole una permutazione individua una partizione in cicli di $\{1, 2, \dots, n\}$. Il quesito (iii) del problema in esame, che contiene gli altri due come casi particolari, può essere pertanto riformulato come segue: qual è la probabilità che il ciclo contenente 1 abbia lunghezza m ?

Definiamo

$$C_m := \{\sigma \in S_n : \text{il ciclo contenente 1 ha lunghezza } m\}.$$

Si noti che $C_1 = \{\sigma \in S_n : \sigma(1) = 1\}$, e che c'è una naturale corrispondenza biunivoca tra C_1 e l'insieme delle permutazioni di $\{2, 3, \dots, n\}$, da cui si deduce che

$$|C_1| = (n-1)! \Rightarrow P(C_1) = \frac{1}{n}.$$

In altre parole, la probabilità che Laura si trovi da sola a fare le pulizie è pari a $\frac{1}{n}$. Questo risponde alla domanda (i).

Consideriamo ora la domanda (ii), cioè calcoliamo $P(C_n)$. Per contare gli elementi di C_n osserviamo che se $\sigma \in C_n$ essa ha un unico ciclo, che si può rappresentare nella forma

$$(1, \sigma(1), \sigma^2(1), \dots, \sigma^{n-1}(1)).$$

Ma la scrittura precedente si può *interpretare* come una permutazione di $\{1, 2, \dots, n\}$ con 1 al primo posto. Abbiamo appena notato che ci sono $(n-1)!$ tali permutazioni, per cui $|C_n| = (n-1)!$. Segue in particolare che $P(C_n) = \frac{1}{n}$, che risponde alla domanda (ii).

In modo più rigoroso, mostriamo che l'applicazione φ definita da

$$(\varphi(\sigma))(k) := \sigma^{k-1}(1),$$

dove si intende che σ^0 sia la funzione identica, è una biiezione tra C_n e C_1 . È evidente che $\varphi(\sigma) \in C_1$ per ogni $\sigma \in C_n$ e che l'applicazione φ è iniettiva. Per mostrare che φ è suriettiva, e dunque biiettiva, basta mostrare che ammette inversa destra, cioè che esiste un'applicazione $\psi : C_1 \rightarrow C_n$ tale che $\varphi \circ \psi = \text{identità su } C_1$. Mostriamo che tale ψ è data da

$$(\psi(\tau))(k) := \tau(\tau^{-1}(k) + 1), \quad \text{dove } n+1 := 1,$$

cioè si ha $\varphi(\psi(\tau)) = \tau$ per ogni $\tau \in C_1$. Per verificare quest'ultimo fatto, per definizione

$$\varphi(\psi(\tau))(1) = (\psi(\tau))^0(1) = 1 = \tau(1),$$

dato che $\tau \in C_1$. Inoltre, assumendo che per $m \leq k$ si abbia $\varphi(\psi(\tau))(m) = \tau(m)$, si ha

$$\begin{aligned} \varphi(\psi(\tau))(k+1) &= (\psi(\tau))^k(1) = (\psi(\tau))[(\psi(\tau))^{k-1}(1)] \\ &= (\psi(\tau))[\varphi(\psi(\tau))(k)] = (\psi(\tau))(\tau(k)) = \tau(\tau^{-1}(\tau(k)) + 1) = \tau(k+1). \end{aligned}$$

Abbiamo dunque mostrato per induzione che $\varphi(\psi(\tau))(k) = \tau(k)$ per ogni $k \in \{1, 2, \dots, n\}$.

A questo punto abbiamo gli strumenti per calcolare $|C_m|$ per ogni valore di m , cioè per rispondere alla domanda (iii). Infatti, gli elementi $\sigma \in C_m$ possono essere determinati dalle seguenti *tre scelte successive*:

- si scelgono gli m elementi del ciclo contenente 1, per cui ci sono $\binom{n-1}{m-1}$ esiti possibili (uno degli m elementi dev'essere 1);
- si sceglie uno dei cicli formati da questi m elementi: come abbiamo appena visto nella risposta alla domanda (ii), ci sono $(m-1)!$ tali cicli;
- si scelgono i valori di σ sui rimanenti $n-m$ elementi: dato che σ permuta in modo arbitrario tali elementi, per questa scelta ci sono $(n-m)!$ esiti possibili.

Per il principio fondamentale del calcolo combinatorio, si ottiene

$$|C_m| = \binom{n-1}{m-1} (m-1)! (n-m)! = (n-1)! \Rightarrow P(C_m) = \frac{1}{n}.$$

Concludendo, la probabilità che la squadra per le pulizie sia composta da m elementi è $\frac{1}{n}$, in particolare non dipende da m . \square

Avendo acquistato un po' di familiarità con le permutazioni, consideriamo il problema seguente.

Problema 2.2. Lo stesso gruppo di n amici decide di usare il metodo proposto da Laura per suddividersi in sottogruppi, corrispondenti alla partizione in cicli determinata della permutazione.

- Qual è la probabilità che si formi un sottogruppo, necessariamente unico, con più di $n/2$ persone?

Soluzione. Sia $m > \frac{n}{2}$, e sia

$$D_m := \{\sigma \in S_n : \sigma \text{ ha un ciclo di lunghezza } m\}.$$

In realtà D_m è ben definito anche per $m \leq n/2$. Tuttavia in questo caso un ciclo di lunghezza m non è necessariamente unico, e l'argomento che vedremo ora per determinare il numero dei suoi elementi non si può applicare. Assumiamo perciò $m > n/2$. Gli elementi di D_m possono essere determinati attraverso le seguenti scelte successive:

- si scelgono gli m elementi che compaiono nel ciclo "grande", per cui ci sono $\binom{n}{m}$ esiti possibili;
- si sceglie il ciclo sugli m elementi fissati: per questa scelta ci sono $(m-1)!$ esiti possibili;

- si permutano in modo arbitrario i rimanenti $n - m$ elementi, per cui ci sono $(n - m)!$ esiti possibili.

Pertanto

$$|D_m| = \binom{n}{m} (m - 1)! (n - m)! = \frac{n!}{m} \Rightarrow P(D_m) = \frac{1}{m}.$$

Si osservi che se fosse $m \leq \frac{n}{2}$, la possibile non unicità dei cicli di lunghezza m conduce a “contare più di una volta” la stessa permutazione, e quindi il precedente conteggio risulta non corretto: questo si evince anche dal fatto che $\sum_{m=1}^n \frac{1}{m} > 1$ per ogni $n \in \mathbb{N}$.

Per rispondere al quesito del problema, dobbiamo calcolare

$$P \left(\bigcup_{\frac{n}{2} < m \leq n} D_m \right) = \sum_{\frac{n}{2} < m \leq n} P(D_m) = \sum_{\frac{n}{2} < m \leq n} \frac{1}{m}.$$

Denotiamo con p_n quest’ultima probabilità, e sia $\lfloor \frac{n}{2} \rfloor$ la parte intera di $\frac{n}{2}$. Intuitivamente, sostituendo la somma con un integrale si ottiene $p_n \approx \int_{\lfloor n/2 \rfloor}^n \frac{1}{x} dx \approx \log n - \log(n/2) = \log 2$. Precisiamo ora questa relazione in modo rigoroso. Usando le disuguaglianze (da verificare per esercizio!), valide per ogni $x > 0$

$$0 \leq x - \log(1 + x) \leq x^2,$$

abbiamo che, se $m > \lfloor \frac{n}{2} \rfloor$

$$0 \leq \frac{1}{m} - \log \left(\frac{m+1}{m} \right) \leq \frac{1}{m^2} \leq \frac{4}{n^2}.$$

Perciò

$$0 \leq \sum_{\frac{n}{2} < m \leq n} \frac{1}{m} - \sum_{\frac{n}{2} < m \leq n} \log \left(\frac{m+1}{m} \right) \leq \frac{2}{n},$$

da cui, essendo

$$\sum_{\frac{n}{2} < m \leq n} \log \left(\frac{m+1}{m} \right) = \log \left(\frac{n+1}{\lfloor \frac{n}{2} \rfloor + 1} \right),$$

si deduce che

$$\log \left(\frac{n+1}{\lfloor \frac{n}{2} \rfloor + 1} \right) \leq p_n \leq \log \left(\frac{n+1}{\lfloor \frac{n}{2} \rfloor + 1} \right) + \frac{2}{n}.$$

In particolare

$$\lim_{n \rightarrow +\infty} p_n = \log 2.$$

In altre parole, per grandi valori di n , la probabilità che si formi un sottogruppo con più di $n/2$ persone è approssimativamente $\log 2 \simeq 0,693147181$ e dunque (approssimativamente) *non dipende da n* , un risultato non evidente a priori. Per $n > 50$, $|p_n - \log 2| \leq 0.01$. \square

Il risultato appena ottenuto permette di trovare una soluzione al seguente difficile problema.

Problema 2.3. Il docente di un corso di Probabilità frequentato da 100 studenti propone ai suoi allievi quanto segue. Si preparano 100 buste, numerate da 1 a 100, e 100 carte, su ciascuna delle quali è scritto il nome di uno studente del corso (senza ripetizioni, si escludano

omonimie). Quindi le carte vengono inserite, casualmente, una per ogni busta. Le buste, chiuse ma non sigillate, vengono quindi disposte sulla cattedra di un'aula. Gli studenti entrano nell'aula uno per volta. Ogni studente apre a suo piacimento 50 buste, e comunica al docente se, tra le buste aperte, c'è quella con il proprio nome. Quindi le richiude ed esce dall'aula, e da quel momento non può più comunicare con i colleghi che ancora devono entrare in aula. Il docente alzerà il voto dell'esame di tre punti a tutti, solo nel caso in cui *ciascuno studente* trovi la busta contenente la carta con il proprio nome. Gli studenti non possono comunicare dopo l'inizio delle aperture, ma possono concordare una strategia *a priori*. Si determini una strategia che conduca al successo (cioè all'aumento di tre punti per tutti) con probabilità almeno 0.3.

Soluzione. È assolutamente non ovvio che questo problema abbia soluzione. Le strategie "banali" falliscono miseramente. Supponiamo, ad esempio, che gli studenti non si accordino per nulla, ad esempio che ognuno di essi scelga a caso, indipendentemente dagli altri, le 50 buste da aprire. In questo caso è facile mostrare che ognuno avrebbe probabilità $\frac{1}{2}$ di trovare il proprio nome e, vista l'indipendenza delle scelte, la probabilità che *tutti* trovino il proprio nome sarebbe $\frac{1}{2^{100}}$: irrisoria! Si può fare naturalmente di peggio: se fossero così sciocchi da accordarsi di aprire tutti le *stesse* 50 buste, la probabilità di successo sarebbe nulla. Quello che non è ovvio è se sia possibile fare meglio.

Per semplificare le notazioni, poniamo $n := 100$ ed etichettiamo i cento nomi con i numeri $1, 2, \dots, n$. Denotiamo inoltre con $\sigma(k)$ il numero (nome) all'interno della busta numero k . Tale σ è evidentemente un elemento di S_n , e la probabilità uniforme su S_n corrisponde al fatto che i nomi nelle buste vengono inseriti a caso. Lo scopo di ogni studente k è di aprire la busta numero j con $\sigma(j) = k$. Supponiamo che gli studenti si accordino per seguire la seguente strategia. Ogni studente k apre per prima la busta k , e ne legge il contenuto $\sigma(k)$. Quindi apre la busta $\sigma(k)$ leggendone il contenuto $\sigma^2(k)$, e così via. Se, nella permutazione σ , l'elemento k appartiene ad un ciclo di lunghezza $m \leq \frac{n}{2}$, la m -sima busta aperta è la busta $\sigma^{m-1}(k)$, il cui contenuto è $\sigma^m(k) = k$: questo significa che lo studente trova la carta col proprio nome! Segue pertanto che se non ci sono in σ cicli di lunghezza maggiore di $n/2$, *ogni studente troverà sicuramente la busta contenente il proprio nome*. Perciò

$$\text{probabilità di successo della strategia} \geq 1 - p_n,$$

dove p_n è la probabilità calcolata nel problema precedente. Avendo visto che $p_n \simeq \log 2 \simeq 0.69$, abbiamo ottenuto quanto richiesto (per scrupolo, per $n = 100$, si calcola $p_n \simeq 0.688$). Sottolineiamo che il limite inferiore ottenuto alla probabilità di successo è approssimativamente indipendente da n , se n è abbastanza grande.

Per capire meglio la strategia, definiamo per $m = 1, \dots, n$ l'evento $B_m := \{\text{lo studente numero } m \text{ trova la carta col proprio nome}\}$, e poniamo $B := \{\text{tutti gli studenti trovano la carta col proprio nome}\} = \bigcap_{m=1}^n B_m$. Non è difficile convincersi del fatto che $P(B_m) = 0.5$ per ogni $m = 1, \dots, n$, *qualunque sia la strategia seguita!* Di conseguenza $P(B) \leq 0.5$. Con la strategia proposta abbiamo mostrato che $P(B) = P(\bigcap_{m=1}^n B_m) \geq 0.3$. In particolare, gli eventi $\{B_m\}_{1 \leq m \leq n}$ sono "molto sovrapposti" (quindi tutt'altro che indipendenti). Il cuore della soluzione consiste proprio nel determinare una strategia tale che, se si verifica il primo evento A_1 , con grande probabilità si verificano tutti gli altri eventi A_m con $m \geq 1$. \square

2.1.2 Punti fissi

Problema 2.4. Una comitiva di n turisti si sta imbarcando per un viaggio aereo. La loro guida ha tutte le carte d'imbarco (nominative), che deve distribuire ai turisti prima dell'imbarco. Per la fretta e la confusione le distribuisce a caso. Qual è la probabilità che qualcuno dei turisti riceva effettivamente la propria carta d'imbarco? Qual è la probabilità che esattamente m turisti ricevano la propria carta d'imbarco?

Soluzione. Al solito, etichettiamo con $\{1, 2, \dots, n\}$ gli n turisti e sia $\sigma(i)$ il numero (nome) sulla carta d'imbarco ricevuta dal turista i . Chiaramente $\sigma \in S_n$. L' i -esimo turista riceve la propria carta d'imbarco se $\sigma(i) = i$, cioè se i è un *punto fisso* della permutazione σ . Dunque, i quesiti del problema si possono riformulare come segue: qual è la probabilità che una permutazione abbia almeno un punto fisso? E qual è la probabilità che abbia esattamente m punti fissi?

Per $m = 0, 1, \dots, n$ e $i = 1, 2, \dots, n$, introduciamo gli eventi

$$A_m := \{\sigma \in S_n : \sigma \text{ ha esattamente } m \text{ punti fissi}\}, \quad C_i := \{\sigma \in S_n : \sigma(i) = i\}.$$

Chiaramente

$$A_0^c = C_1 \cup C_2 \cup \dots \cup C_n.$$

Per la formula di inclusione-esclusione (Proposizione 1.9)

$$P(A_0^c) = \sum_{k=1}^n \sum_{\substack{J \subseteq \{1, 2, \dots, n\} \\ \text{tale che } |J|=k}} (-1)^{k+1} P\left(\bigcap_{i \in J} C_i\right). \quad (2.1)$$

Fissiamo dunque $k \in \{1, 2, \dots, n\}$, e sia $J \subseteq \{1, 2, \dots, n\}$ tale che $|J| = k$. Si ha

$$\bigcap_{i \in J} C_i = \{\sigma \in S_n : \sigma(i) = i \text{ per ogni } i \in J\}.$$

Le permutazioni che lasciano fissi gli elementi di J sono in naturale corrispondenza biunivoca con le permutazioni di $\{1, 2, \dots, n\} \setminus J$, e quindi

$$\left| \bigcap_{i \in J} C_i \right| = (n - k)! \Rightarrow P\left(\bigcap_{i \in J} C_i\right) = \frac{(n - k)!}{n!}.$$

Poichè i sottoinsiemi J di $\{1, 2, \dots, n\}$ con k elementi sono $\binom{n}{k}$, si ha

$$\sum_{\substack{J \subseteq \{1, 2, \dots, n\} \\ \text{tale che } |J|=k}} P\left(\bigcap_{i \in J} C_i\right) = \binom{n}{k} \frac{(n - k)!}{n!} = \frac{1}{k!}.$$

Inserendo quest'ultima uguaglianza in (2.1) otteniamo

$$P(A_0^c) = \sum_{k=1}^n \frac{(-1)^{k+1}}{k!} = 1 - \sum_{k=0}^n \frac{(-1)^k}{k!}.$$

È inoltre ben noto che

$$\left| \sum_{k=0}^n \frac{(-1)^k}{k!} - e^{-1} \right| \leq \frac{1}{(n+1)!}.$$

Quindi l'approssimazione

$$P(A_0^c) \simeq 1 - e^{-1} \simeq 0.632$$

è eccellente per valori non troppo piccoli di n (già per $n = 6$ i due numeri hanno le prime tre cifre decimali uguali). Dunque, la probabilità che almeno un passeggero riceva la sua carta di imbarco è “quasi” indipendente dal numero di passeggeri!

Resta da determinare $P(A_m)$ per $m \geq 1$. Nel seguito usiamo la notazione

$$q_n := \sum_{k=0}^n \frac{(-1)^k}{k!},$$

che, come appena visto, è la probabilità che dell'insieme delle permutazioni di un insieme di n elementi che non hanno alcun punto fisso, cioè $P(A_0) = q_n$.

Per $J \subseteq \{1, 2, \dots, n\}$, $|J| = m$, sia

$$B_J := \{\sigma \in S_n : \sigma(j) = j \text{ per ogni } j \in J, \sigma(i) \neq i \text{ per ogni } i \notin J\}.$$

Ogni elemento di B_J può essere identificato con una permutazione degli elementi di J^c che non ha alcun punto fisso. Per quanto appena visto, ci sono $q_{n-m}(n-m)!$ tali permutazioni. Pertanto

$$|B_J| = q_{n-m}(n-m)! \Rightarrow P(B_J) = q_{n-m} \frac{(n-m)!}{n!}.$$

Infine essendo

$$A_m = \bigcup_{J:|J|=m} B_J,$$

ed essendo la precedente l'unione di insiemi disgiunti,

$$P(A_m) = \binom{n}{m} q_{n-m} \frac{(n-m)!}{n!} = \frac{q_{n-m}}{m!}.$$

Se m non è troppo vicino a n

$$P(A_m) \simeq \frac{e^{-1}}{m!}.$$

□

2.2 La passeggiata aleatoria semplice

Consideriamo un moto *discreto* sull'insieme dei numeri interi \mathbb{Z} che avvenga con le seguenti regole:

- la posizione iniziale, cioè all'istante $k = 0$, è 0;
- se $x \in \mathbb{Z}$ è la posizione all'istante k , allora le posizioni possibili all'istante $k + 1$ sono $x + 1$ e $x - 1$;

- fissato l'istante finale n , tutti i cammini *possibili* che terminano all'istante n sono equiprobabili.

Un cammino possibile è dunque identificabile con un vettore (s_0, s_1, \dots, s_n) , dove $s_0 = 0$ e, per $k = 0, 1, \dots, n-1$, $|s_{k+1} - s_k| = 1$. È facile vedere che l'applicazione $(s_0, s_1, \dots, s_n) \mapsto (x_1, x_2, \dots, x_n)$ data da $x_k := s_k - s_{k-1}$ fornisce un biiezione tra l'insieme dei cammini possibili e l'insieme $\{-1, 1\}^n$, la cui inversa è data, per $k \geq 1$, da $s_k = \sum_{j=1}^k x_j$. In particolare, ci sono 2^n cammini possibili che terminano all'istante n . Le variabili x_k rappresentano gli *incrementi* del cammino, mentre le variabili s_k rappresentano le *posizioni* del cammino nei diversi istanti.

Data la biiezione appena citata, è indifferente lavorare con le posizioni o con gli incrementi: per semplicità, scegliamo come spazio campionario Ω_n del nostro modello lo spazio degli incrementi, cioè $\Omega_n := \{-1, 1\}^n$. Come suggerito dalle regole citate sopra, muniamo Ω_n della probabilità uniforme, che indichiamo con P . Lo spazio di probabilità (Ω_n, P) viene chiamato *passeggiata aleatoria semplice e simmetrica* (di n passi). L'aggettivo *semplice* sta ad indicare che gli incrementi possono assumere solo i valori ± 1 , mentre *simmetrica* indica che le sequenze di incrementi sono tutte equiprobabili. Si tratta del più semplice modello per un moto aleatorio, tuttavia di rilevanza teorica e applicativa fondamentale. La seguente è una lista di domande "classiche" relative al nostro modello.

- Con quale probabilità e con quale frequenza la passeggiata raggiunge un determinato valore?
- Quanto tempo trascorre la passeggiata aleatoria al di sopra di un determinato valore?
- Quali sono i valori "tipici" di s_n , per n grande? (Il senso concreto dell'aggettivo "tipici" sarà chiarito in seguito.)

Prima di analizzare in dettaglio alcune di queste domande, c'è una questione che merita di essere approfondita. Consideriamo un evento che dipende solo dai primi n incrementi della passeggiata aleatoria, come ad esempio "la posizione s_n al tempo n è uguale a 0". Per descrivere questo evento, è naturale considerare il sottoinsieme di Ω_n dato da

$$A_n := \{(x_1, \dots, x_n) \in \Omega_n : x_1 + x_2 + \dots + x_n = 0\}. \quad (2.2)$$

La scelta di Ω_n non è tuttavia obbligata: è altrettanto legittimo adottare come spazio campionario Ω_N , per un qualunque valore di $N \geq n$, e definire l'analogo sottoinsieme di Ω_N in termini delle prime n variabili x_1, \dots, x_n :

$$A_N := \{(x_1, \dots, x_N) \in \Omega_N : x_1 + x_2 + \dots + x_n = 0\}.$$

Questa "ambiguità" non crea problemi, perché le probabilità degli eventi A_n e A_N è la stessa. Più precisamente, indicando per chiarezza con P_n e P_N le probabilità su Ω_n e Ω_N rispettivamente, si ha $P_n(A_n) = P_N(A_N)$ per ogni $n \geq N$. La dimostrazione è semplice: dato che $A_N = A_n \times \{-1, 1\}^{N-n}$, si può scrivere

$$P_N(A_N) = \frac{|A_N|}{|\Omega_N|} = \frac{|A_n| 2^{N-n}}{2^N} = \frac{|A_n|}{2^n} = P_n(A_n).$$

Si noti che non si è usata in alcun modo la forma esplicita dell'evento A_n , data dall'equazione (2.2), ma solo il fatto che $A_n \subseteq \Omega_n$ e che $A_N = A_n \times \{-1, 1\}^{N-n}$. Abbiamo ottenuto

un'importante conclusione: per calcolare la probabilità di un evento che dipende solo dai primi n incrementi (equivalentemente, dalle prime n posizioni) della passeggiata aleatoria, si può scegliere come spazio campionario Ω_N , per un qualunque valore di $N \geq n$.

Data questa arbitrarietà nella scelta dello spazio Ω_N , risulta *naturale* (almeno per un matematico...) considerare lo spazio campionario dato dai cammini di lunghezza infinita $\Omega_\infty := \{-1, 1\}^{\mathbb{N}}$, che contiene in modo canonico Ω_N per ogni $N \in \mathbb{N}$. Il problema è di definire la “giusta” probabilità su Ω_∞ , che estenda (in un senso da precisare) la probabilità uniforme su Ω_n . Infatti lo spazio Ω_∞ è infinito (dunque la probabilità uniforme non ha senso) e non è neppure numerabile, dunque la nozione stessa di probabilità che abbiamo introdotto nel capitolo 1 non si può applicare. Questo problema ammette una soluzione positiva, che però richiede una nozione più generale di spazio di probabilità, e sarà affrontato nel capitolo 4.

2.2.1 Considerazioni preliminari

Indichiamo con $\mathbf{x} := (x_1, x_2, \dots, x_n)$ l'elemento generico di Ω_n . Per $n \in \mathbb{N}$ e $m \in \mathbb{Z}$, introduciamo l'evento

$$\{s_n = m\} := \{\mathbf{x} \in \Omega_n : x_1 + x_2 + \dots + x_n = m\}. \quad (2.3)$$

Sottolineiamo che “ $\{s_n = m\}$ ” è semplicemente una *notazione* (che riprenderemo e generalizzeremo nel capitolo 3) per indicare l'evento definito dal membro destro in (2.3). Per la probabilità dell'evento $\{s_n = m\}$ scriveremo semplicemente $P(s_n = m)$. È facile vedere che se n è pari, necessariamente s_n è pari e, analogamente, se n è dispari, necessariamente s_n è dispari, cioè $P(s_n = m) = 0$ se n e m non hanno la stessa parità. Pertanto è sufficiente considerare probabilità del tipo

$$P(s_{2n} = 2m), \quad P(s_{2n+1} = 2m + 1),$$

dove $n \in \mathbb{N}$ e $m \in \mathbb{Z}$. Inoltre, essendo gli incrementi di modulo 1, $P(s_{2n} = 2m) > 0$ se e solo se $|m| \leq n$, mentre $P(s_{2n+1} = 2m + 1) > 0$ se e solo se $-n - 1 \leq m \leq n$.

Consideriamo ora un elemento $\mathbf{x} = (x_1, x_2, \dots, x_{2n}) \in \Omega_{2n}$. Sia k il numero di incrementi positivi di \mathbf{x} , cioè $k := |\{i : x_i = 1\}|$. Essendo $2n - k$ gli incrementi negativi, segue che la posizione finale del cammino corrispondente è $s_{2n} = x_1 + \dots + x_{2n} = k - (2n - k) = 2(k - n)$. Pertanto $\mathbf{x} \in \Omega_{2n}$ è un elemento dell'evento $\{s_{2n} = 2m\}$ se e solo se il numero di incrementi positivi di \mathbf{x} è $k = n + m$. Da questo segue facilmente che

$$|\{s_{2n} = 2m\}| = \binom{2n}{n+m},$$

e quindi

$$P(s_{2n} = 2m) = \frac{|\{s_{2n} = 2m\}|}{|\Omega_{2n}|} = \frac{1}{2^{2n}} \binom{2n}{n+m}. \quad (2.4)$$

Lasciamo al lettore verificare, in modo del tutto analogo, che se si considera lo spazio Ω_{2n+1} dei cammini di lunghezza dispari $2n + 1$, si ha

$$P(s_{2n+1} = 2m + 1) = \frac{1}{2^{2n+1}} \binom{2n+1}{n+m+1}. \quad (2.5)$$

2.2.2 Il problema della ricorrenza

Ci interessiamo ora alla probabilità che la passeggiata aleatoria ritorni al punto di partenza: per $n \in \mathbb{N}$ poniamo

$$R_n := P(s_k = 0 \text{ per qualche } k = 1, 2, \dots, n). \quad (2.6)$$

L'obiettivo è di capire il comportamento di R_n per $n \rightarrow \infty$.

Ricordando che $s_k = 0$ è possibile solo per k pari, definiamo

$$\begin{aligned} u_{2n} &:= P(s_{2n} = 0), \\ f_{2n} &:= P(s_2 \neq 0, \dots, s_{2(n-1)} \neq 0, s_{2n} = 0). \end{aligned}$$

Si noti che u_{2n} è la probabilità che la passeggiata aleatoria valga 0 al passo $2n$, mentre f_{2n} è la probabilità che la passeggiata aleatoria ritorni a 0 *per la prima volta* al passo $2n$. Dire che la passeggiata aleatoria visita zero in un qualche passo $k \leq 2n$ è equivalente a dire che il primo ritorno a zero avviene prima di $2n$ passi: vale cioè l'uguaglianza di eventi

$$\{s_k = 0 \text{ per qualche } k = 1, \dots, 2n\} = \bigcup_{m=1}^n \{s_2 \neq 0, \dots, s_{2(m-1)} \neq 0, s_{2m} = 0\},$$

e inoltre gli eventi che appaiono nell'unione sono a due a due disgiunti. Si ha pertanto l'uguaglianza

$$R_{2n} = \sum_{k=1}^n f_{2k}. \quad (2.7)$$

Si noti che, grazie a (2.4), si ha

$$u_{2n} = \frac{1}{2^{2n}} \binom{2n}{n}.$$

Nel prossimo paragrafo determineremo una formula esplicita per f_{2n} :

$$f_{2n} = \frac{1}{2n} u_{2n-2} = \frac{1}{2n-1} u_{2n},$$

che tuttavia non è particolarmente utile per determinare il comportamento asintotico, per n grande, delle somme parziali $\sum_{k=1}^n f_{2k}$ (e dunque di R_{2n} , grazie a (2.7)). Sarà invece fondamentale il seguente Lemma.

Lemma 2.5. *Per ogni $n > 0$*

$$u_{2n} = \sum_{k=1}^n f_{2k} u_{2(n-k)}.$$

Dimostrazione. Sia $A := \{s_{2n} = 0\}$, per cui $u_{2n} = P(A)$. L'evento A si può scrivere come unione dei seguenti n eventi disgiunti:

$$\begin{aligned} A_k &:= \{s_2 \neq 0, \dots, s_{2k-2} \neq 0, s_{2k} = 0, s_{2n} = 0\}, \\ A &= \bigcup_{k=1}^n A_k. \end{aligned}$$

Contiamo i cammini in A_k . La cardinalità di A_k è uguale al numero di cammini di lunghezza $2k$ che ritornano a 0 la prima volta dopo $2k$ passi ($2^{2k} f_{2k}$) moltiplicato il numero di cammini di lunghezza $2n - 2k$ che terminano in 0 ($2^{2(n-k)} u_{2(n-k)}$). Pertanto

$$P(A_k) = \frac{1}{2^{2n}} |A_k| = \frac{1}{2^{2n}} 2^{2k} f_{2k} 2^{2(n-k)} u_{2(n-k)} = f_{2k} u_{2(n-k)}.$$

Essendo $P(A) = \sum_{k=1}^n P(A_k)$, la conclusione segue facilmente. \square

Lemma 2.6. *Siano $(a_n)_{n \geq 0}$ e $(b_n)_{n \geq 1}$ due successioni di numeri reali positivi tali che $a_0 > 0$ e, per ogni $n \geq 1$,*

$$a_n = \sum_{k=1}^n b_k a_{n-k}.$$

Supponiamo che $a_n \leq 1$ per ogni $n \geq 1$. Allora $\sum_{n=1}^{+\infty} b_n \leq 1$ e

$$\sum_{n=1}^{+\infty} b_n = 1 \iff \sum_{n=1}^{+\infty} a_n = +\infty.$$

Dimostrazione. Usando il fatto che, per somme infinite ad addendi positivi è lecito permutare l'ordine degli addendi, abbiamo, posto $s := \sum_{n=1}^{+\infty} a_n$,

$$s = \sum_{n=1}^{+\infty} a_n = \sum_{n=1}^{+\infty} \sum_{k=1}^n b_k a_{n-k} = \sum_{k=1}^{\infty} b_k \sum_{n=k}^{+\infty} a_{n-k} = \sum_{k=1}^{+\infty} b_k \sum_{m=0}^{+\infty} a_m = (a_0 + s) \sum_{k=1}^{+\infty} b_k.$$

Ciò implica immediatamente che, se $s < +\infty$, allora

$$\sum_{k=1}^{+\infty} b_k = \frac{s}{a_0 + s} < 1.$$

Resta da dimostrare che, se $s = +\infty$, si ha $\sum_{k=1}^{+\infty} b_k = 1$. Possiamo scrivere

$$\sum_{n=1}^N a_n = \sum_{n=1}^N \sum_{k=1}^n b_k a_{n-k} = \sum_{k=1}^N b_k \sum_{n=k}^N a_{n-k} = \sum_{k=1}^N b_k \sum_{n=0}^{N-k} a_n. \quad (2.8)$$

Trattandosi di somme a termini positivi, valgono le seguenti disuguaglianze:

$$\sum_{k=1}^N b_k \sum_{n=0}^{N-k} a_n \leq \sum_{k=1}^N b_k \sum_{n=0}^N a_n = \sum_{k=1}^N b_k \left(a_0 + \sum_{n=1}^N a_n \right) \quad (2.9)$$

e, per $m \leq N$,

$$\sum_{k=1}^N b_k \sum_{n=0}^{N-k} a_n \geq \sum_{k=1}^m b_k \sum_{n=0}^{N-k} a_n \geq \sum_{k=1}^m b_k \sum_{n=0}^{N-m} a_n \geq \sum_{k=1}^m b_k \left(a_0 + \sum_{n=1}^N a_n - m \right), \quad (2.10)$$

dove, nell'ultimo passaggio, abbiamo usato il fatto che $a_n \leq 1$ per ogni $n \geq 1$. Supponiamo ora $\sum_{n=0}^{+\infty} a_n = +\infty$. Da (2.8) e (2.9) segue che

$$\sum_{k=1}^N b_k \geq \frac{\sum_{n=1}^N a_n}{a_0 + \sum_{n=1}^N a_n},$$

da cui, passando al limite per $N \rightarrow +\infty$, si ottiene

$$\sum_{k=1}^{+\infty} b_k \geq 1. \quad (2.11)$$

Usando (2.8) e (2.10) otteniamo

$$\sum_{k=1}^m b_k \leq \frac{\sum_{n=1}^N a_n}{a_0 + \sum_{n=1}^N a_n - m},$$

da cui, passando prima al limite per $N \rightarrow +\infty$ e poi per $m \rightarrow +\infty$, otteniamo

$$\sum_{k=1}^{+\infty} b_k \leq 1, \quad (2.12)$$

che, assieme a (2.11) conclude la dimostrazione. \square

Usando il Lemma 2.6 con $a_n = u_{2n}$ e $b_n = f_{2n}$, segue che

$$\lim_{n \rightarrow +\infty} R_{2n} = \sum_{n=1}^{+\infty} f_{2n} = 1 \iff \sum_n u_{2n} = +\infty.$$

L'affermazione $\lim_{n \rightarrow +\infty} R_{2n} = 1$ si può esprimere dicendo che la probabilità che la passeggiata aleatoria torni al punto di partenza entro i primi n passi tende a 1 per $n \rightarrow +\infty$, e viene chiamata *proprietà di ricorrenza*. Per stabilire se tale affermazione sia vera o falsa, dobbiamo dunque studiare la convergenza della serie $\sum_n u_{2n}$. A tale scopo usiamo la seguente celebre formula di approssimazione.

Proposizione 2.7. (*Formula di Stirling*) Per ogni $n \geq 1$

$$n! = \left(\frac{n}{e}\right)^n \sqrt{2\pi n} e^{\frac{\theta(n)}{12n}},$$

dove $0 \leq \theta(n) \leq 1$.

Dimostrazione. Diamo ora una dimostrazione, piuttosto elementare, dell'esistenza di una costante $C > 0$ tale che

$$n! = C \left(\frac{n}{e}\right)^n \sqrt{n} e^{\frac{\theta(n)}{12n}}. \quad (2.13)$$

Questo sarà sufficiente per i risultati di questo paragrafo. Nel prossimo paragrafo dimostreremo che in realtà $C = \sqrt{2\pi}$.

Notiamo anzitutto che (2.13) è equivalente a dimostrare che esiste una costante $c (= \log C)$ tale che per ogni $n \geq 1$

$$0 \leq \log n! - \left(n + \frac{1}{2}\right) \log n + n - c \leq \frac{1}{12n}. \quad (2.14)$$

Poniamo $d_n := \log n! - \left(n + \frac{1}{2}\right) \log n + n$. È semplice (esercizio!) mostrare che (2.14) segue dai seguenti due fatti, con $c = \lim_{n \rightarrow \infty} d_n$:

- (i) La successione d_n è decrescente.

(ii) La successione $d_n - \frac{1}{12n}$ è crescente.

Quindi, non ci rimane che dimostrare le affermazioni (i) e (ii).

Con semplici calcoli si ottiene:

$$d_n - d_{n+1} = \left(n + \frac{1}{2}\right) \log \frac{n+1}{n} - 1 = \frac{2n+1}{2} \log \frac{1 + \frac{1}{2n+1}}{1 - \frac{1}{2n+1}} - 1.$$

A questo punto si usa la serie di Taylor $\log(1+t) = \sum_{k=1}^{\infty} (-1)^{k+1} \frac{t^k}{k}$, convergente per $|t| < 1$, per ottenere

$$\log \frac{1+t}{1-t} = \log(1+t) - \log(1-t) = \sum_{k=1}^{\infty} (-1)^{k+1} \frac{t^k}{k} + \sum_{k=1}^{\infty} \frac{t^k}{k} = 2 \sum_{k=0}^{\infty} \frac{t^{2k+1}}{2k+1},$$

che converge anch'essa per $|t| < 1$. Usando tale serie per $t = \frac{1}{2n+1}$ si trova

$$d_n - d_{n+1} = (2n+1) \sum_{k=0}^{\infty} \frac{1}{2k+1} \frac{1}{(2n+1)^{2k+1}} - 1 = \sum_{k=1}^{\infty} \frac{1}{2k+1} \frac{1}{(2n+1)^{2k}} \geq 0, \quad (2.15)$$

essendo quest'ultima una serie a termini positivi. Ciò dimostra (i).

Usando di nuovo (2.15) e il fatto che $2k+1 \geq 3$ per $k \geq 1$, si ottiene

$$\begin{aligned} d_n - d_{n+1} &= \sum_{k=1}^{\infty} \frac{1}{2k+1} \frac{1}{(2n+1)^{2k}} \leq \frac{1}{3} \sum_{k=1}^{\infty} \frac{1}{(2n+1)^{2k}} = \frac{1}{3} \frac{(2n+1)^{-2}}{1 - (2n+1)^{-2}} \\ &= \frac{1}{3[(2n+1)^2 - 1]} = \frac{1}{12n(n+1)} = \frac{1}{12n} - \frac{1}{12(n+1)}, \end{aligned}$$

dove abbiamo usato la somma della serie geometrica

$$\sum_{n=1}^{\infty} x^n = \frac{1}{1-x} \quad \text{per } |x| < 1.$$

In altri termini, abbiamo dimostrato che

$$d_n - \frac{1}{12n} \leq d_{n+1} - \frac{1}{12(n+1)},$$

cioè la relazione (ii). □

Dimostriamo finalmente la ricorrenza della passeggiata aleatoria semplice e simmetrica su \mathbb{Z} .

Teorema 2.8.

$$\lim_{n \rightarrow +\infty} R_{2n} = 1.$$

Dimostrazione. Come osservato prima, la tesi è equivalente a

$$\sum_n u_{2n} = +\infty. \quad (2.16)$$

Usando la formula di Stirling, abbiamo

$$u_{2n} = \frac{1}{2^{2n}} \binom{2n}{n} = \frac{1}{2^{2n}} \frac{(2n)!}{(n!)^2} = \frac{1}{2^{2n}} \frac{C \left(\frac{2n}{e}\right)^{2n} \sqrt{2ne} \frac{\theta(2n)}{24n}}{C^2 \left(\frac{n}{e}\right)^{2n} n e \frac{\theta(n)}{6n}} = \frac{\sqrt{2}}{C\sqrt{n}} \exp \left[\frac{\theta(2n)}{24n} - \frac{\theta(n)}{6n} \right],$$

da cui segue

$$\lim_{n \rightarrow +\infty} \frac{u_{2n}}{1/\sqrt{n}} = \frac{\sqrt{2}}{C}.$$

Da ciò, per il criterio del confronto asintotico tra serie, si ricava (2.16), e questo conclude la dimostrazione. \square

Concludiamo questo paragrafo con la nozione di passeggiata aleatoria semplice e simmetrica multi-dimensionale. Per $d \geq 1$, consideriamo i cammini di lunghezza n “generati” dallo spazio di incrementi

$$\Omega_n^d = \{\mathbf{x} = (x_1, x_2, \dots, x_n) : x_i \in \{-1, 1\}^d \text{ per } i = 1, 2, \dots, n\}.$$

Ciò significa che consideriamo cammini uscenti dall’origine di \mathbb{Z}^d e la cui posizione al tempo k è $s_k := x_1 + x_2 + \dots + x_k \in \mathbb{Z}^d$, dove $x_i \in \{-1, 1\}^d$. Tutti questi cammini si assumono equiprobabili, cioè la probabilità P su Ω_n^d è quella uniforme. Si noti che se $A \subseteq \Omega_n = \{-1, 1\}^n$ allora $A^d = A \times A \times \dots \times A \subseteq \Omega_n^d$ e vale la formula

$$|A^d| = |A|^d.$$

Consideriamo allora l’evento $S_{2n}^{(d)} := \{s_{2n} = 0\}$ (in Ω_N^d , con $N \geq 2n$), dove 0 denota l’origine di \mathbb{Z}^d , e denotiamo con $u_{2n}^{(d)}$ la sua probabilità. In particolare, $u_{2n}^{(1)} = u_{2n}$. Poichè $s_{2n} = 0$ se e solo se tutte le d componenti sono uguali a zero, si ha

$$S_{2n}^{(d)} = (S_{2n}^{(1)})^d.$$

Pertanto

$$u_{2n}^{(d)} = P(S_{2n}^{(d)}) = \frac{|S_{2n}^{(d)}|}{|\Omega_N^d|} = \frac{|S_{2n}^{(1)}|^d}{|\Omega_N|^d} = (u_{2n})^d. \quad (2.17)$$

In modo analogo al caso unidimensionale, possiamo definire

$$f_{2k}^{(d)} := P(s_2 \neq 0, \dots, s_{2(k-1)} \neq 0, s_{2k} = 0).$$

Allora la quantità $R_{2n}^{(d)}$, definita da

$$R_{2n}^{(d)} := \sum_{k=1}^n f_{2k}^{(d)},$$

è la probabilità che la passeggiata aleatoria torni all’origine entro $2n$ passi. La relazione

$$u_{2n}^{(d)} = \sum_{k=1}^n f_{2k}^{(d)} u_{2(n-k)}^{(d)}$$

si dimostra esattamente come nel caso $d = 1$. Dunque, applicando il Lemma 2.6, abbiamo che la passeggiata aleatoria è ricorrente, cioè $\lim_{n \rightarrow \infty} R_{2n}^{(d)} = 1$, se e solo se $\sum_n u_{2n}^{(d)} = +\infty$. D'altra parte, essendo $u_{2n}^{(d)} = (u_{2n})^d$ e avendo visto che u_{2n} è asintoticamente equivalente a $n^{-1/2}$, deduciamo che $u_{2n}^{(d)}$ è asintoticamente equivalente a $n^{-d/2}$. Poichè la serie di termine generale $n^{-\alpha}$ converge se e solo se $\alpha > 1$, possiamo concludere quanto segue.

Teorema 2.9. *La passeggiata aleatoria semplice e simmetrica in dimensione d è ricorrente per $d = 1, 2$ e non è ricorrente per $d \geq 3$.*

Un modo suggestivo per esprimere la non ricorrenza della passeggiata aleatoria in dimensione ≥ 3 , consiste nell'affermare che "la passeggiata aleatoria in dimensione ≥ 3 ha una probabilità strettamente positiva di non ritornare mai all'origine". Per dare un significato a questa affermazione sarebbe però necessario introdurre lo spazio campionario delle passeggiate aleatorie di lunghezza infinita. Quindi, per il momento, la descrizione rigorosa della non ricorrenza è: "se $d \geq 3$, esiste una costante $\epsilon > 0$ tale che per ogni $n \in \mathbb{N}$ la probabilità che la passeggiata in dimensione ≥ 3 non sia mai tornata all'origine nei primi n passi è maggiore di ϵ ".

2.2.3 La legge dell'arcoseno

Consideriamo il seguente evento:

$$A_{2k,2n} = \{|\{i : 0 \leq i \leq 2n, s_i \geq 0\}| = 2k\}.$$

In altre parole, $A_{2k,2n}$ è l'insieme delle traiettorie di lunghezza $2n$ che assumono valore positivo $2k$ volte (e quindi valore negativo $2n - 2k$ volte; non è difficile convincersi che $|\{i : 0 \leq i \leq 2n, s_i \geq 0\}|$ è necessariamente un numero pari). Definiamo

$$P_{2k,2n} := P(A_{2k,2n}).$$

Quanto vale questa probabilità? E per quali valori di k essa assume il valore massimo? Data la simmetria della passeggiata aleatoria, si potrebbe pensare che la cosa più probabile sia che essa trascorra circa metà del tempo sul semiasse positivo; cioè che $P_{2k,2n}$ sia massimo per $k \simeq n/2$, almeno quando n è grande. *Niente di più falso!* È vero esattamente il contrario: $P_{2k,2n}$ assume il valore massimo per $k = 0$ e $k = n$. Inoltre, come vedremo, è assai probabile che la passeggiata aleatoria trascorra "gran parte" del tempo sullo stesso semiasse. Per dare un'immagine suggestiva di questo fenomeno, supponiamo di giocare ripetutamente al seguente gioco: viene lanciata una moneta, se esce *testa* vinciamo un Euro, altrimenti perdiamo un Euro. È molto probabile che noi si sia gran parte del tempo in attivo (cioè il denaro vinto supera quello perso) e poco tempo in passivo, oppure gran parte del tempo in passivo e poco tempo in attivo, mentre è molto meno probabile che il tempo trascorso in attivo sia grosso modo uguale a quello trascorso in passivo.

Cominciamo col calcolare la probabilità che la passeggiata aleatoria resti sempre sul semiasse positivo (che, per simmetria, sarà uguale alla probabilità che la passeggiata aleatoria resti sempre sul semiasse negativo).

Proposizione 2.10. *Per ogni $n \geq 1$*

$$P_{2n,2n} = u_{2n}.$$

La dimostrazione della Proposizione 2.10 richiede il seguente Lemma, di natura tecnica ma che contiene uno degli argomenti più importanti della teoria delle passeggiate aleatorie: il *principio di riflessione*.

Lemma 2.11. Per ogni $n \geq 1$

(i)

$$P(s_1 \geq 0, s_2 \geq 0, \dots, s_{2n-1} \geq 0, s_{2n} = 0) = \frac{1}{n+1} u_{2n}.$$

(ii)

$$P(s_1 > 0, s_2 > 0, \dots, s_{2n-1} > 0, s_{2n} = 0) = \frac{1}{4n} u_{2n-2}.$$

(iii)

$$f_{2n} = \frac{1}{2n} u_{2n-2} = u_{2n-2} - u_{2n},$$

dove, per $n = 1$, si pone $u_0 = 1$.

Dimostrazione. (i) Consideriamo gli eventi

$$A := \{s_1 \geq 0, s_2 \geq 0, \dots, s_{2n-1} \geq 0, s_{2n} = 0\},$$

e $A_0 := \{s_{2n} = 0\}$. Si noti che

$$B := A_0 \setminus A = \{s_{2n} = 0, s_{2k+1} = -1 \text{ per qualche } k = 0, 1, \dots, n-1\}. \quad (2.18)$$

Infine, sia

$$A_2 := \{s_{2n} = -2\}.$$

Mostriamo ora che $|B| = |A_2|$, e pertanto $P(B) = P(A_2)$. Per far ciò esibiamo una corrispondenza biunivoca tra B e A_2 . L'argomento usato per costruire tale biiezione viene chiamato *principio di riflessione*.

Sia $\mathbf{x} \in B$. Per definizione di B , $s_i = -1$ per qualche $i = 1, \dots, 2n-1$. Indichiamo con $i(\mathbf{x})$ il minimo indice i per cui $s_i = -1$ (ovviamente tale indice dipende dal particolare \mathbf{x}). Sia $\mathbf{y} \in \Omega_{2n}$ così definito

$$y_i = \begin{cases} x_i & \text{se } i \leq i(\mathbf{x}) \\ -x_i & \text{se } i > i(\mathbf{x}). \end{cases}$$

Per rendere più chiaro il senso di questa costruzione, basti dire che la passeggiata relativa agli incrementi \mathbf{y} coincide con quella relativa agli incrementi \mathbf{x} fino a $i = i(\mathbf{x})$, e da quel punto in poi è ottenuta *per riflessione* rispetto all'asse $i = -1$, come illustrato nella figura seguente.

Si vede subito che $\mathbf{y} \in A_2$. Questa costruzione definisce allora una mappa $\varphi : B \rightarrow A_2$, in cui si pone $\varphi(\mathbf{x}) = \mathbf{y}$. Non è sorprendente il fatto che, essendo questa mappa ottenuta da una riflessione, si tratti di una funzione invertibile. Non è difficile convincersi che la funzione inversa è data, per $\mathbf{y} \in A_2$

$$(\varphi^{-1}\mathbf{y})_i = \begin{cases} y_i & \text{se } i \leq i(\mathbf{y}) := \min\{i : y_1 + \dots + y_i = -1\} \\ -y_i & \text{altrimenti.} \end{cases}$$

Abbiamo dunque dimostrato che $|B| = |A_2|$. La probabilità di A_2 è facile da calcolare. Infatti $\mathbf{x} \in A_2$ se e solo se il numero di incrementi negativi di \mathbf{x} è $n + 1$. Pertanto

$$|A_2| = \binom{2n}{n+1}.$$

Ne segue che

$$P(B) = P(A_2) = \frac{1}{2^{2n}} \binom{2n}{n+1}.$$

Usando (2.18)

$$P(A) = P(A_0) - P(B) = \frac{1}{2^{2n}} \left[\binom{2n}{n} - \binom{2n}{n+1} \right] = \frac{1}{2^{2n}} \binom{2n}{n} \left[1 - \frac{n}{n+1} \right] = \frac{1}{n+1} u_{2n}.$$

(ii) Sia $C := \{s_1 > 0, s_2 > 0, \dots, s_{2n-1} > 0, s_{2n} = 0\}$. Notare che

$$C = \{s_1 = 1, s_2 \geq 1, \dots, s_{2n-2} \geq 1, s_{2n-1} = 1, s_{2n} = 0\}.$$

Se $\mathbf{x} \in C$, si ha che $x_1 = 1$ e $x_{2n} = -1$. Ne segue che

$$|C| = |\{(x_1, \dots, x_{2n-2}) : s_1 \geq 0, \dots, s_{2n-3} \geq 0, s_{2n-2} = 0\}|.$$

Perciò, per quanto visto al punto (i),

$$|C| = 2^{2n-2} \frac{1}{n} u_{2n-2},$$

e quindi

$$P(C) = 2^{-2n} |C| = \frac{1}{4n} u_{2n-2}.$$

(iii) Per simmetria

$$\begin{aligned} f_{2n} &= P(s_1 \neq 0, \dots, s_{2n-1} \neq 0, s_{2n} = 0) \\ &= P(s_1 > 0, \dots, s_{2n-1} > 0, s_{2n} = 0) + P(s_1 < 0, \dots, s_{2n-1} < 0, s_{2n} = 0) \\ &= 2P(s_1 > 0, \dots, s_{2n-1} > 0, s_{2n} = 0) = \frac{1}{2n} u_{2n-2}, \end{aligned}$$

dove abbiamo usato il risultato al punto (ii). Infine, l'uguaglianza

$$\frac{1}{2n} u_{2n-2} = u_{2n-2} - u_{2n}$$

si verifica con un banale calcolo algebrico. □

Dimostrazione della Proposizione 2.10. Sia

$$A := \{s_1 \geq 0, s_2 \geq 0, \dots, s_{2n} \geq 0\},$$

per cui $P_{2n,2n} = P(A)$. Notare che $\mathbf{x} \in A^c$ se e solo se $s_i = -1$ per qualche $i = 1, \dots, 2n - 1$. Quindi, per $\mathbf{x} \in A^c$, è ben definito

$$i(\mathbf{x}) := \min\{i : s_i = -1\}.$$

Posto, per $k = 1, 2, \dots, n$, $B_k = \{\mathbf{x} \in A^c : i(\mathbf{x}) = 2k - 1\}$, è chiaro che i B_k sono disgiunti e

$$A^c = \bigcup_{k=1}^n B_k.$$

Contiamo ora gli elementi di B_k . Osserviamo che

$$B_k = \{s_1 \geq 0, s_2 \geq 0, \dots, s_{2k-3} \geq 0, s_{2k-2} = 0, s_{2k-1} = -1\}.$$

Gli incrementi (x_1, \dots, x_{2n}) di un elemento di B_k devono dunque avere le seguenti proprietà:

- i primi $2k - 2$ incrementi sono tali che $\{s_1 \geq 0, s_2 \geq 0, \dots, s_{2k-3} \geq 0, s_{2k-2} = 0\}$. Per il Lemma 2.11, tali incrementi possono essere scelti in $2^{2k-2} \frac{1}{k} u_{2k-2}$ modi.
- L'incremento x_{2k-1} vale necessariamente -1 .
- I rimanenti $2(n - k) + 1$ incrementi $x_{2k}, x_{2k+1}, \dots, x_{2n}$ sono arbitrari, e quindi possono essere scelti in $2^{2(n-k)+1}$ modi.

Perciò

$$|B_k| = 2^{2k-2} \frac{1}{k} u_{2k-2} 2^{2(n-k)+1} = 2^{2n-1} \frac{1}{k} u_{2k-2},$$

da cui si ottiene

$$P(B_k) = 2^{-2n} |B_k| = \frac{1}{k} u_{2k-2} = f_{2k} = u_{2k-2} - u_{2k},$$

dove abbiamo usato la parte (iii) del Lemma 2.11. Ma allora

$$P(A^c) = \sum_{k=1}^n P(B_k) = \sum_{k=1}^n [u_{2k-2} - u_{2k}] = 1 - u_{2n},$$

da cui si conclude immediatamente la dimostrazione. \square

Teorema 2.12. Per ogni $n \geq 1$ e $0 \leq k \leq n$

$$P_{2k,2n} = u_{2k} u_{2n-2k}.$$

Dimostrazione. Per la Proposizione 2.10, la tesi è vera per $k = n$ e quindi, per simmetria, per $k = 0$. In particolare, la tesi è sempre vera per $n = 1$. Procediamo per induzione su n . Sia $n > 1$. Poichè, come abbiamo appena detto, già sappiamo che la tesi è vera per $k = 0, n$, possiamo assumere $1 \leq k \leq n - 1$. Si ricordi l'evento $A_{2k,2n} = \{|\{i : 0 \leq i \leq 2n, s_i \geq 0\}| = 2k\}$, per cui $P(A_{2k,2n}) = P_{2k,2n}$. Siano

$$A_{2k,2n}^+ := \{\mathbf{x} \in A_{2k,2n} : x_1 = 1\} \quad A_{2k,2n}^- := \{\mathbf{x} \in A_{2k,2n} : x_1 = -1\}.$$

Denotiamo con

$$E_r = \{\mathbf{x} : s_{2k} = 0, s_i \neq 0 \text{ per ogni } i = 1, 2, \dots, 2k - 1\}$$

l'insieme delle traiettorie che tornano all'origine per la prima volta dopo $2r$ passi, e

$$E_{k,r}^\pm = A_{2k,2n}^\pm \cap E_r.$$

Notare che una traiettoria in $A_{2k,2n}^+$ deve necessariamente tornare a zero entro $2k$ passi, altrimenti trascorrerebbe più di $2k$ istanti nel semiasse positivo. Pertanto

$$A_{2k,2n}^+ = \bigcup_{r=1}^k E_{k,r}^+,$$

dove la precedente unione è evidentemente disgiunta. Contiamo ora gli elementi di $E_{k,r}^+$, i cui incrementi sono caratterizzati dalle seguenti proprietà:

- gli incrementi $(x_1, x_2, \dots, x_{2r})$ sono tali che $x_1 = 1$ e il primo ritorno a 0 avviene dopo $2r$ passi. Questi incrementi si possono scegliere in $\frac{1}{2}f_{2r}2^{2r}$ modi possibili.
- I successivi incrementi $(x_{2r+1}, \dots, x_{2n})$ devono essere tali che la traiettoria corrispondente trascorre $2k - 2r$ istanti nel semiasse positivo (la traiettoria complessiva ha già trascorso $2r$ istanti nel semiasse positivo prima del primo ritorno all'origine). Quindi questi incrementi si possono scegliere in $P_{2k-2r,2n-2r}2^{2n-2r}$ modi.

Ne segue che

$$|E_{k,r}^+| = \frac{1}{2}f_{2r}2^{2r}P_{2k-2r,2(n-r)}2^{2n-2r} = \frac{1}{2}f_{2r}P_{2k-2r,2(n-r)}2^{2n},$$

e pertanto

$$P(E_{k,r}^+) = \frac{1}{2}f_{2r}P_{2k-2r,2(n-r)}.$$

Essendo $r \geq 1$ e quindi $n - r < n$, possiamo usare l'ipotesi induttiva $P_{2k-2r,2(n-r)} = u_{2(k-r)}u_{2(n-k)}$. Ma allora

$$P(A_{2k,2n}^+) = \sum_{r=1}^k P(E_{k,r}^+) = \frac{1}{2}u_{2(n-k)} \sum_{r=1}^k f_{2r}u_{2(k-r)} = \frac{1}{2}u_{2(n-k)}u_{2k},$$

dove, in quest'ultima uguaglianza, abbiamo usato il Lemma 2.5. In modo del tutto analogo (esercizio!), partendo da

$$A_{2k,2n}^- = \bigcup_{r=1}^{n-k} E_{k,r}^-,$$

si mostra che anche $P(A_{2k,2n}^-) = \frac{1}{2}u_{2(n-k)}u_{2k}$ e quindi, essendo $P(A_{2k,2n}) = P(A_{2k,2n}^+) + P(A_{2k,2n}^-)$, si conclude. \square

Una facile conseguenza è la seguente.

Corollario 2.13. *La probabilità $P_{2k,2n}$ assume il valore massimo per $k = 0$ e $k = n$.*

Dimostrazione. Usando il Teorema 2.12, facendo le dovute semplificazioni si trova, per $0 \leq k \leq n - 1$

$$\frac{P_{2k,2n}}{P_{2(k+1),2n}} = \frac{k+1}{2k+1} \frac{2n-2k-1}{n-k},$$

da cui è facile ricavare

$$\frac{P_{2k,2n}}{P_{2(k+1),2n}} < 1 \iff k > \frac{n-1}{2}.$$

Segue che la funzione $k \mapsto P_{2k,2n}$ decresce per $k \leq \frac{n}{2}$ e cresce per $k \geq \frac{n}{2}$. Pertanto il massimo è assunto in $k = 0$ o in $k = n$. Tuttavia, per simmetria, $P_{0,2n} = P_{2n,2n}$. \square

Per $\mathbf{x} \in \Omega_{2n}$, sia $t(\mathbf{x})$ il tempo trascorso nel semiasse positivo dalla corrispondente traiettoria. Pertanto

$$P_{2k,2n} = P(\{\mathbf{x} : t(\mathbf{x}) = 2k\}).$$

Ci occupiamo ora di determinare il comportamento asintotico, per $n \rightarrow +\infty$, di probabilità della forma

$$P(\{\mathbf{x} : 2\beta n \leq t(\mathbf{x}) \leq 2\alpha n\}),$$

dove $0 \leq \beta < \alpha \leq 1$, cioè la probabilità che la *frazione* di tempo che la traiettoria trascorre nel semiasse positivo sia compresa tra β e α . Cominciamo con una semplice considerazione. Assumendo che esista il limite

$$\lim_{n \rightarrow +\infty} P(\{\mathbf{x} : 2\beta n \leq t(\mathbf{x}) \leq 2\alpha n\}), \quad (2.19)$$

esso è uguale al limite

$$\lim_{n \rightarrow +\infty} P(\{\mathbf{x} : 2\beta n < t(\mathbf{x}) \leq 2\alpha n\}).$$

Infatti, per il Corollario 2.13,

$$\begin{aligned} P(\{\mathbf{x} : 2\beta n \leq t(\mathbf{x}) \leq 2\alpha n\}) - P(\{\mathbf{x} : 2\beta n < t(\mathbf{x}) \leq 2\alpha n\}) \\ = P(\{\mathbf{x} : t(\mathbf{x}) = 2\beta n\}) \leq P_{2n,2n} = u_{2n} \xrightarrow{n \rightarrow +\infty} 0. \end{aligned}$$

Definiamo ora

$$\rho_n(\alpha) := P(\{\mathbf{x} : t(\mathbf{x}) \leq 2\alpha n\}).$$

Si noti che

$$\begin{aligned} P(\{\mathbf{x} : 2\beta n \leq t(\mathbf{x}) \leq 2\alpha n\}) &= P(\{\mathbf{x} : 2\beta n < t(\mathbf{x}) \leq 2\alpha n\}) \\ &= P(\{\mathbf{x} : t(\mathbf{x}) \leq 2\alpha n\} \setminus \{\mathbf{x} : t(\mathbf{x}) \leq 2\beta n\}) \\ &= P(\{\mathbf{x} : t(\mathbf{x}) \leq 2\alpha n\}) - P(\{\mathbf{x} : t(\mathbf{x}) \leq 2\beta n\}) = \rho_n(\alpha) - \rho_n(\beta). \end{aligned}$$

Dunque, per calcolare il limite di (2.19) è sufficiente determinare, per $\alpha \in [0, 1]$, il limite $\lim_{n \rightarrow +\infty} \rho_n(\alpha)$.

Teorema 2.14. (Legge dell'arcoseno) *Per ogni $\alpha \in [0, 1]$ si ha*

$$\lim_{n \rightarrow +\infty} \rho_n(\alpha) = \frac{2}{\pi} \arcsin(\sqrt{\alpha}).$$

Dimostrazione. In questa dimostrazione useremo il fatto, dimostrato nel paragrafo precedente,

$$\lim_{n \rightarrow +\infty} \frac{u_{2n}}{1/\sqrt{n}} = \frac{\sqrt{2}}{C}, \quad (2.20)$$

dove C è la costante che appare nella dimostrazione della Formula di Stirling. In realtà, come fatto collaterale, dimostreremo qui che $C = \sqrt{2\pi}$, completando così la dimostrazione della Formula di Stirling.

Sia, per il momento, $\frac{1}{2} < \alpha < 1$. Abbiamo

$$\rho_n(\alpha) - \rho_n(1/2) = \sum_{n \leq 2k \leq 2\alpha n} P_{2k,2n} = \sum_{n \leq 2k \leq 2\alpha n} u_{2k} u_{2n-2k}.$$

Per (2.20), per ogni $\epsilon > 0$ esiste n_0 tale che per ogni $m \geq n_0$

$$(1 - \epsilon) \frac{\sqrt{2}}{C\sqrt{m}} \leq u_{2m} \leq (1 + \epsilon) \frac{\sqrt{2}}{C\sqrt{m}}.$$

Si osservi che se $n \leq 2k \leq 2\alpha n$, allora $k \geq \frac{n}{2}$ e $n - k \geq (1 - \alpha)n$. Essendo $\alpha < 1$, se n è sufficientemente grande abbiamo $k \geq n_0$, $n - k \geq n_0$. Pertanto, per ogni $n \leq 2k \leq 2\alpha n$

$$(1 - \epsilon)^2 \frac{2}{C^2 \sqrt{k(n-k)}} \leq u_{2k} u_{2n-2k} \leq (1 + \epsilon)^2 \frac{2}{C^2 \sqrt{k(n-k)}}.$$

Sommando su k e osservando che $\frac{1}{\sqrt{k(n-k)}} = \frac{1}{n} \frac{1}{\sqrt{\frac{k}{n}(1-\frac{k}{n})}}$, abbiamo

$$(1 - \epsilon)^2 \frac{2}{C^2} \frac{1}{n} \sum_{\frac{n}{2} \leq k \leq \alpha n} \frac{1}{\sqrt{\frac{k}{n}(1-\frac{k}{n})}} \leq \rho_n(\alpha) - \rho_n(1/2) \leq (1 + \epsilon)^2 \frac{2}{C^2} \frac{1}{n} \sum_{\frac{n}{2} \leq k \leq \alpha n} \frac{1}{\sqrt{\frac{k}{n}(1-\frac{k}{n})}}. \quad (2.21)$$

Si osservi che

$$\frac{1}{n} \sum_{\frac{n}{2} \leq k \leq \alpha n} \frac{1}{\sqrt{\frac{k}{n}(1-\frac{k}{n})}}$$

è una *somma di Riemann* per l'integrale

$$\int_{\frac{1}{2}}^{\alpha} \frac{1}{\sqrt{x(1-x)}} dx = 2 \left[\arcsin(\sqrt{\alpha}) - \frac{\pi}{4} \right].$$

Pertanto, da (2.21)

$$\lim_{n \rightarrow +\infty} [\rho_n(\alpha) - \rho_n(1/2)] = \frac{2}{C^2} \int_{\frac{1}{2}}^{\alpha} \frac{1}{\sqrt{x(1-x)}} dx = \frac{4}{C^2} \left[\arcsin(\sqrt{\alpha}) - \frac{\pi}{4} \right]. \quad (2.22)$$

Facciamo ora alcune considerazioni. Usando la semplice relazione di simmetria $P_{2k,2n} = P_{2(n-k),2n}$ e il fatto che $\sum_{k=0}^n P_{2k,2n} = 1$, si ha

$$\begin{aligned} \rho_n(\alpha) &= \sum_{k \leq \alpha n} P_{2k,2n} = \sum_{k \leq \alpha n} P_{2(n-k),2n} \\ &= 1 - \sum_{k > \alpha n} P_{2(n-k),2n} = 1 - \sum_{k < (1-\alpha)n} P_{2k,2n} = \begin{cases} 1 - \rho_n(1-\alpha) & \text{se } \alpha n \text{ non è intero} \\ 1 - \rho_n(1-\alpha) + P_{2\alpha n,2n} & \text{se } \alpha n \text{ è intero.} \end{cases} \end{aligned} \quad (2.23)$$

Inoltre, come abbiamo già osservato in precedenza, $P_{2\alpha n,2n}$ tende a 0 se n tende all'infinito. Perciò, segue da (2.23) che

$$\lim_{n \rightarrow +\infty} [\rho_n(\alpha) + \rho_n(1-\alpha)] = 1. \quad (2.24)$$

Posto $\alpha = 1/2$, (2.24) implica che

$$\lim_{n \rightarrow +\infty} \rho_n(1/2) = 1/2. \quad (2.25)$$

Segue da (2.22) e (2.25) che il limite $\lim_{n \rightarrow +\infty} \rho_n(\alpha)$ esiste, e

$$\lim_{n \rightarrow +\infty} \rho_n(\alpha) = \frac{1}{2} + \frac{4}{C^2} \left[\arcsin(\sqrt{a}) - \frac{\pi}{4} \right]. \quad (2.26)$$

Ci è rimasto il problema di identificare il valore di c . Osserviamo che la funzione $\alpha \mapsto \rho_n(\alpha)$ è nondecrecente. Quindi anche la funzione $\alpha \mapsto \lim_{n \rightarrow +\infty} \rho_n(\alpha)$ è nondecrecente, e perciò ammette limite per $\alpha \uparrow 1$. Inoltre, per ogni $\alpha < 1$, $\rho_n(\alpha) < 1$. Ne segue che

$$\lim_{\alpha \uparrow 1} \lim_{n \rightarrow +\infty} \rho_n(\alpha) \leq 1. \quad (2.27)$$

Vogliamo mostrare che tale limite è proprio uguale a 1. Procediamo come segue. Notare che

$$1 - \rho_n(\alpha) = \sum_{\alpha n < k \leq n} P_{2k, 2n} = \sum_{\alpha n < k \leq n} u_{2k} u_{2n-2k}. \quad (2.28)$$

Abbiamo già osservato che, per la formula di Stirling,

$$u_{2n} \sim \frac{\sqrt{2}}{C\sqrt{n}},$$

cioè che la successione $\frac{C\sqrt{n}u_{2n}}{\sqrt{2}}$ tende a 1. Poiché le successioni convergenti sono limitate, deduciamo che esiste una costante $c > 0$ tale che $\sqrt{n}u_{2n} \leq c$, cioè

$$u_{2n} \leq \frac{c}{\sqrt{n}} \quad (2.29)$$

per ogni n . Sostituendo (2.29) in (2.28) e tenendo conto del fatto che $u_0 = 1$, otteniamo

$$1 - \rho_n(\alpha) \leq c \sum_{\alpha n < k < n} \frac{1}{\sqrt{k(n-k)}} + u_{2n}. \quad (2.30)$$

Facciamo ora tendere n all'infinito in (2.30). Ripetendo l'argomento usato per mostrare (2.22) e tenuto conto che $u_{2n} \rightarrow 0$, otteniamo

$$1 - \lim_{n \rightarrow +\infty} \rho_n(\alpha) \leq c \int_{\alpha}^1 \frac{1}{x(1-x)} dx. \quad (2.31)$$

Poiché

$$\lim_{\alpha \uparrow 1} \int_{\alpha}^1 \frac{1}{x(1-x)} dx = 0,$$

otteniamo da (2.31)

$$1 - \lim_{\alpha \uparrow 1} \lim_{n \rightarrow +\infty} \rho_n(\alpha) \leq 0,$$

cioè

$$\lim_{\alpha \uparrow 1} \lim_{n \rightarrow +\infty} \rho_n(\alpha) \geq 1. \quad (2.32)$$

Mettendo assieme (2.32) e (2.27) otteniamo

$$\lim_{\alpha \uparrow 1} \lim_{n \rightarrow +\infty} \rho_n(\alpha) = 1,$$

come desiderato. A questo punto prendiamo il limite $\alpha \uparrow 1$ in (2.26); poiché $\arcsin(\sqrt{\alpha}) \rightarrow \arcsin(1) = \pi/2$, otteniamo

$$1 = \frac{1}{2} + \frac{4}{C^2} \frac{\pi}{4},$$

da cui $C = \sqrt{2\pi}$. Sostituendo in (2.26), otteniamo, per $\frac{1}{2} \leq \alpha < 1$

$$\lim_{n \rightarrow +\infty} \rho_n(\alpha) = \frac{2}{\pi} \arcsin(\sqrt{\alpha}). \quad (2.33)$$

Per estendere (2.33) anche per $\alpha < 1/2$, basta usare (2.24). \square

2.2.4 I teoremi limite di De Moivre

I risultati di questo paragrafo hanno lo scopo di determinare i valori più probabili per s_n , e il comportamento asintotico per $n \rightarrow +\infty$ della probabilità che tali valori vengano assunti.

I due seguenti risultati sono quelli che vengono oggi chiamati i Teoremi Limite di De Moivre. Una parte consistente della moderna Teoria della Probabilità ha lo scopo di generalizzare questi risultati e di analizzare le numerose conseguenze teoriche e applicative.

Teorema 2.15. (*Teorema limite locale di De Moivre*) Sia $A > 0$, $n \geq 1$ e m un numero intero tale che $|m| \leq A\sqrt{n}$. Sia $\rho(x) := \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$. Allora esiste una costante $C > 0$, dipendente solo da A (e non da m e n), tale che

$$\left| \frac{P(s_{2n} = 2m)}{\sqrt{\frac{2}{n}} \rho(2m/\sqrt{2n})} - 1 \right| \leq \frac{C}{\sqrt{n}}, \quad (2.34)$$

e

$$\left| \frac{P(s_{2n+1} = 2m+1)}{\sqrt{\frac{2}{2n+1}} \rho((2m+1)/\sqrt{2n+1})} - 1 \right| \leq \frac{C}{\sqrt{n}}, \quad (2.35)$$

Dimostrazione. Dimostriamo solo (2.34), poiché (2.35) si mostra in modo del tutto analogo. Useremo ancora la Formula di Stirling vista in 2.7:

$$\log N! = N \log \frac{N}{e} + \frac{1}{2} \log(2\pi N) + \frac{\theta(N)}{12N}, \quad (2.36)$$

dove $0 \leq \theta(N) \leq 1$. Osserviamo ora che è sufficiente dimostrare (2.34) solo per valori di n abbastanza grandi (perché?). In particolare, non è restrittivo supporre che $A\sqrt{n} \leq n/2$. Scriviamo ora, usando (2.4)

$$\log P(s_{2n} = 2m) = -2n \log 2 + \log((2n)!) - \log((n+m)!) - \log((n-m)!).$$

L'idea è ora di applicare la formula di Stirling (2.36) a tutti i fattoriali nella formula precedente:

$$\begin{aligned} \log P(s_{2n} = 2m) &= -2n \log 2 + 2n \log \frac{2n}{e} + \frac{1}{2} \log(4\pi n) + \frac{\theta(2n)}{24n} \\ &\quad - (n+m) \log \frac{n+m}{e} - \frac{1}{2} \log(2\pi(n+m)) - \frac{\theta(n+m)}{12(n+m)} \\ &\quad - (n-m) \log \frac{n-m}{e} - \frac{1}{2} \log(2\pi(n-m)) - \frac{\theta(n-m)}{12(n-m)}. \end{aligned}$$

Con facili semplificazioni si ha pertanto

$$\log P(s_{2n} = 2m) = (n+m) \log \frac{n}{n+m} + (n-m) \log \frac{n}{n-m} + \frac{1}{2} \log \frac{n}{\pi(n^2-m^2)} + R_1(n, m), \quad (2.37)$$

dove $R_1(n, m) = \frac{\theta(2n)}{24n} - \frac{\theta(n+m)}{12(n+m)} - \frac{\theta(n-m)}{12(n-m)}$. Poichè, come osservato prima, non è restrittivo assumere $m \leq n/2$, abbiamo che, per un'opportuna costante $c_1 > 0$,

$$|R_1(n, m)| \leq \frac{c_1}{n}. \quad (2.38)$$

Ora dobbiamo trattare i vari termini di (2.37). Anzitutto

$$0 \leq \frac{1}{2} \log \frac{n}{\pi(n^2-m^2)} - \frac{1}{2} \log \frac{1}{\pi n} = \frac{1}{2} \log \left(1 + \frac{m^2}{n^2-m^2} \right) \leq \frac{1}{2} \frac{m^2}{n^2-m^2} \leq \frac{c_2}{n}, \quad (2.39)$$

per un'opportuna costante c_2 , dove abbiamo usato il fatto che, per ogni $x > 0$, $\log(1+x) \leq x$. Dunque, per (2.37), (2.38) e (2.39), possiamo scrivere

$$\log P(s_{2n} = 2m) = (n+m) \log \frac{n}{n+m} + (n-m) \log \frac{n}{n-m} + \frac{1}{2} \log \frac{1}{\pi n} + R_2(n, m), \quad (2.40)$$

con

$$|R_2(n, m)| \leq \frac{c_3}{n} \quad (2.41)$$

per qualche $c_3 > 0$. Per trattare i primi due addendi del membro destro di (2.40), consideriamo la funzione, definita per $x \in (0, 1)$

$$g(x) = x \log 2x + (1-x) \log 2(1-x).$$

Si noti che $g(1/2) = g'(1/2) = 0$, e $g''(1/2) = 4$. Quindi, usando la formula di Taylor con, ad esempio, resto di Lagrange, si ha che per $x \in [1/4, 3/4]$

$$g(x) = 2 \left(x - \frac{1}{2} \right)^2 + r(x), \quad (2.42)$$

dove

$$|r(x)| \leq c_4 \left| x - \frac{1}{2} \right|^3. \quad (2.43)$$

Ora abbiamo

$$(n+m) \log \frac{n}{n+m} + (n-m) \log \frac{n}{n-m} = -2ng \left(\frac{n+m}{2n} \right).$$

Come si è visto prima, possiamo assumere $|m| \leq n/2$, cioè $\frac{n+m}{2n} \in [1/4, 3/4]$. Perciò, da (2.42),

$$(n+m) \log \frac{n}{n+m} + (n-m) \log \frac{n}{n-m} = -4n \left(\frac{m}{2n} \right)^2 + R_3(n, m),$$

dove, da (2.43),

$$|R_3(n, m)| = 2n \left| r \left(\frac{n+m}{2n} \right) \right| \leq 2c_4 n \left(\frac{m}{2n} \right)^3 \leq \frac{c_5}{\sqrt{n}}.$$

Sostituendo in (2.40), otteniamo

$$\log P(s_{2n} = 2m) = -\frac{1}{2} \left(\frac{2m}{\sqrt{2n}} \right)^2 + \frac{1}{2} \log \frac{1}{\pi n} + R_4(n, m), \quad (2.44)$$

con $|R_4(n, m)| \leq \frac{c_6}{\sqrt{n}}$ per un opportuno $c_6 > 0$. Usando il fatto che allora (perché?)

$$\left| e^{R_4(n, m)} - 1 \right| \leq \frac{c_7}{\sqrt{n}},$$

esponenziando ambo i membri di (2.44) si ricava (2.34) immediatamente. \square

Il teorema limite locale di De Moivre fornisce pertanto il comportamento asintotico di $P(s_{2n} = 2m)$ per $m \in [-A\sqrt{n}, A\sqrt{n}]$. Si noti che da (2.34) segue che per tutti gli m in un tale intervallo la probabilità $P(s_{2n} = 2m)$ ha lo stesso ordine di infinitesimo $O(1/\sqrt{n})$.

Il seguente teorema è la prima versione di quello che è probabilmente il più importante teorema del calcolo delle probabilità, e per questo viene chiamato *centrale*. Si tratta di un corollario abbastanza semplice del teorema limite locale e implica, tra l'altro (vedi Osservazione 2.17) che la probabilità che $s_{2n} \notin [-A\sqrt{n}, A\sqrt{n}]$ è piccola *uniformemente* in n se A è sufficientemente grande. Questo si può esprimere dicendo che gli scostamenti *tipici* da 0 di s_{2n} sono di ordine \sqrt{n} , mentre scostamenti di ordine maggiore sono estremamente improbabili.

Teorema 2.16. (*Teorema limite centrale di De Moivre*) *Siano $a, b \in \mathbb{R}$ con $a < b$. Allora*

$$\lim_{n \rightarrow +\infty} P(s_n \in [a\sqrt{n}, b\sqrt{n}]) = \int_a^b \rho(x) dx. \quad (2.45)$$

Dimostrazione. Come per il Teorema 2.15, si distingue i casi di n pari e n dispari. Trattiamo nel dettaglio il caso di n pari, lasciando al lettore l'altro caso, peraltro del tutto analogo. Mostriamo dunque (2.45) con $2n$ in luogo di n . Iniziamo con l'osservare che l'espressione

$$\sum_{m: a\sqrt{2n} \leq 2m \leq b\sqrt{2n}} \sqrt{\frac{2}{n}} \rho(2m/\sqrt{2n})$$

è una somma di Riemann per l'integrale $\int_a^b \rho(x) dx$ rispetto ad una suddivisione dell'intervallo $[a, b]$ in intervalli di ampiezza $\sqrt{\frac{2}{n}}$. Pertanto

$$\lim_{n \rightarrow +\infty} \sum_{m: a\sqrt{2n} \leq 2m \leq b\sqrt{2n}} \sqrt{\frac{2}{n}} \rho(2m/\sqrt{2n}) = \int_a^b \rho(x) dx. \quad (2.46)$$

Inoltre, è evidente che l'evento $\{s_n \in [a\sqrt{n}, b\sqrt{n}]\}$ si può decomporre come unione di eventi disgiunti come segue

$$\{s_n \in [a\sqrt{n}, b\sqrt{n}]\} = \bigcup_{m: a\sqrt{2n} \leq 2m \leq b\sqrt{2n}} \{s_{2n} = 2m\},$$

e quindi

$$P(s_n \in [a\sqrt{n}, b\sqrt{n}]) = \sum_{m: a\sqrt{2n} \leq 2m \leq b\sqrt{2n}} P(s_{2n} = 2m). \quad (2.47)$$

Per (2.46) e (2.47), la dimostrazione è completa se mostriamo che

$$\lim_{n \rightarrow +\infty} \sum_{m: a\sqrt{2n} \leq 2m \leq b\sqrt{2n}} \left| P(s_{2n} = 2m) - \sqrt{\frac{2}{n}} \rho(2m/\sqrt{2n}) \right| = 0. \quad (2.48)$$

Da (2.34) segue che esiste una costante $C > 0$ tale che per ogni m per cui $a\sqrt{2n} \leq 2m \leq b\sqrt{2n}$ si ha

$$\left| P(s_{2n} = 2m) - \sqrt{\frac{2}{n}} \rho(2m/\sqrt{2n}) \right| \leq \frac{C}{\sqrt{n}} \sqrt{\frac{2}{n}} \rho(2m/\sqrt{2n}).$$

Notando che il numero di addendi della somma in (2.48) è minore o uguale a $\sqrt{\frac{n}{2}}(b-a)$ e che $\rho(x) \leq 1/\sqrt{2\pi}$ per ogni $x \in \mathbb{R}$, si ha

$$\sum_{m: a\sqrt{2n} \leq 2m \leq b\sqrt{2n}} \left| P(s_{2n} = 2m) - \sqrt{\frac{2}{n}} \rho(2m/\sqrt{2n}) \right| \leq \frac{C(b-a)}{\sqrt{2\pi n}},$$

da cui segue (2.48). □

Osservazione 2.17. È ben noto che

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

Fissiamo $\epsilon > 0$. Esiste $a > 0$ tale che

$$\int_{-a}^a \rho(x) dx \geq 1 - \frac{\epsilon}{2}. \quad (2.49)$$

Inoltre, per il Teorema 2.16, esiste $N > 0$ tale che per $n \geq N$

$$\left| P(|s_n| \leq a\sqrt{n}) - \int_{-a}^a \rho(x) dx \right| \leq \frac{\epsilon}{2}.$$

Usando (2.49) deduciamo che, per $n \geq N$,

$$P(|s_n| > a\sqrt{n}) \leq \epsilon.$$

D'altra parte è ovvio che per ogni $n < N$ si ha che $P(|s_n| > \sqrt{N}\sqrt{n}) = 0$. Quindi, posto $A = \max(a, \sqrt{N})$, abbiamo

$$P(|s_n| > A\sqrt{n}) \leq \epsilon.$$

per ogni $n \geq 1$, cioè la probabilità di avere uno scostamento dalla posizione iniziale maggiore di $A\sqrt{n}$ è piccola *uniformemente* rispetto a n .

2.3 Il modello di Ising in meccanica statistica

Descriviamo ora un celebre modello in meccanica statistica, il *modello di Ising* per un materiale ferromagnetico. Sia Λ un sottoinsieme finito di \mathbb{Z}^d . I punti di Λ vanno interpretati come i *nodi* (detti anche *siti*) di un cristallo regolare. Ogni nodo è occupato da un atomo, il moto dei cui elettroni produce un campo magnetico. In questo modello semplificato, si assume che tale

campo magnetico, che chiameremo *spin*, assuma solo due valori, $+1$ e -1 . Una *configurazione* per tale sistema è

$$\sigma = (\sigma_x)_{x \in \Lambda},$$

dove $\sigma_x = \pm 1$ è lo spin nel nodo $x \in \Lambda$. In altre parole $\Omega = \{-1, 1\}^\Lambda$ è l'insieme di tutte le configurazioni.

L'energia (potenziale) associata ad una configurazione è dovuta all'interazione tra gli spin dei nodi in Λ e all'interazione con l'esterno. In questa presentazione, assumiamo che l'interazione sia *locale*: l'interazione tra gli spin in Λ avviene solo tra siti *primi vicini*, la cui distanza è pari a 1, mentre l'interazione con l'esterno riguarda solo i nodi del "bordo" di Λ , cioè $\partial\Lambda = \{x \in \Lambda : \exists y \in \Lambda^c \text{ tale che } |x - y| = 1\}$. Più precisamente, l'energia (o Hamiltoniana) di una configurazione σ è data da

$$H_\Lambda^\tau(\sigma) := - \sum_{\substack{x, y \in \Lambda \\ |x - y| = 1}} \sigma_x \sigma_y - \sum_{x \in \partial\Lambda} \tau_x \sigma_x. \quad (2.50)$$

Si noti che il primo termine in H_Λ^τ descrive l'interazione tra gli spin primi vicini in Λ , mentre il secondo termine può essere interpretato come risultante da campi magnetici di valore τ_x agenti sugli spin del bordo di Λ . Assumiamo per semplicità che $\tau_x = \pm 1$, cioè $\tau \in \{-1, 1\}^{\partial\Lambda}$.

Si noti che se fosse $\tau_x \equiv 0$, l'energia H_Λ^τ avrebbe esattamente due minimi assoluti, dati rispettivamente da $\sigma_x \equiv 1$ e $\sigma_x \equiv -1$; più in generale, l'energia di una configurazione σ sarebbe uguale a quella della configurazione $-\sigma$. La presenza di un campo magnetico al bordo $\tau \in \{-1, 1\}^{\partial\Lambda}$, rompe tale simmetria: in particolare, se $\tau_x \equiv +1$, l'unico minimo di H_Λ^τ è la configurazione con $\sigma_x \equiv +1$. In ogni caso, una configurazione ha un valore tanto più basso dell'energia quanto più gli spin della configurazione sono allineati tra di loro.

Una quantità che gioca un ruolo fisico fondamentale è la temperatura. Il "moto termico" degli atomi si traduce in un "disturbo aleatorio" sugli spin: il sistema ha una "preferenza" per le configurazioni a bassa energia, ma tale preferenza è tanto più debole tanto più è alta la temperatura. Queste considerazioni intuitive hanno una traduzione precisa in meccanica statistica considerando la *misura di Gibbs* associata all'Hamiltoniana H_Λ , descritta nell'Esempio 1.6, che ora riprendiamo in dettaglio. Se T è la temperatura assoluta, indichiamo con $\beta = 1/(k_B T)$ la temperatura inversa, dove k_B è la costante di Boltzmann. È conveniente semplificare le notazioni ponendo $k_B = 1$ (il che equivale a misurare la temperatura in unità di k_B), di modo che $\beta = 1/T$. Secondo l'ipotesi di Gibbs, se il sistema è in equilibrio ad una temperatura inversa $\beta > 0$ con campo magnetico al bordo τ , la probabilità di osservare una configurazione di spin σ è data da

$$\mu_{\Lambda, \beta}^\tau(\{\sigma\}) := \frac{1}{Z_\Lambda^\tau} \exp[-\beta H_\Lambda^\tau],$$

dove

$$Z_{\Lambda, \beta}^\tau := \sum_{\sigma \in \Omega} \exp[-\beta H_\Lambda^\tau].$$

In questo modo $\sum_{\sigma \in \Omega} \mu_{\Lambda, \beta}^\tau(\{\sigma\}) = 1$. Dunque, come in (1.4), $\mu_{\Lambda, \beta}^\tau$ si può estendere ad una probabilità su Ω ponendo, per $A \subset \Omega$:

$$\mu_{\Lambda, \beta}^\tau(A) := \sum_{\sigma \in A} \mu_{\Lambda}^\tau(\{\sigma\}).$$

Come abbiamo visto nell'Esempio 1.6, per ogni $\sigma \in \Omega$ si ha che

$$\lim_{\beta \rightarrow 0} \mu_{\Lambda, \beta}^{\tau}(\{\sigma\}) = \frac{1}{|\Omega|},$$

cioè, nel limite di temperatura infinita, tutte le configurazioni diventano equiprobabili. Inoltre, per ogni $\sigma \in \Omega$ che non sia un minimo assoluto di H_{Λ}^{τ} si ha

$$\lim_{\beta \rightarrow +\infty} \mu_{\Lambda}^{\tau}(\{\sigma\}) = 0,$$

cioè, nel limite di temperatura zero, il sistema tende a “congelarsi” nelle configurazioni che minimizzano l'energia.

Supponiamo ora di fissare il reticolo $\Lambda = \Lambda_n = \{-n, -n+1, \dots, 0, \dots, n-1, n\}^d$ e le condizioni al bordo $\tau_x \equiv 1$. Poniamo $\Omega_n := \{-1, 1\}^{\Lambda_n}$ e scriveremo $\mu_{n, \beta}^+$ in luogo di $\mu_{\Lambda_n, \beta}^{\tau}$, $Z_{n, \beta}^+$ in luogo di $Z_{\Lambda_n, \beta}^{\tau}$ e H_n^+ in luogo di $H_{\Lambda_n}^{\tau}$. Introduciamo l'evento

$$A := \{\sigma \in \Omega_n : \sigma_0 = +1\}, \quad (2.51)$$

dove $0 = (0, 0, \dots, 0)$ indica l'origine in \mathbb{Z}^d . Quindi $\mu_{n, \beta}^+(A)$ è la probabilità che lo spin nell'origine sia positivo. Come suggerito in precedenza, la presenza al bordo di un campo magnetico positivo “favorisce” gli spin positivi rispetto a quelli negativi: di conseguenza, è intuitivamente plausibile che si abbia

$$\mu_{n, \beta}^+(A) > 1/2.$$

Questa disuguaglianza è effettivamente vera, per ogni valore fissato di $n \in \mathbb{N}$ e $\beta > 0$, ma ne omettiamo per brevità la dimostrazione. Essa esprime il fatto che lo spin nell'origine “risente” del campo magnetico al bordo, e, con probabilità $> 1/2$, si allinea allo stesso. Facendo crescere n , aumenta la distanza tra l'origine e il campo magnetico al bordo, la cui influenza, si può congetturare, diventi sempre meno rilevante. In altre parole, potrebbe accadere che

$$\lim_{n \rightarrow +\infty} \mu_{n, \beta}^+(A) = 1/2. \quad (2.52)$$

Viceversa, potrebbe accadere che l'influenza del campo magnetico al bordo sull'origine sia rilevante anche per n grande, ossia che esista $\epsilon > 0$ tale che, *per ogni* n ,

$$\mu_{n, \beta}^+(A) > \frac{1}{2} + \epsilon. \quad (2.53)$$

Se accade (2.53), si dice che (per il valore di β dato) si ha *magnetizzazione spontanea*.

Per i ferromagneti reali, la magnetizzazione spontanea è un fenomeno effettivamente osservato, purché la temperatura sia non troppo elevata. Il problema che ci poniamo è di stabilire se il modello di Ising, almeno per questo aspetto, è un buon modello per un ferromagnete reale. Il risultato interessante è che la risposta dipende dalla dimensione d dello spazio: per $d \geq 2$ si ha magnetizzazione spontanea a basse temperature, mentre per $d = 1$ non si ha magnetizzazione spontanea per nessun valore della temperatura. Non dovrebbe sorprendere il fatto che l'analisi del modello diventa via via più difficile al crescere della dimensione d . Benché la dimensione “fisica” sia $d = 3$, ci occuperemo per semplicità solo dei casi $d = 1$ e $d = 2$.

2.3.1 Il caso $d = 1$

Per $d = 1$ si ha $\Lambda_n = \{-n, -n+1, \dots, n-1, n\}$ e $\Omega_n = \{-1, 1\}^{\Lambda_n}$. Con un conto esplicito possiamo scrivere:

$$\begin{aligned}\mu_{n,\beta}^+(A) &= \sum_{\sigma \in A} \mu_{n,\beta}^+(\{\sigma\}) = \frac{1}{Z_{n,\beta}^+} \sum_{\sigma \in \Omega_n: \sigma_0=1} \exp \left[\beta \left(\sigma_{-n} + \sum_{k=-n}^{n-1} \sigma_k \sigma_{k+1} + \sigma_n \right) \right] \\ &= \frac{1}{Z_{n,\beta}^+} \sum_{\substack{\sigma_{-n}, \dots, \sigma_{-1} \\ \sigma_1, \dots, \sigma_n}} \exp \left[\beta \left(\sigma_{-n} + \sum_{k=-n}^{-2} \sigma_k \sigma_{k+1} + \sigma_{-1} \right) \right] \exp \left[\beta \left(\sigma_1 + \sum_{k=1}^{n-1} \sigma_k \sigma_{k+1} + \sigma_n \right) \right] \\ &= \frac{1}{Z_{n,\beta}^+} \left(\sum_{\sigma_{-n}, \dots, \sigma_{-1}} e^{\beta(\sigma_{-n} + \sum_{k=-n}^{-2} \sigma_k \sigma_{k+1} + \sigma_{-1})} \right) \left(\sum_{\sigma_1, \dots, \sigma_n} e^{\beta(\sigma_1 + \sum_{k=1}^{n-1} \sigma_k \sigma_{k+1} + \sigma_n)} \right).\end{aligned}$$

Si noti ora che le due somme contenute in quest'ultima espressione sono uguali, cambiando solo i nomi delle variabili sommate ($\sigma_i \leftrightarrow \sigma_{-i}$). Dunque:

$$\mu_{n,\beta}^+(A) = \frac{1}{Z_{n,\beta}^+} \left\{ \sum_{\sigma_1, \dots, \sigma_n} \exp \left[\beta \left(\sigma_1 + \sum_{k=1}^{n-1} \sigma_k \sigma_{k+1} + \sigma_n \right) \right] \right\}^2. \quad (2.54)$$

Per semplificare questa espressione, introduciamo un operatore lineare T , che agisce sullo spazio vettoriale delle funzioni f da $\{-1, +1\}$ in \mathbb{R} nel modo seguente: la funzione Tf , sempre da $\{-1, +1\}$ in \mathbb{R} , è definita da

$$(Tf)(s) := \sum_{s'=\pm 1} e^{\beta s s'} f(s') = e^{\beta s} f(1) + e^{-\beta s} f(-1).$$

Una funzione $f : \{-1, +1\} \rightarrow \mathbb{R}$ può essere identificata con il vettore colonna $\begin{pmatrix} f(-1) \\ f(+1) \end{pmatrix}$. In questo modo, la trasformazione $f \rightarrow Tf$ corrisponde alla trasformazione lineare sui vettori di dimensione due data dalla matrice $T_{s,s'} := e^{\beta s s'}$, cioè

$$T = \begin{pmatrix} e^{\beta} & e^{-\beta} \\ e^{-\beta} & e^{\beta} \end{pmatrix}. \quad (2.55)$$

Posto $\varphi(s) = e^{\beta s}$, possiamo riscrivere la relazione (2.54) come

$$\mu_{n,\beta}^+(A) = \frac{1}{Z_{n,\beta}^+} \left[\sum_{\sigma_1, \dots, \sigma_n} T_{1,\sigma_1} \cdot T_{\sigma_1,\sigma_2} \cdots T_{\sigma_{n-1},\sigma_n} \cdot \varphi(\sigma_n) \right]^2 = \frac{1}{Z_{n,\beta}^+} [(T^n \varphi)(1)]^2. \quad (2.56)$$

Si osservi che, se avessimo voluto calcolare $\mu_n^+(A^c)$, avremmo dovuto sommare sulle configurazioni per le quali $\sigma_0 = -1$. L'unica differenza, rispetto all'espressione in (2.54), è che l'addendo σ_1 nell'esponenziale sarebbe stato sostituito da $-\sigma_1$. Usando le notazioni or ora introdotte, possiamo concludere che

$$\mu_{n,\beta}^+(A^c) = \frac{1}{Z_{n,\beta}^+} [(T^n \varphi)(-1)]^2.$$

Quest'ultima uguaglianza, assieme a (2.56) e al fatto che $\mu_{n,\beta}^+(A) + \mu_{n,\beta}^+(A^c) = 1$, ci dà

$$Z_{n,\beta}^+ = [(T^n \varphi)(1)]^2 + [(T^n \varphi)(-1)]^2.$$

Mettendo tutto assieme:

$$\mu_{n,\beta}^+(A) = \frac{[(T^n \varphi)(1)]^2}{[(T^n \varphi)(1)]^2 + [(T^n \varphi)(-1)]^2}. \quad (2.57)$$

Usiamo ora un po' di algebra lineare. La matrice T definita in (2.55) ha come autovalori $\lambda_1 = 2 \cosh(\beta)$, $\lambda_2 = 2 \sinh(\beta)$, corrispondenti agli autovettori $v_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ e $v_2 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$. Identificando vettori e funzioni come sopra indicato, possiamo esprimere la funzione φ come

$$\varphi = \cosh(\beta) v_1 - \sinh(\beta) v_2,$$

da cui, usando la linearità di T , si ottiene

$$(T^n \varphi)(\pm 1) = 2^n \cosh^{n+1}(\beta) \pm 2^n \sinh^{n+1}(\beta).$$

In conclusione, abbiamo calcolato l'espressione *esatta* di $\mu_{n,\beta}^+(A)$:

$$\mu_{n,\beta}^+(A) = \frac{[\cosh^{n+1}(\beta) + \sinh^{n+1}(\beta)]^2}{[\cosh^{n+1}(\beta) + \sinh^{n+1}(\beta)]^2 + [\cosh^{n+1}(\beta) - \sinh^{n+1}(\beta)]^2}.$$

Si noti che effettivamente $\mu_{n,\beta}^+(A) > \frac{1}{2}$, per ogni valore fissato di $n \in \mathbb{N}$ e $\beta > 0$ (mentre $\mu_{n,\beta}^+(A) = \frac{1}{2}$ per ogni $n \in \mathbb{N}$, se $\beta = 0$). Dato che $\cosh(\beta) > \sinh(\beta) > 0$ per ogni $\beta > 0$, lasciamo al lettore il compito di dedurre dalla formula precedente che, per ogni $\beta > 0$, si ha

$$\lim_{n \rightarrow +\infty} \mu_{n,\beta}^+(A) = \frac{1}{2}.$$

Questo mostra che in dimensione 1 non c'è magnetizzazione spontanea per nessun $\beta > 0$.

2.3.2 Il caso $d = 2$

In dimensione due non tenteremo di effettuare calcoli esatti con il modello di Ising. In realtà, molti calcoli esatti sono possibili: la loro complessità va però al di là del livello di questo libro. Dimosteremo l'esistenza di magnetizzazione spontanea in $d = 2$ (a basse temperature) mediante un argomento geometrico-combinatorio semplice ed efficace, generalizzabile a molti modelli più complessi: si tratta del celebre *argomento di Peierls*.

In questo caso il reticolo $\Lambda_n = \{-n, -n+1, \dots, n-1, n\}^2$ è formato dai punti a coordinate intere del quadrato di lato $2n$ avente l'origine al centro. Poniamo $\Omega_n := \{-1, 1\}^{\Lambda_n}$. Per comodità di calcolo, conviene modificare leggermente la definizione dell'energia, ponendo

$$\tilde{H}_n^+(\sigma) := - \sum_{\substack{x,y \in \Lambda_n \\ |x-y|=1}} (\sigma_x \sigma_y - 1) - \sum_{x \in \partial \Lambda_n} (\sigma_x - 1). \quad (2.58)$$

Si noti che, con riferimento all'energia originale H_n^+ definita in (2.50), per ogni $\sigma \in \Omega$ si ha $\tilde{H}_n^+(\sigma) = H_n^+(\sigma) + c_n$, dove $c_n = |\{x, y \in \Lambda_n : |x - y| = 1\}| + |\partial\Lambda_n|$ è una costante che non dipende da σ . Di conseguenza possiamo scrivere

$$\mu_{n,\beta}^+(\sigma) = \frac{1}{Z_{n,\beta}^+} \exp[-\beta H_n^+(\sigma)] = \frac{1}{\tilde{Z}_{n,\beta}^+} \exp[-\beta \tilde{H}_n^+(\sigma)],$$

dove

$$\tilde{Z}_{n,\beta}^+ := \sum_{\sigma \in \Omega_n} \exp[-\beta \tilde{H}_n^+(\sigma)]. \quad (2.59)$$

In altre parole, la nuova energia \tilde{H}_n^+ è fisicamente equivalente a quella originale, cioè determina la stessa misura di Gibbs. La ragione per introdurre l'energia \tilde{H}_n^+ è che essa si può riscrivere come

$$\tilde{H}_n^+(\sigma) = 2|\{(x, y) : |x - y| = 1 \text{ e } \sigma_x \neq \sigma_y\}| + 2|\{x \in \partial\Lambda_n : \sigma_x = -1\}|. \quad (2.60)$$

Vedremo tra poco l'utilità di tale espressione.

Data una configurazione di spin σ su Λ_n , completiamola ad una configurazione di spin su tutto \mathbb{Z}^2 assegnando spin $+1$ a tutti i punti esterni a Λ_n (in realtà, per la costruzione che segue, è sufficiente assegnare spin $+1$ ai punti di Λ_n^c che distano 1 da qualche punto di Λ_n). Per ogni coppia di punti x, y tali che $|x - y| = 1$ e $\sigma_x \neq \sigma_y$ disegniamo quindi nel piano un segmento di lunghezza 1, ortogonale al segmento congiungente x e y , il cui punto medio sia $\frac{x+y}{2}$. La figura ottenuta dall'unione di tutti i segmenti disegnati è detta *contour* (si veda la Figura 2.1 per un esempio).

Introduciamo una notazione importante: definiamo *poligonale chiusa autoevitante* (p.c.a.) l'unione $\bigcup_{i=1}^{k-1} \overline{P_i P_{i+1}}$ dei segmenti che congiungono in successione k punti P_1, \dots, P_k del piano, dove $P_k = P_1$ (poligonale chiusa), $P_i \neq P_j$ se $\{i, j\} \neq \{1, k\}$ (poligonale autoevitante) e inoltre $P_i = (x_i \pm \frac{1}{2}, y_i \pm \frac{1}{2})$ con $(x_i, y_i) \in \Lambda_n$ e $|P_{i+1} - P_i| = 1$, per ogni $i = 1, \dots, k$. Un esempio di poligonale chiusa autoevitante è tratteggiato in rosso nella Figura 2.1.

Il punto fondamentale è che ogni contour \mathcal{C} si può sempre scrivere come unione $\bigcup_{i=1}^m \gamma_i$ di p.a.c. *disgiunte*, dove, con leggero abuso di notazione, intendiamo disgiunte anche due poligonali che si intersecano in un numero finito di punti. Viceversa, un'unione di p.a.c. disgiunte è sempre un contour ammissibile, cioè esiste una configurazione di spin $\sigma \in \Omega_n$ che lo determina, e tale configurazione è unica[†]. Una dimostrazione formale di queste affermazioni non è difficile ma è piuttosto lunga e noiosa e sarà pertanto omessa[‡]. Abbiamo ottenuto una caratterizzazione esplicita dell'insieme dei contour ammissibili, che indicheremo con Ξ_n , che è in corrispondenza biunivoca con lo spazio delle configurazioni Ω_n . Sottolineiamo che la decomposizione $\mathcal{C} = \bigcup_{i=1}^m \gamma_i$ con γ_i p.c.a. disgiunte in generale non è unica (si veda ad esempio la Figura 2.1).

Da qui in poi identificheremo una configurazione σ di spin con il contour \mathcal{C} corrispondente: in particolare, scriveremo $\mu_{n,\beta}^+(\{\mathcal{C}\})$, $\tilde{H}_n^+(\mathcal{C})$, ecc. Definiamo la lunghezza $\ell(\mathcal{C})$ di un contour

[†]La configurazione si costruisce assegnando il valore $+$ (a causa delle condizioni al bordo positive) agli spin "esterni", cioè che non sono racchiusi da alcuna poligonale; quindi assegnando il valore $-$ agli spin dentro le poligonali che "confinano" con spin di valore $+$, e così via.

[‡]Il punto fondamentale è il seguente: per come è costruito un contour, da ogni punto $(x \pm \frac{1}{2}, y \pm \frac{1}{2})$ con $(x, y) \in \Lambda_n$ partono necessariamente 0, 2 oppure 4 segmenti del contour. È questa proprietà che permette di decomporre il contour in poligonali chiuse autoevitanti.

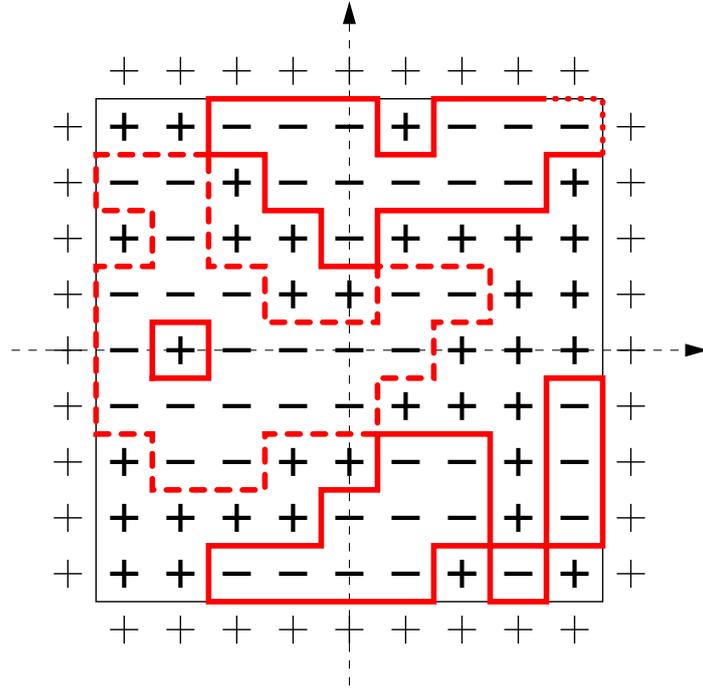


Figura 2.1: Una configurazione di spin σ per il modello di Ising nel piano, sul reticolo $\Lambda_n = \{-n, \dots, n\}^2$ con $n = 4$, con condizioni al bordo positive. In rosso è tracciato il contour \mathcal{C} corrispondente. La parte di contour tratteggiata è una poligonale chiusa autoevitante che contiene l'origine. I due segmenti a puntini, nell'angolo in alto a destra, danno contributo 1 alla lunghezza $\ell(\mathcal{C})$ del contour (che in questo caso è pari a 83).

\mathcal{C} come il numero dei segmenti di lunghezza 1 che lo compongono (lunghezza geometrica), eccetto nel caso in cui una o più coppie di segmenti del contour descrivano uno dei 4 angoli del reticolo Λ_n : in questo caso conveniamo che ciascuna coppia di tali segmenti dia contributo 1 alla lunghezza $\ell(\mathcal{C})$ (si veda la Figura 2.1). Con queste convenzioni, ricordando la relazione (2.60) si ottiene la rappresentazione basilare

$$\tilde{H}_n^+(\mathcal{C}) = 2\ell(\mathcal{C}).$$

Possiamo quindi scrivere

$$\mu_{n,\beta}^+(\{\mathcal{C}\}) = \frac{1}{\tilde{Z}_n^+} e^{-2\beta\ell(\mathcal{C})}, \quad \tilde{Z}_{n,\beta}^+ = \sum_{\mathcal{C}' \in \Xi_n} e^{-2\beta\ell(\mathcal{C}')}. \quad (2.61)$$

Sia ora \mathcal{C} un contour corrispondente a una configurazione di spin σ in cui $\sigma_0 = -1$. Dato che l'“isola” di spin -1 a cui l'origine appartiene dev'essere separata dal “mare” di spin $+1$ che ricopre tutto Λ_n^c , il contour \mathcal{C} dev'essere della forma $\gamma \cup \mathcal{C}'$, dove γ è una p.c.a. avente l'origine al suo interno e \mathcal{C}' è un contour tale che $\gamma \cap \mathcal{C}' = \emptyset$ (o, più precisamente, l'intersezione deve consistere di un numero finito di punti). Ricordando la definizione (2.51) dell'evento A , in termini dei contour vale la seguente inclusione:

$$A^c \subseteq \{\mathcal{C} \in \Xi_n : \mathcal{C} = \gamma \cup \mathcal{C}', \gamma \text{ p.c.a. che racchiude } 0, \mathcal{C}' \in \Xi_n, \gamma \cap \mathcal{C}' = \emptyset\}.$$

Si noti che in generale la decomposizione $\mathcal{C} = \gamma \cup \mathcal{C}'$ non è unica (si veda ancora la Figura 2.1) ma questo non sarà un problema. Osserviamo anche che, essendo γ e \mathcal{C}' disgiunti, si ha chiaramente $\ell(\gamma \cup \mathcal{C}') = \ell(\gamma) + \ell(\mathcal{C}')$. Possiamo dunque scrivere

$$\begin{aligned} \mu_{n,\beta}^+(A^c) &= \frac{1}{\tilde{Z}_{n,\beta}^+} \sum_{\mathcal{C} \in A^c} e^{-2\beta \ell(\mathcal{C})} \leq \frac{1}{\tilde{Z}_{n,\beta}^+} \sum_{\substack{\gamma \text{ p.c.a.} \\ \text{che racchiude } 0}} e^{-2\beta \ell(\gamma)} \sum_{\substack{\mathcal{C}' \in \Xi_n \\ \gamma \cap \mathcal{C}' = \emptyset}} e^{-2\beta \ell(\mathcal{C}')} \\ &\leq \sum_{\substack{\gamma \text{ p.c.a.} \\ \text{che racchiude } 0}} e^{-2\beta \ell(\gamma)} \left(\frac{1}{\tilde{Z}_{n,\beta}^+} \sum_{\mathcal{C}' \in \Xi_n} e^{-2\beta \ell(\mathcal{C}')} \right) = \sum_{\substack{\gamma \text{ p.c.a.} \\ \text{che racchiude } 0}} e^{-2\beta \ell(\gamma)}, \end{aligned}$$

dove si è usata la seconda relazione in (2.61). Disintegrando rispetto ai possibili valori di $\ell(\gamma)$ otteniamo

$$\mu_n^+(A^c) \leq \sum_{m=1}^{\infty} K_m e^{-2\beta m}, \quad \text{dove } K_m := |\{\gamma \text{ p.c.a. che racchiude } 0 : \ell(\gamma) = m\}|.$$

Ci resta da stimare K_m . Se γ è una poligonale chiusa autoevitante con $\ell(\gamma) = m$, la sua lunghezza geometrica è compresa tra m e $m+4$; se γ racchiude 0 al suo interno, necessariamente γ è interamente contenuta nel quadrato $Q := [-\frac{m+4}{2}, \frac{m+4}{2}] \times [-\frac{m+4}{2}, \frac{m+4}{2}]$. Costruiamo ora una curva nel modo seguente:

- scegliamo un punto in Q della forma $(x \pm \frac{1}{2}, y \pm \frac{1}{2})$, con $x, y \in \mathbb{Z}$ (per questa scelta abbiamo al massimo $(\frac{m+4}{2} + \frac{m+4}{2} + 1)^2 = (m+5)^2$ possibilità);
- scegliamo una delle quattro direzioni possibili e tracciamo un segmento di lunghezza uno in quella direzione;
- a questo punto, per il tratto successivo, scegliamo una delle tre direzioni che non ci fanno tornare al punto da cui proveniamo;
- iteriamo la procedura per l passi.

Tra le curve costruite in questo modo in un numero di passi compreso tra m e $m+4$ ci sono in particolare tutte le possibili p.c.a. γ con $\ell(\gamma) = m$. Di conseguenza

$$K_m \leq \sum_{l=m}^{m+4} (m+5)^2 4^l 3^{l-1} \leq 5 \cdot ((m+5)^2 4^3 3^{m+3}).$$

Semplifichiamo questa espressione con la stima (molto rozza) $x^2 \leq 3^x$ per ogni $x \in \mathbb{N}$:

$$K_m \leq C \cdot 9^m, \quad \text{dove } C := 3^8 \cdot 20.$$

In questo modo otteniamo.

$$\mu_{n,\beta}^+(A^c) \leq C \sum_{m=1}^{+\infty} 9^m e^{-2\beta m} = C \sum_{m=1}^{+\infty} e^{-c(\beta)m} = C \frac{e^{-c(\beta)}}{1 - e^{-c(\beta)}},$$

dove abbiamo posto $c(\beta) := 2\beta - \log 9$. Si noti che la stima ottenuta *non dipende da n* . Visto che $\lim_{\beta \rightarrow \infty} c(\beta) = +\infty$ e dato che C è una costante fissata, segue che[†] esiste $\beta_0 \in (0, \infty)$

[†]In effetti abbiamo dimostrato un'affermazione più forte: per ogni $\eta > 0$ esiste $\beta_0(\eta) \in (0, \infty)$ tale che per ogni $\beta > \beta_0(\eta)$ e per ogni $n \in \mathbb{N}$ si ha $\mu_n^+(A) > 1 - \eta$.

tale che per ogni $\beta > \beta_0$ e per ogni $n \in \mathbb{N}$ si ha

$$\mu_{n,\beta}^+(A^c) \leq \frac{1}{4},$$

ovvero

$$\mu_{n,\beta}^+(A) \geq \frac{3}{4}.$$

Abbiamo dunque mostrato che, per grandi valori di β (cioè a temperatura sufficientemente bassa) nel modello di Ising in dimensione 2 ha luogo il fenomeno della magnetizzazione spontanea. È possibile mostrare (non lo faremo) che, al contrario, per valori piccoli di β non si ha magnetizzazione spontanea.

2.4 Il modello di Hardy-Weinberg in genetica

Le nozioni di probabilità condizionata e indipendenza permettono, come vedremo in questo paragrafo, la formulazione e l'analisi rigorosa di un noto modello per la trasmissione ereditaria dei caratteri, in *modello di Hardy*. Consideriamo una popolazione di una determinata specie, e supponiamo di essere interessati ad uno specifico *carattere*. Assumiamo che questo carattere si manifesti in solo due modi, che chiameremo *dominante* e *recessivo*, e che esso sia caratterizzato da una coppia (non ordinata) di *geni*, ognuno dei quali può assumere uno dei due valori a, A . Dunque, ogni individuo della specie possiede una delle tre possibili coppie di geni AA, aA e aa . Nel terzo caso il carattere manifestato sarà quello recessivo, negli altri due quello dominante. Le tre coppie AA, aA e aa sono chiamate *genotipi*.

Il modello di Hardy ha per oggetto l'evoluzione della frequenza dei genotipi in popolazioni sessuate. In un determinato istante consideriamo gli individui della popolazione, che chiameremo di *generazione 0*. Assumiamo che tale popolazione sia numerosa, usiamo le seguenti notazioni:

- $u \in (0, 1)$ è la *frazione* di individui della generazione 0 di genotipo AA ;
- $2v \in (0, 1)$ è la *frazione* di individui della generazione 0 di genotipo aA
- $w \in (0, 1)$ è la *frazione* di individui della generazione 0 di genotipo aa ;

Evidentemente $u + 2v + w = 1$. Tali numeri hanno un'evidente interpretazione probabilistica: se si sceglie a caso un individuo della generazione 0, la probabilità che esso sia di genotipo AA (risp. aA, aa) è u (risp. $2v, w$).

Consideriamo in un istante successivo, gli individui che sono stati generati da una coppia della generazione 0; l'insieme di essi verrà chiamata *generazione 1*. Supponiamo di scegliere a caso un individuo della generazione 1, e consideriamo, per $g \in \{AA, aA, aa\}$, gli eventi

$$D_g := \text{“l'individuo è di genotipo } g\text{”}.$$

Siano

$$\begin{aligned} u_1 &= P(D_{AA}) \\ 2v_1 &= P(D_{aA}) \\ w_1 &= P(D_{aa}). \end{aligned}$$

In altre parole, u_1 , $2v_1$ e w_1 sono le frequenze dei tre genotipi nella generazione 1. Introduciamo anche i seguenti eventi, per $g \in \{AA, aA, aa\}$ e $x \in \{A, a\}$:

$$\begin{aligned} F_{g,x} &:= \text{“l'individuo scelto ha padre di genotipo } g, \text{ e ha da lui ereditato il gene } x\text{”} \\ M_{g,x} &:= \text{“l'individuo scelto ha madre di genotipo } g, \text{ e ha da lei ereditato il gene } x\text{”}. \end{aligned}$$

Inoltre poniamo

$$F_g := \text{“l'individuo scelto ha padre di genotipo } g\text{”} = F_{g,A} \cup F_{g,a},$$

e analogamente per M_g . Il modello di Hardy è definito dalle seguenti ipotesi.

(i) Per ogni scelta di $g, h \in \{AA, aA, aa\}$ e $x, y \in \{A, a\}$, gli eventi $F_{g,x}$ e $M_{h,y}$ sono indipendenti.

(ii)

$$P(F_{AA}) = P(M_{AA}) = u \quad P(F_{aA}) = P(M_{aA}) = 2v \quad P(F_{aa}) = P(M_{aa}) = w.$$

(iii)

$$P(F_{AA,A}|F_{AA}) = P(M_{AA,A}|M_{AA}) = 1 \quad P(F_{aA,A}|F_{aA}) = P(M_{aA,A}|M_{aA}) = \frac{1}{2} \quad P(F_{aa,A}) = 0$$

Le condizioni (i) e (ii) sono dette di *accoppiamento casuale*, e implicano in particolare che la fertilità è indipendente dal sesso e dal genotipo, e che la frequenza dei genotipi non dipende dal sesso. La condizione (iii) è invece una semplice istanza delle *Leggi di Mendel*.

Lo scopo di quanto segue è quello di mostrare che le ipotesi precedenti implicano una relazione funzionale tra le frequenze dei genotipi nella generazione 0 e quelle nella generazione 1. Si noti anzitutto che

$$D_{AA} = [F_{AA,A} \cap M_{AA,A}] \cup [F_{aA,A} \cap M_{AA,A}] \cup [F_{AA,A} \cap M_{aA,A}] \cup [F_{aA,A} \cap M_{aA,A}].$$

Evidentemente, le unioni precedenti sono tra eventi disgiunti. Inoltre ognuna delle intersezioni nella formula precedente sono, per l'ipotesi (i), tra eventi indipendenti. Pertanto:

$$u_1 = P(D_{AA}) = P(F_{AA,A})P(M_{AA,A}) + P(F_{aA,A})P(M_{AA,A}) + P(F_{AA,A})P(M_{aA,A}) + P(F_{aA,A})P(M_{aA,A}). \quad (2.62)$$

Inoltre, per l'ipotesi (iii)

$$P(F_{AA,A}) = P(M_{AA,A}) = P(F_{AA}) = u \quad P(F_{aA,A}) = P(M_{aA,A}) = \frac{1}{2}P(F_{aA}) = v.$$

Perciò

$$u_1 = u^2 + 2uv + v^2 = (u + v)^2.$$

Per simmetria

$$w_1 = (v + w)^2.$$

Essendo

$$u_1 + w_1 = (u + v)^2 + (v + w)^2 = (u + 2v + w)^2 - 2(u + v)(v + w) = 1 - 2(u + v)(v + w)$$

si ha

$$2v_1 = 1 - u_1 - w_1 = 2(u + v)(v + w).$$

Quindi, se definiamo la funzione di tre variabili

$$T(u, v, w) = ((u + v)^2, (u + v)(v + w), (v + w)^2),$$

abbiamo ottenuto la relazione

$$(u_1, v_1, w_1) = T(u, v, w).$$

In altre parole, la mappa T fornisce la relazione tra le frequenze dei genotipi di una generazione e quelle della successiva. A questo punto possiamo anche affermare che le frequenze dei genotipi nella generazione successiva alla generazione 1, diciamo la generazione 2, sono date da

$$(u_2, v_2, w_2) = T(u_1, v_1, w_1) = T(T(u, v, w))$$

Si osservi che la prima componente di $T(T(u, v, w))$ è

$$[(u + v)^2 + (u + v)(v + w)]^2 = [(u + v)(u + 2v + w)]^2 = (u + v)^2,$$

essendo $u + 2v + w = 1$, e quindi è uguale alla prima componente di $T(u, v, w)$. In modo analogo si mostra che la terza componente di $T(T(u, v, w))$ coincide con la terza componente di $T(u, v, w)$. Pertanto, essendo $u_2 + 2v_2 + w_2 = u_1 + 2v_1 + w_1 = 1$, possiamo concludere che

$$T(T(u, v, w)) = T(u, v, w).$$

Questo implica che *la frequenza dei genotipi dalla generazione 1 in poi rimane costante*: la popolazione raggiunge un equilibrio dopo una sola generazione!

Quello che abbiamo fin qui descritto è la versione standard del modello di Hardy-Weinberg. Sono state proposte numerose modifiche e generalizzazioni, per tener conto di fenomeni quali la dipendenza dal sesso dei caratteri e la selezione naturale. Vedremo ora brevemente una di queste versioni modificate, che rappresenta un semplice modello per la *selezione naturale*. Questo modello si ottiene dal precedente modificando l'ipotesi (ii) come segue:

(ii)'

$$P(F_{AA}) = P(M_{AA}) = \frac{u}{u+2v} \quad P(F_{Aa}) = P(M_{Aa}) = \frac{2v}{u+2v} \quad P(F_{aa}) = P(M_{aa}) = 0.$$

Il significato di (ii)' è evidente: gli individui che manifestano il carattere recessivo, cioè di genotipo aa non si riproducono. Lasciamo al lettore controllare che la relazione in (2.62) diventa:

$$u_1 = \left(\frac{u}{u+2v}\right)^2 + \left(\frac{u}{u+2v}\right)\left(\frac{v}{u+2v}\right) + \left(\frac{v}{u+2v}\right)^2 = \left(\frac{u+v}{u+2v}\right)^2.$$

Analogamente

$$w_1 = \left(\frac{v}{u+2v}\right)^2, \quad v_1 = \frac{v(u+v)}{(u+2v)^2}.$$

Posto, come prima, $(u_1, v_1, w_1) =: T(u, v, w)$, possiamo definire induttivamente

$$(u_n, v_n, w_n) = T(u_{n-1}, v_{n-1}, w_{n-1}). \tag{2.63}$$

Ne segue che $u_n, 2v_n$ e w_n sono le frequenze dei tre genotipi nell' n -esima generazione. Con un calcolo paziente ma elementare, si verifica per induzione che

$$\begin{aligned} u_n &= \left(\frac{u+nv}{u+(n+1)v} \right)^2 \\ v_n &= \frac{v(u+nv)}{(u+(n+1)v)^2} \\ w_n &= \frac{v^2}{(u+(n+1)v)^2}. \end{aligned}$$

Notare che asintoticamente in n

$$v_n \simeq \frac{1}{n+1} \quad w_n \simeq \frac{1}{(n+1)^2}.$$

Abbiamo dunque quantificato, come effetto della selezione naturale, la progressiva diminuzione dei genotipi che contengono il gene a .

2.5 Un'applicazione alla finanza: il modello di Black-Scholes discreto

In questo paragrafo si utilizza la nozione di variabile casuale, sviluppata nel capitolo 3.

Consideriamo un mercato finanziario molto semplificato, in cui sia presente un unico titolo rischioso, ad esempio un'azione, e in cui non vi sia inflazione e non vi siano costi di transazione. Sia $X_0 \equiv x$ il valore odierno del titolo, che viene aggiornato una volta al giorno. Le variabili casuali X_1, X_2, \dots rappresentano i valori del titolo nei giorni successivi.

Gli istituti finanziari, ad esempio le banche, offrono ai loro clienti le cosiddette *opzioni*. Un esempio di opzione è l'*opzione call europea*: il cliente, che ha investito nel titolo, acquista il diritto di vendere al giorno N all'istituto finanziario una (o più) unità del titolo ad un prezzo $\pi = \pi(X_N)$ dato da

$$\pi(X_N) = \begin{cases} X_N & \text{se } X_N \geq a \\ a & \text{se } X_N < a \end{cases}$$

dove a è un prezzo minimo prefissato. In questo modo il cliente limita il rischio di perdite.

Una delle domande fondamentali in finanza è: quale è il *giusto* prezzo per una tale opzione? La nozione di giusto prezzo è legata a quella di *strategia di copertura*. Supponiamo che al momento di vendita dell'opzione, l'istituto finanziario utilizzi un capitale $V_0 = V$ per attivare un *portafoglio*, di cui una parte investita in a_0 unità del titolo rischioso, la parte restante c_0 investita in un titolo non rischioso, cioè a tasso di rendimento costante. Per semplicità di calcolo assumiamo che tale tasso sia uguale a zero. In altre parole

$$V_0 = a_0 X_0 + c_0.$$

Il giorno successivo, essendo variato da X_0 a X_1 il prezzo del titolo rischioso, il capitale diventa

$$V_1 = V_0 + a_0(X_1 - X_0).$$

A questo punto l'istituto finanziario può modificare la porzione di capitale investita nel titolo rischioso, passando da a_0 ad a_1 unità possedute, con il rimanente $c_1 = V_1 - a_1 X_1$ mantenuta a rendimento costante. Dopo due giorni il capitale disponibile diventa:

$$V_2 = V_1 + a_1(X_2 - X_1) = V_0 + a_0(X_1 - X_0) + a_1(X_2 - X_1),$$

Iterando il procedimento, si ottiene che il capitale dopo n giorni è

$$V_n = V_0 + \sum_{i=0}^{n-1} a_i(X_{i+1} - X_i).$$

I coefficienti a_0, a_1, \dots, a_{N-1} rappresentano la strategia di investimento. Affinché si tratti di una strategia effettivamente implementabile essa dev'essere *non-anticipativa*; in altre parole la scelta di a_n può dipendere dall'evoluzione del prezzo solo fino all'istante n , cioè $a_n = a_n(X_0, X_1, \dots, X_n)$ (ciò significa che la strategia si può basare solo sull'informazione effettivamente disponibile). È chiaro che in questo caso anche $V_n = V_n(X_0, X_1, \dots, X_n)$.

Definizione 2.18. Per un dato capitale iniziale V_0 , una strategia non-anticipativa a_0, a_1, \dots, a_{N-1} si dice *strategia di copertura* se, per qualunque evoluzione X_1, \dots, X_N del prezzo, si ha

$$V_N(X_0, X_1, \dots, X_N) = \pi(X_N).$$

Nel caso in cui vi sia un'unico valore di V_0 per cui esiste una strategia di copertura, diremo che V_0 è il *prezzo dell'opzione*.

Si noti che è anche naturale assumere che una strategia sia tale che, in ogni giorno n , $a_n \geq 0$ e $a_n X_n \leq V_n$. Se fosse $a_n < 0$, l'istituto avrebbe operato un *vendita allo scoperto*, mentre se $a_n X_n > V_n$ avrebbe dovuto ricorrere ad ulteriore capitale, cioè ad un prestito. Nel semplice modello di mercato che studieremo tra poco, mostreremo che la strategia di copertura è unica, e soddisfa alle suddette proprietà, che quindi non assumeremo a priori.

Lo stabilire l'esistenza ed, eventualmente, l'unicità del capitale iniziale e della strategia per la copertura dell'opzione, costituisce un problema di estrema rilevanza economica e che si presta ad una trattazione matematica elegante ed efficace. La nozione di copertura si può evidentemente estendere ad una qualunque funzione $\pi(X_N)$. I modelli di mercato in cui ad ogni tale funzione è associata una strategia di copertura vengono chiamati mercati *completi*. La completezza di un modello di mercato ammette varie caratterizzazioni matematiche, interessanti e utili. In queste note non affronteremo questo problema; ci limiteremo a semplificare ulteriormente il nostro modello di mercato con un unico titolo, e mostrare che in esso l'opzione europea ammette un prezzo. A tale scopo, definiamo, per $n = 1, 2, \dots, N$, $R_n = X_n/X_{n-1}$, e assumiamo che le variabili casuali R_n possano assumere con probabilità strettamente positiva esattamente due valori c e C , con $0 < c < 1 < C$. Ciò significa che ogni giorno il prezzo del titolo può soltanto o contrarsi di un fattore c o dilatarsi di un fattore C . Osserviamo che $X_n = R_n R_{n-1} \cdots R_1 x > 0$ se $x > 0$.

Possiamo ora analizzare nel dettaglio il problema della copertura. Consideriamo il capitale $V_{N-1} = V_{N-1}(X_0, \dots, X_{N-1})$ a disposizione dell'istituto al giorno $N-1$, e sia a_{N-1} l'ultimo passo della strategia. Noto il valore X_{N-1} del prezzo del titolo, il prezzo X_N del giorno successivo può assumere i due valori cX_{N-1} e CX_{N-1} . Se la strategia in esame copre l'opzione, allora dev'essere vero che

$$\pi(X_N) = V_{N-1} + a_{N-1}(X_N - X_{N-1})$$

sia per $X_N = cX_{N-1}$ che per $X_N = CX_{N-1}$. In altre parole

$$\begin{aligned} \pi(cX_{N-1}) &= V_{N-1} - a_{N-1}(1-c)X_{N-1} \\ \pi(CX_{N-1}) &= V_{N-1} + a_{N-1}(C-1)X_{N-1}, \end{aligned}$$

dalla quale si ricava

$$a_{N-1} = \frac{\pi(CX_{N-1}) - \pi(cX_{N-1})}{(C-c)X_{N-1}} \quad (2.64)$$

$$V_{N-1} = \frac{1-c}{C-c}\pi(CX_{N-1}) + \frac{C-1}{C-c}\pi(cX_{N-1}). \quad (2.65)$$

Dunque, il capitale V_{N-1} e la strategia a_{N-1} sono determinati in modo *univoco* dalla richiesta di copertura dell'opzione. A questo punto il procedimento si può iterare. Per coprire l'opzione è necessario disporre del capitale V_{N-1} dato da (2.65) al giorno $N-1$. Ragionando come prima, il capitale V_{N-2} e la strategia a_{N-2} devono soddisfare le relazioni

$$\begin{aligned} V_{N-1}(cX_{N-2}) &= V_{N-2} - a_{N-2}(1-c)X_{N-2} \\ V_{N-1}(CX_{N-2}) &= V_{N-2} + a_{N-2}(C-1)X_{N-2}, \end{aligned}$$

da cui

$$\begin{aligned} a_{N-2} &= \frac{V_{N-1}(CX_{N-2}) - V_{N-1}(cX_{N-2})}{(C-c)X_{N-2}} \\ V_{N-2} &= \frac{1-c}{C-c}V_{N-1}(CX_{N-2}) + \frac{C-1}{C-c}V_{N-1}(cX_{N-2}). \end{aligned}$$

Iterando, per $n \geq 0$, troveremo le formule ricorsive

$$a_n = \frac{V_{n+1}(CX_n) - V_{n+1}(cX_n)}{(C-c)X_n} \quad (2.66)$$

$$V_n = \frac{1-c}{C-c}V_{n+1}(CX_n) + \frac{C-1}{C-c}V_{n+1}(cX_n). \quad (2.67)$$

Dovendo essere $V_N(X_N) = \pi(X_N)$, questo determina in modo univoco l'intera strategia a_0, a_1, \dots, a_{N-1} , e il prezzo $V = V_0(x)$ dell'opzione! È necessario, in realtà, verificare che quest'unica strategia possibile sia effettivamente autofinanziante, cioè che vale la seguente proposizione.

Proposizione 2.19. *La strategia $a_n = a_n(X_n)$ e i corrispondenti valori del capitale $V_n = V_n(X_n)$ dati da (2.66) e (2.67) soddisfano le relazioni*

$$0 \leq a_n X_n \leq V_n$$

per ogni $0 \leq n < N$.

Dimostrazione. Si noti che le equazioni (2.66) e (2.67) sono formule ricorsive le cui soluzioni sono funzioni di una variabile $a_n(x), V_n(x)$:

$$a_n(x) = \frac{V_{n+1}(Cx) - V_{n+1}(cx)}{(C-c)x} \quad (2.68)$$

$$V_n(x) = \frac{1-c}{C-c}V_{n+1}(Cx) + \frac{C-1}{C-c}V_{n+1}(cx). \quad (2.69)$$

con condizione "terminale" $V_N(x) = \pi(x)$. Vogliamo dimostrare che, per ogni $0 \leq n < N$ e ogni $x > 0$, si ha

$$0 \leq xa_n(x) \leq V_n(x). \quad (2.70)$$

Anzitutto notiamo che $V_N(x) = \pi(x)$ è una funzione non negativa e crescente di $x > 0$. È facile vedere usando (2.69), per induzione “all’indietro” su n che tali proprietà sono trasmesse a tutte le $V_n(x)$ per $n \leq N - 1$. In particolare, essendo $Cx > cx$ per ogni $x > 0$, questo implica $a_n(x) \geq 0$ per ogni $0 \leq n \leq N - 1$. Resta pertanto da dimostrare la disuguaglianza $xa_n(x) \leq V_n(x)$ che, per (2.68) e (2.69), equivale a

$$\frac{V_{n+1}(Cx) - V_{n+1}(cx)}{C - c} \leq \frac{1 - c}{C - c} V_{n+1}(Cx) + \frac{C - 1}{C - c} V_{n+1}(cx) \iff V_{n+1}(Cx) \leq \frac{C}{c} V_{n+1}(cx).$$

Quest’ultima disuguaglianza è verificata se mostriamo che, per ogni $0 \leq n \leq N$, ogni $x > 0$ e $\rho > 1$

$$V_n(\rho x) \leq \rho V_n(x). \quad (2.71)$$

Tale disuguaglianza è verificata per $n = N$ (verificarlo distinguendo i tre casi $x \geq a$, $x < a$, $\rho x \geq a$, $\rho x < a$). Inoltre, di nuovo per induzione all’indietro, la proprietà (2.71) si trasmette alle altre V_n . \square

Le equazioni ricorsive (2.64) e (2.65) determinano quindi tanto il prezzo dell’opzione, come funzione del valore del titolo al momento in cui l’opzione viene acquistata, quanto la strategia autofinanziante che copre l’opzione. Va notato che tanto il prezzo quanto la strategia *non* dipendono dalla distribuzione di (X_1, X_2, \dots, X_N) , ma solo dall’ipotesi fatta che il valore X_{n+1} del titolo al giorno $n + 1$ possa essere o cX_n oppure CX_n . Il valore $V(x)$ del prezzo ammette però un’interpretazione probabilistica che si può estendere a modelli di mercato assai più complessi, e che ha grande rilevanza sia teorica che applicativa. Ricordiamo che, come osservato in precedenza, $X_n = R_n R_{n-1} \cdots R_1 x$, dove $R_n = X_n / X_{n-1}$ è la variabile casuale che “aggiorna” il valore del titolo al giorno n .

Teorema 2.20. *Siano Q_1, Q_2, \dots, Q_N variabili casuali indipendenti, che assumono soltanto i valori c e C , e di media $E(Q_n) = 1$. Allora la soluzione $V_n(x)$ dell’equazione ricorsiva (2.69) è, per $n \leq N - 1$,*

$$V_n(x) = E[\pi(Q_{n+1} \cdots Q_N x)].$$

In particolare, il prezzo dell’opzione è

$$V(x) = V_0(x) = E[\pi(Q_1 Q_2 \cdots Q_N x)].$$

Osservazione 2.21. Posto $Z_n = Q_1 Q_2 \cdots Q_n x$, il prezzo dell’opzione è dunque $V = E[\pi(Z_N)]$. La variabile casuale Z_n si può interpretare come il valore al giorno n di un titolo avente lo stesso valore iniziale x del titolo reale, ma che evolve in un mercato *fittizio* in cui le variabili casuali di “aggiornamento” sono indipendenti e hanno media 1 (*mercato neutrale*). Il prezzo dell’opzione è dato dal valor medio del costo finale a carico dell’istituto finanziario ($\pi(Z_N)$) calcolato non rispetto al mercato “reale” ma rispetto al corrispondente mercato neutrale.

Dimostrazione del Teorema 2.20. Si noti, anzitutto, che le richieste di assumere solo i valori c e C e di avere media 1 caratterizzano univocamente la distribuzione (comune) dell’ Q_n . Infatti dev’essere

$$\begin{aligned} cP(Q_n = c) + CP(Q_n = C) &= 1 \\ P(Q_n = c) + P(Q_n = C) &= 1 \end{aligned}$$

da cui segue

$$P(Q_n = c) = \frac{C-1}{C-c}, \quad P(Q_n = C) = \frac{1-c}{C-c}.$$

Inoltre, usando l'indipendenza delle Q_n , posto $W_n(x) := E[\pi(Q_{n+1} \cdots Q_N x)]$, per $n \leq N-2$ si ha

$$\begin{aligned} W_n(x) &= E[\pi(Q_{n+1} \cdots Q_N x)] \\ &= \sum_{q_{n+1} \in \{c, C\}} \sum_{q_{n+2} \in \{c, C\}} \cdots \sum_{q_N \in \{c, C\}} \pi(q_{n+1} q_{n+2} \cdots q_N x) P(Q_{n+1} = q_{n+1}, Q_{n+2} = q_{n+2}, \dots, Q_N = q_N) \\ &= \sum_{q_{n+1} \in \{c, C\}} \sum_{q_{n+2} \in \{c, C\}} \cdots \sum_{q_N \in \{c, C\}} \pi(q_{n+1} q_{n+2} \cdots q_N x) P(Q_{n+1} = q_{n+1}) P(Q_{n+2} = q_{n+2}) \cdots P(Q_N = q_N) \\ &= \sum_{q_{n+1} \in \{c, C\}} P(Q_{n+1} = q_{n+1}) \sum_{q_{n+2} \in \{c, C\}} \cdots \sum_{q_N \in \{c, C\}} \pi(q_{n+1} q_{n+2} \cdots q_N x) P(Q_{n+2} = q_{n+2}) \cdots P(Q_N = q_N) \\ &= \frac{C-1}{C-c} \sum_{q_{n+2} \in \{c, C\}} \cdots \sum_{q_N \in \{c, C\}} \pi(q_{n+2} \cdots q_N c x) P(Q_{n+2} = q_{n+2}) \cdots P(Q_N = q_N) \\ &\quad + \frac{1-c}{C-c} \sum_{q_{n+2} \in \{c, C\}} \cdots \sum_{q_N \in \{c, C\}} \pi(q_{n+2} \cdots q_N C x) P(Q_{n+2} = q_{n+2}) \cdots P(Q_N = q_N) \\ &= \frac{C-1}{C-c} W_{n+1}(cx) + \frac{1-c}{C-c} W_{n+1}(Cx). \end{aligned}$$

Lo stesso argomento per $n = N-1$ fornisce

$$W_{N-1}(x) = \frac{C-1}{C-c} \pi(cx) + \frac{1-c}{C-c} \pi(Cx).$$

Dunque $W_n(x)$ risolve l'equazione ricorsiva (2.69), e pertanto $W_n(x) = V_n(x)$. \square

Capitolo 3

Variabili casuali discrete

3.1 Variabili casuali e loro distribuzioni

In molti degli esempi dei capitoli precedenti, abbiamo calcolato delle probabilità di eventi che si potevano esprimere in termini di una funzione dell'esito di un esperimento aleatorio.

Esempio 3.1. Riprendiamo qui l'esempio 1.37, in cui lo spazio campionario è

$$\Omega = \{x = (x_1, x_2, \dots, x_N) : x_i \in \{0, 1\}\} = \{0, 1\}^N,$$

e la probabilità P è data da

$$P(\{y\}) = p^{\sum_{i=1}^N y_i} (1-p)^{N-\sum_{i=1}^N y_i}.$$

Abbiamo calcolato la probabilità dell'evento

$$A = \{x \in \Omega : \sum_{i=1}^N x_i = n\},$$

dove $0 \leq n \leq N$, e dell'evento

$$B = \{x \in \Omega : x_n = 1, x_k = 0 \forall k < n\}.$$

Posto $X(x) = \sum_{i=1}^N x_i$, e

$$Y(x) = \begin{cases} \min\{k : x_k = 1\} & \text{se } \{k : x_k = 1\} \neq \emptyset \\ N+1 & \text{altrimenti,} \end{cases}$$

possiamo riscrivere

$$A = \{x \in \Omega : X(x) = n\}$$

e

$$B = \{x \in \Omega : Y(x) = n\}.$$

Definizione 3.2. Sia (Ω, P) uno spazio di probabilità discreto, e E un insieme. Una funzione $X : \Omega \rightarrow E$ si dice *variabile casuale* o *variabile aleatoria* a valori in E .

Introduciamo un po' di terminologia e di notazioni. Con riferimento alla definizione precedente, se $E = \mathbb{R}$ diciamo che X è una variabile casuale *scalare*. Se $E = \mathbb{R}^n$ diciamo che X è una variabile casuale *vettoriale* o *vettore aleatorio* di dimensione n . Se $E = \mathbb{C}$, diciamo che X è una variabile casuale *complessa*.

Se $A \subseteq E$, allora $X^{-1}(A) = \{\omega \in \Omega : X(\omega) \in A\}$ è un sottoinsieme di Ω . Scriveremo $\{X \in A\}$ e $P(X \in A)$ in luogo di $X^{-1}(A)$ e $P(X^{-1}(A))$ rispettivamente. Nel caso in cui $A = \{x\}$, con $x \in E$, scriveremo $\{X = x\}$ e $P(X = x)$ invece di $X^{-1}(\{x\})$ e $P(X^{-1}(\{x\}))$. Se $(A_n)_{n \geq 0}$ è una successione di sottoinsiemi disgiunti di E , allora gli eventi $\{X \in A_n\}$ sono disgiunti (verificarlo!). Ne segue che

$$P\left(X \in \bigcup_n A_n\right) = \sum_n P(X \in A_n). \quad (3.1)$$

Si noti infine che, essendo Ω un insieme finito o numerabile, anche l'immagine di X , $X(\Omega) = \{X(\omega) : \omega \in \Omega\}$, è un insieme finito o numerabile (in generale, la cardinalità dell'immagine di una funzione è minore o uguale a quella del dominio della stessa). Perciò, se $A \subseteq E$, allora $A \cap X(\Omega)$ è finito o numerabile. Essendo (perché?) $P(X \in A) = P(X \in A \cap X(\Omega))$, usando (3.1) si ha

$$\begin{aligned} P(X \in A) &= P\left(X \in \bigcup_{x \in A \cap X(\Omega)} \{x\}\right) \\ &= \sum_{x \in A \cap X(\Omega)} P(X = x). \end{aligned}$$

Poichè, evidentemente, $P(X = x) = 0$ se $x \notin X(\Omega)$, possiamo sinteticamente scrivere l'identità precedente come

$$P(X \in A) = \sum_{x \in A} P(X = x).$$

Possiamo riassumere le precedenti osservazioni come segue.

Definizione 3.3. Sia X una variabile casuale discreta a valori in E . La funzione

$$\begin{aligned} \mu_X : \mathcal{P}(E) &\rightarrow [0, 1] \\ A &\mapsto P(X \in A) \end{aligned}$$

si dice *distribuzione* o *legge* della variabile casuale X .

Proposizione 3.4. La distribuzione μ_X di una variabile casuale discreta X a valori in E gode delle seguenti proprietà:

(i) se (A_n) è una successione di sottoinsiemi disgiunti di E , allora

$$\mu_X\left(\bigcup_n A_n\right) = \sum_n \mu_X(A_n);$$

(ii) per ogni sottoinsieme A di E

$$\mu_X(A) = \sum_{x \in A} P(X = x) = \sum_{x \in A} \mu_X(\{x\}).$$

Osservazione 3.5. Nel caso in cui E è un insieme finito o numerabile, segue dalla Proposizione 3.4 che la coppia (E, μ_X) è uno spazio di probabilità discreto. Va tuttavia notato che, in generale, non facciamo alcuna assunzione sulla cardinalità di E . Abbiamo però osservato che $X(\Omega)$ è sicuramente un insieme finito o numerabile. Denotando con $\tilde{\mu}_X$ la restrizione di μ_X ai sottoinsiemi di $X(\Omega)$, si ha che $(X(\Omega), \tilde{\mu}_X)$ è uno spazio di probabilità discreto.

3.2 Densità discreta

Sia X una variabile casuale discreta a valori in un insieme E .

Definizione 3.6. La funzione

$$\begin{aligned} p_X : E &\rightarrow [0, 1] \\ x &\mapsto P(X = x) \end{aligned}$$

si dice *densità discreta* della variabile casuale X .

Nel seguito l'aggettivo “discreta” verrà ommesso, fino a quando non introdurremo, nel capitolo 5, la nozione “continua” di densità.

Osservazione 3.7. La densità p_X e la distribuzione μ_X di una variabile casuale discreta, si determinano l'un l'altra tramite le relazioni, per $x \in E$, $A \subseteq E$

$$p_X(x) = \mu_X(\{x\}), \quad \mu_X(A) = \sum_{x \in A} p_X(x). \quad (3.2)$$

Consideriamo ora la situazione in cui una variabile casuale X prende valori in un insieme della forma $E = E_1 \times E_2 \times \cdots \times E_n$. In questo caso possiamo scrivere, per ogni $\omega \in \Omega$,

$$X(\omega) = (X_1(\omega), X_2(\omega), \dots, X_n(\omega)),$$

dove X_i , per $i = 1, \dots, n$, è una variabile casuale a valori in E_i . La densità $p_X : E \rightarrow [0, 1]$ di X è spesso denotata con p_{X_1, \dots, X_n} , e viene chiamata *densità congiunta* delle variabili casuali X_1, X_2, \dots, X_n . Consideriamo ora un sottoinsieme $I = \{i_1, \dots, i_m\} \subseteq \{1, 2, \dots, n\}$ e conveniamo $i_1 < i_2 < \cdots < i_m$. Possiamo dunque considerare la variabile casuale $(X_{i_1}, X_{i_2}, \dots, X_{i_m})$ a valori in $E_{i_1} \times E_{i_2} \times \cdots \times E_{i_m}$. La seguente proposizione stabilisce che la densità di tale variabile casuale è determinabile a partire dalla densità della variabile casuale X .

Proposizione 3.8. La densità $p_{X_{i_1}, X_{i_2}, \dots, X_{i_m}}$ della variabile casuale $(X_{i_1}, X_{i_2}, \dots, X_{i_m})$ è data dalla relazione

$$p_{X_{i_1}, X_{i_2}, \dots, X_{i_m}}(x_{i_1}, x_{i_2}, \dots, x_{i_m}) = \sum_{j \notin I} \sum_{x_j \in E_j} p_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n). \quad (3.3)$$

Dimostrazione. Dimostriamo il caso $n = 2$, $m = 1$, $I = \{1\}$. Il caso generale segue o usando lo stesso argomento, con notazioni più pesanti, oppure osservando che è sempre possibile ridursi a questo caso osservando che, posto $E_I = \times_{i \in I} E_i$ e $X_I = (X_{i_1}, X_{i_2}, \dots, X_{i_m})$, a meno di una permutazione di coordinate, l'insieme E si può identificare con $E_I \times E_{I^c}$, e la variabile X con la coppia di variabili (X_I, X_{I^c}) a valori rispettivamente in E_I e E_{I^c} .

Dunque, si tratta di dimostrare che, per ogni $x_1 \in E_1$,

$$p_{X_1}(x_1) = \sum_{x_2 \in E_2} p_{X_1, X_2}(x_1, x_2). \quad (3.4)$$

Si osservi che l'evento $\{X_1 = x_1\}$ può essere espresso come unione al più numerabile di eventi disgiunti come segue

$$\{X_1 = x_1\} = \bigcup_{x_2 \in X_2(\Omega)} \{X_1 = x_1, X_2 = x_2\}.$$

Pertanto, per la σ -additività,

$$\begin{aligned} p_{X_1}(x_1) = P(X_1 = x_1) &= \sum_{x_2 \in X_2(\Omega)} P(X_1 = x_1, X_2 = x_2) = \sum_{x_2 \in X_2(\Omega)} P((X_1, X_2) = (x_1, x_2)) \\ &= \sum_{x_2 \in X_2(\Omega)} p_{X_1, X_2}(x_1, x_2) = \sum_{x_2 \in E_2} p_{X_1, X_2}(x_1, x_2) \end{aligned}$$

dove, per l'ultima uguaglianza, abbiamo usato il fatto che se $x_2 \notin X_2(\Omega)$, allora l'evento $\{(X_1, X_2) = (x_1, x_2)\}$ è vuoto e ha dunque probabilità zero. \square

Nel caso in cui $k = 1$, la Proposizione 3.8 consente di esprimere la densità di una componente di un vettore aleatorio in termini della densità congiunta. Le densità delle componenti vengono chiamate *densità marginali*. Come mostriamo nei due seguenti esempi, non è possibile ricostruire la densità congiunta a partire dalla sola conoscenza delle densità marginali.

Esempio 3.9. Ad un esame partecipano n studenti, ed $m < n$ di essi portano con loro dei testi il cui uso non è consentito, nella speranza di non venire controllati. I due docenti addetti alla sorveglianza, Aldo e Giovanni, decidono di eseguire dei controlli casuali. Aldo controlla h studenti, mentre Giovanni ne controlla altri k , distinti da quelli controllati da Aldo. Supponiamo che se uno studente che ha con se testi non consentiti viene controllato, venga senz'altro espulso dall'aula. Sia X_A (risp. X_G) il numero di studenti espulsi da Aldo (risp. Giovanni). Si determinino le densità congiunte e marginali di X_A e X_G .

Supponiamo di "etichettare" gli studenti con numeri tra 1 e n , assegnando etichetta tra 1 e m a quelli che hanno portato testi non permessi. Come spazio campionario, possiamo scegliere

$$\Omega = \{(A, G) : A, G \subseteq \{1, 2, \dots, n\}, |A| = h, |G| = k, A \cap G = \emptyset\},$$

munito della probabilità uniforme P . Le variabili casuali X_A e X_G si possono allora definire come segue:

$$X_A(A, G) = |A \cap \{1, 2, \dots, m\}|, \quad X_G(A, G) = |G \cap \{1, 2, \dots, m\}|.$$

Siano x_A, x_G due interi fissati, con $1 \leq x_A \leq h$, $1 \leq x_G \leq k$. Anzitutto osserviamo che

$$\begin{aligned} p_{X_A, X_G}(x_A, x_G) &= P(X_A = x_A, X_G = x_G) \\ &= P(X_G = x_G | X_A = x_A) P(X_A = x_A). \end{aligned}$$

Da quanto visto nell'esempio 1.19

$$P(X_A = x_A) = p_{X_A}(x_A) = \frac{\binom{m}{x_A} \binom{n-m}{h-x_A}}{\binom{n}{h}}, \quad (3.5)$$

con la solita convenzione che $\binom{a}{b} = 0$ se $b > a$. Inoltre, condizionatamente al verificarsi dell'evento $\{X_A = x_A\}$, Giovanni sceglie k studenti tra gli $n - h$ non controllati da Aldo, dei quali $m - x_A$ hanno testi non ammessi. Perciò

$$P(X_G = x_G | X_A = x_A) = \frac{\binom{m-x_A}{x_G} \binom{n-h-m+x_A}{k-x_G}}{\binom{n-h}{k}}.$$

Mettendo tutto assieme:

$$p_{X_A, X_G}(x_A, x_G) = \frac{\binom{m}{x_A} \binom{n-m}{h-x_A}}{\binom{n}{h}} \frac{\binom{m-x_A}{x_G} \binom{n-h-m+x_A}{k-x_G}}{\binom{n-h}{k}}. \quad (3.6)$$

La densità marginale p_{X_A} è già stata calcolata in (3.5). Poichè il problema è completamente simmetrico nel ruolo di Aldo e Giovanni, esattamente come in (3.5) troviamo che

$$p_{X_G}(x_G) = \frac{\binom{m}{x_G} \binom{n-m}{k-x_G}}{\binom{n}{k}}. \quad (3.7)$$

Si noti infine che, dalla proposizione 3.8, sappiamo che

$$p_{X_G}(x_G) = \sum_{x_A=1}^h p_{X_A, X_G}(x_A, x_G).$$

Sostituendo i valori ottenuti per le densità in quest'ultima relazione, si ottiene una relazione combinatoria assolutamente non banale!

Esempio 3.10. Nello stesso contesto dell'esempio 3.9, supponiamo che Aldo e Giovanni effettuino i controlli in momenti successivi, senza comunicare tra loro, e che soltanto alla fine dei controlli di entrambi venga comunicata agli studenti "imbroglianti" la loro espulsione. In questo caso, dunque, non si esclude la possibilità che uno stesso studente sia controllato sia da Aldo che da Giovanni. Lo spazio campionario che descrive questa procedura è

$$\Omega = \{(A, G) : A, G \subseteq \{1, 2, \dots, n\}, |A| = h, |G| = k\},$$

munito della probabilità uniforme. Le variabili casuali X_A e X_G sono definite come nell'esempio 3.9. Inoltre, la diversa procedura di controllo è del tutto ininfluenza per il calcolo delle densità marginali, che risultano identiche a quelle calcolate nell'esempio precedente.

Per il calcolo della densità congiunta, si osservi che $\Omega = \Omega_A \times \Omega_G$, dove

$$\Omega_A = \{A \subseteq \{1, 2, \dots, n\} : |A| = h\}$$

$$\Omega_G = \{G \subseteq \{1, 2, \dots, n\} : |G| = k\}.$$

Se $H \subseteq \Omega_A$ e $K \subseteq \Omega_G$, gli eventi in Ω $H \times \Omega_G$ e $\Omega_A \times K$ sono indipendenti. Infatti:

$$\begin{aligned} P[(H \times \Omega_G) \cap (\Omega_A \times K)] &= P(H \times K) \\ &= \frac{|H||K|}{|\Omega_A||\Omega_G|} \\ &= \frac{|H||\Omega_G|}{|\Omega_A||\Omega_G|} \frac{|\Omega_A||K|}{|\Omega_A||\Omega_G|} \\ &= P(H \times \Omega_G)P(\Omega_A \times K). \end{aligned}$$

Si noti ora che l'evento $\{X_A = x_A\}$ è della forma $H \times \Omega_G$ con $H = \{A : |A \cap \{1, \dots, m\}| = x_A\}$, e similmente $\{X_G = x_G\} = \Omega_A \times K$ con $K = \{G : |G \cap \{1, \dots, m\}| = x_G\}$. Dunque i due eventi sono indipendenti. Ma allora

$$\begin{aligned} p_{X_A, X_G}(x_A, x_G) &= P(X_A = x_A, X_G = x_G) \\ &= P(X_A = x_A)P(X_G = x_G) \\ &= p_{X_A}(x_A)p_{X_G}(x_G), \end{aligned}$$

che è diversa dalla densità congiunta trovata nell'esempio precedente, pur essendo le stesse le densità marginali

3.3 Indipendenza di variabili casuali

Cominciamo col definire l'indipendenza per una famiglia finita di variabili casuali.

Definizione 3.11. Siano X_1, X_2, \dots, X_n variabili casuali definite nello stesso spazio di probabilità (Ω, P) , a valori rispettivamente negli insiemi E_1, E_2, \dots, E_n . Esse si dicono indipendenti se per ogni scelta di $A_1 \subseteq E_1, A_2 \subseteq E_2, \dots, A_n \subseteq E_n$ si ha

$$P(X_1 \in A_1, X_2 \in A_2, \dots, X_n \in A_n) = \prod_{i=1}^n P(X_i \in A_i). \quad (3.8)$$

Osserviamo che se X_1, X_2, \dots, X_n sono variabili casuali indipendenti e $\{i_1, i_2, \dots, i_k\} \subset \{1, 2, \dots, n\}$, allora $X_{i_1}, X_{i_2}, \dots, X_{i_k}$ sono variabili casuali indipendenti. Infatti se, in (3.8) si sceglie $A_j = E_j$ per $j \notin \{i_1, i_2, \dots, i_k\}$ si ottiene

$$\begin{aligned} P(X_{i_1} \in A_{i_1}, X_{i_2} \in A_{i_2}, \dots, X_{i_k} \in A_{i_k}) &= P(X_1 \in A_1, X_2 \in A_2, \dots, X_n \in A_n) \\ &= \prod_{i=1}^n P(X_i \in A_i) \\ &= \prod_{j=1}^k P(X_{i_j} \in A_{i_j}). \end{aligned}$$

Più in generale, per famiglie non necessariamente finite di variabili casuali, si dà la seguente definizione che, grazie all'osservazione appena fatta, è consistente con la Definizione 3.11

Definizione 3.12. Sia I un insieme qualsiasi di indici, e $\{X_i : i \in I\}$ una famiglia di variabili casuali a valori negli insiemi $E_i, i \in I$. Si dice che le variabili casuali di tale famiglia sono indipendenti se, per ogni $J \subset I$ finito e per ogni scelta di $A_j \subseteq E_j, j \in J$, si ha

$$P\left(\bigcap_{j \in J} \{X_j \in A_j\}\right) = \prod_{j \in J} P(X_j \in A_j).$$

Il semplice confronto tra la Definizione 3.12 e quella di indipendenza tra eventi, ci fornisce la seguente proprietà.

Proposizione 3.13. Siano $\{X_i : i \in I\}$ variabili casuali come nella definizione 3.12. Le seguenti affermazioni sono equivalenti:

- (i) La variabili casuali $\{X_i : i \in I\}$ sono indipendenti;
- (ii) per ogni scelta di $A_i \subseteq E_i$, gli eventi $\{X_i \in A_i\}, i \in I$ sono indipendenti.

Poichè la distribuzione congiunta di n variabili casuali è completamente determinata dalla loro densità congiunta, non è sorprendente che l'indipendenza si possa caratterizzare in termini della densità congiunta.

Proposizione 3.14. Siano X_1, X_2, \dots, X_n variabili casuali definite nello stesso spazio di probabilità (Ω, P) , a valori rispettivamente negli insiemi E_1, E_2, \dots, E_n . Denotiamo con p_{X_1, \dots, X_n} la loro densità congiunta, e con p_{X_i} le densità marginali. Allora X_1, X_2, \dots, X_n sono indipendenti se e solo se

$$p_{X_1, \dots, X_n}(x_1, \dots, x_n) = \prod_{i=1}^n p_{X_i}(x_i) \quad (3.9)$$

per ogni $(x_1, \dots, x_n) \in E_1 \times \dots \times E_n$.

Dimostrazione. Assumiamo che X_1, X_2, \dots, X_n siano indipendenti. Allora

$$p_{X_1, \dots, X_n}(x_1, \dots, x_n) = P(X_1 \in \{x_1\}, \dots, X_n \in \{x_n\}) = \prod_{i=1}^n P(X_i \in \{x_i\}) = \prod_{i=1}^n p_{X_i}(x_i).$$

Viceversa, assumiamo che valga (3.9). Per semplificare le espressioni che seguono, assumiamo $n = 2$; l'argomento che usiamo è però valido in generale. Siano $A_1 \in E_1, A_2 \in E_2$ fissati ma arbitrari. Si ha

$$\begin{aligned} P(X_1 \in A_1, X_2 \in A_2) &= P((X_1, X_2) \in A_1 \times A_2) = \sum_{x_1 \in A_1, x_2 \in A_2} p_{X_1, X_2}(x_1, x_2) \\ &= \sum_{x_1 \in A_1, x_2 \in A_2} p_{X_1}(x_1)p_{X_2}(x_2) = \sum_{x_1 \in A_1} p_{X_1}(x_1) \sum_{x_2 \in A_2} p_{X_2}(x_2) = P(X_1 \in A_1)P(X_2 \in A_2). \end{aligned}$$

□

Osservazione 3.15. Siano X, Y due variabili casuali (ma l'argomento è generalizzabile a n variabili casuali) la cui densità congiunta si fattorizza nella forma

$$p_{X,Y}(x, y) = \alpha(x)\beta(y).$$

Allora

$$p_X(x) = \sum_y p_{X,Y}(x, y) = b\alpha(x),$$

dove $b = \sum_y \beta(y)$. Analogamente

$$p_Y(y) = a\beta(y),$$

con $a = \sum_x \alpha(x)$. Inoltre

$$1 = \sum_{x,y} p_{X,Y}(x, y) = ab.$$

Di conseguenza

$$p_{X,Y}(x, y) = \alpha(x)\beta(y) = p_X(x)p_Y(y),$$

e dunque X e Y sono indipendenti.

Esempio 3.16. Si considerino gli esempi 3.9 e 3.10. Le variabili casuali X_A, X_G dell'Esempio 3.10 sono indipendenti, mentre quelle dell'Esempio 3.9 non sono indipendenti.

Osservazione 3.17. Siano X_1, X_2, \dots, X_n variabili casuali indipendenti, e $I = \{i_1, \dots, i_h\}$, $J = \{j_1, \dots, j_k\}$ due sottoinsiemi non vuoti e *disgiunti* di $\{1, 2, \dots, n\}$. Denotiamo con X_I la variabile casuale

$$X_I = (X_{i_1}, \dots, X_{i_h}),$$

e, analogamente, X_J . Mostriamo che X_I, X_J sono variabili casuali indipendenti. Scriviamo x_I in luogo di $(x_{i_1}, \dots, x_{i_h})$, e analogamente x_J . Allora, tenendo conto che le variabili casuali $\{X_l : l \in I \cup J\}$ sono indipendenti, si ha

$$\begin{aligned} p_{X_I, X_J}(x_I, x_J) &= p_{X_{i_1}, \dots, X_{i_h}, X_{j_1}, \dots, X_{j_k}}(x_{i_1}, \dots, x_{i_h}, x_{j_1}, \dots, x_{j_k}) \\ &= \prod_{r=1}^h p_{X_{i_r}}(x_{i_r}) \prod_{s=1}^k p_{X_{j_s}}(x_{j_s}) = p_{X_I}(x_I) p_{X_J}(x_J), \end{aligned}$$

da cui segue l'indipendenza.

Come vedremo, il risultato nell'Osservazione 3.17 viene usato congiuntamente a quello della seguente proposizione, che stabilisce che l'indipendenza si conserva per trasformazioni. Il risultato che segue è enunciato e dimostrato per due variabili casuali, ma è facilmente estendibile a ogni $n \geq 2$.

Proposizione 3.18. Siano X, Y due variabili casuali definite nello stesso spazio di probabilità (Ω, P) , e a valori negli insiemi E e F rispettivamente. Siano inoltre H, K due insiemi, e $f : E \rightarrow H$, $g : F \rightarrow K$ funzioni arbitrarie. Se le variabili casuali X e Y sono indipendenti, allora anche le variabili casuali $f(X)$ e $g(Y)$ sono indipendenti.

Dimostrazione. Basta osservare che, se $A \subseteq H$, $B \subseteq K$,

$$\begin{aligned} P(f(X) \in A, g(Y) \in B) &= P(X \in f^{-1}(A), Y \in g^{-1}(B)) \\ &= P(X \in f^{-1}(A))P(Y \in g^{-1}(B)) = P(f(X) \in A)P(g(Y) \in B). \end{aligned}$$

□

Esempio 3.19. Siano $X_1, \dots, X_n, X_{n+1}, \dots, X_{n+m}$ variabili casuali scalari indipendenti. Allora $X_1 + \dots + X_n$ e $X_{n+1} + \dots + X_{n+m}$ sono indipendenti. Basta applicare la proposizione precedente con $I = \{1, 2, \dots, n\}$, $J = \{n+1, \dots, n+m\}$, $f(x_1, \dots, x_n) = x_1 + \dots + x_n$, $g(x_{n+1}, \dots, x_{n+m}) = x_{n+1} + \dots + x_{n+m}$.

3.4 Il valor medio

La nozione di *media aritmetica* di un insieme finito di numeri reali $\{x_1, x_2, \dots, x_n\}$ è nota e molto naturale. Una delle sue possibili interpretazioni è quella che si ottiene associando ad ogni x_i un punto materiale; tali punti materiali vengono posizionati su una retta, ognuno nel punto corrispondente alla coordinata x_i . Se tali punti materiali hanno tutti la stessa massa, il punto di coordinate $\frac{1}{n} \sum_{i=1}^n x_i$ è il baricentro di tale sistema di punti materiali. Nel caso in cui i punti non abbiano tutti la stessa massa, il baricentro si ottiene attraverso una media “pesata”: se m_i è la massa del punto materiale in x_i , scegliendo l’unità di misura per la massa in modo tale che $\sum_{i=1}^n m_i = 1$, il baricentro del sistema ha coordinata

$$\mu = \sum_{i=1}^n m_i x_i.$$

In ambito probabilistico, la nozione di *valor medio* corrisponde alla nozione di baricentro, una volta interpretate le x_i come i valori assunti da una variabile casuale, e le m_i come le rispettive probabilità che i corrispondenti valori vengano assunti. Questo ci conduce alla seguente definizione.

Definizione 3.20. Sia X una variabile casuale a valori in \mathbb{R} o in \mathbb{C} , definita in uno spazio di probabilità discreto (Ω, P) . Si consideri la somma

$$\sum_{\omega \in \Omega} |X(\omega)| P(\{\omega\}). \quad (3.10)$$

Se tale somma ha un valore finito allora diremo che la variabile casuale X *ammette valor medio*. In tal caso, la quantità

$$E(X) \equiv \sum_{\omega \in \Omega} X(\omega) P(\{\omega\}) \quad (3.11)$$

si dice *valor medio*, o *media*, o *valor atteso* della variabile casuale X .

Osservazione 3.21. Nel caso in cui Ω sia un insieme finito, la somma (3.10) ha, ovviamente, valore finito. Nel caso in cui Ω sia numerabile, la condizione di finitezza della serie in (3.10) corrisponde alla sommabilità di (3.11).

Osservazione 3.22. Se X è una variabile casuale che assume solo valori reali positivi, la somma (3.11) è sempre definita, anche se può assumere il valore $+\infty$. In questo caso denoteremo con $E(X)$ il valore della somma, anche quando questo è $+\infty$. Con questa convenzione, per ogni variabile casuale a valori reali o complessi, la somma in (3.10) è $E(|X|)$. Dunque, talvolta scriveremo “ $E(|X|) < +\infty$ ” in luogo di “la variabile X ammette valor medio”. Si noti anche che, dalla definizione, X ammette valor medio se e solo se $|X|$ ammette valor medio.

Osservazione 3.23. Data una costante $c \in \mathbb{C}$, essa si può identificare con la variabile casuale X_c che assume solo il valore c : $X_c(\omega) = c \forall \omega$. È ovvio dalla definizione che $E(X_c) = c$. D’ora innanzi indicheremo con c sia il numero complesso che la variabile casuale X_c .

Le seguenti proprietà formali del valor medio derivano immediatamente dalla precedente definizione.

Proposizione 3.24. *Siano X, Y due variabili casuali discrete a valori in \mathbb{C} o \mathbb{R} , definite nello stesso spazio di probabilità (Ω, P) . Allora valgono le seguenti proprietà:*

- (i) *(Monotonia) Se X, Y sono a valori reali, entrambe ammettono valor medio e $X(\omega) \leq Y(\omega)$ per ogni $\omega \in \Omega$, allora $E(X) \leq E(Y)$.*
- (ii) *Se X ammette valor medio, allora*

$$|E(X)| \leq E(|X|).$$

- (iii) *(Linearità) Se X e Y ammettono valor medio e $a, b \in \mathbb{C}$, allora la variabile casuale $aX + bY$ definita da*

$$(aX + bY)(\omega) = aX(\omega) + bY(\omega),$$

ammette valor medio e

$$E(aX + bY) = aE(X) + bE(Y).$$

Le dimostrazioni sono del tutto elementari, e lasciate al lettore. Tali proprietà sono però assolutamente essenziali sia sul piano teorico che su quello computazionale. La prima conseguenza è il fatto che l’insieme delle variabili casuali che ammettono valor medio ha una struttura algebrica e topologica di grande utilità. Fissato uno spazio di probabilità discreto (Ω, P) , sia

$$L^1(\Omega, P) := \{X : \Omega \rightarrow \mathbb{R} \text{ tali che } X \text{ ammetta valor medio}\}$$

(trattiamo qui solo il caso di variabili casuali a valori reali, ma lo stesso si può fare per variabili casuali a valori complessi). La Proposizione 3.24 (iii) garantisce che $L^1(\Omega, P)$ sia uno spazio vettoriale su \mathbb{R} (dove lo “zero” dello spazio è la funzione costantemente uguale a 0). Per $X \in L^1(\Omega, P)$, definiamo

$$\|X\|_1 := E(|X|). \tag{3.12}$$

Teorema 3.25. *Supponiamo che lo spazio di probabilità discreto (Ω, P) sia tale che*

$$P(\{\omega\}) > 0 \text{ per ogni } \omega \in \Omega. \tag{3.13}$$

Allora la funzione $\|\cdot\|_1 : L^1(\Omega, P) \rightarrow \mathbb{R}^+$ è una norma. Inoltre lo spazio normato $(L^1(\Omega, P), \|\cdot\|_1)$ è completo, cioè è uno spazio di Banach.

Dimostrazione. Ricordiamo che dimostrare che $\|\cdot\|_1$ è una norma, significa mostrare:

$$\|X\|_1 = 0 \iff X \equiv 0; \quad (3.14)$$

$$X \in L^1(\Omega, P), \lambda \in \mathbb{R} \Rightarrow \|\lambda X\|_1 = |\lambda| \|X\|_1; \quad (3.15)$$

$$X, Y \in L^1(\Omega, P) \Rightarrow \|X + Y\|_1 \leq \|X\|_1 + \|Y\|_1. \quad (3.16)$$

Le proprietà (3.15) e (3.16) (quest'ultima è detta *disuguaglianza triangolare*) sono conseguenze abbastanza semplici della definizione di valor medio. La proprietà (3.14) deriva invece facilmente dall'ipotesi (3.13). Le dimostrazioni sono lasciate per esercizio.

Concentriamoci invece sulla questione della completezza. Ricordiamo che una successione $(X_n)_{n \geq 1}$ di elementi di $L^1(\Omega, P)$ si dice *successione di Cauchy* se

$$\lim_{n \rightarrow +\infty} \sup_{l, m \geq n} \|X_m - X_l\|_1 = 0. \quad (3.17)$$

Mostrare che uno spazio normato è completo significa dimostrare che ogni successione di Cauchy ammette un limite nella convergenza *indotta* dalla norma, cioè esiste (unico) $X \in L^1(P)$ tale che

$$\lim_{n \rightarrow +\infty} \|X_n - X\|_1 = 0 \quad (3.18)$$

(l'unicità del limite è una semplice conseguenza di (3.14)). Fissato $\omega \in \Omega$ è chiaro che

$$\|X_m - X_l\|_1 = \sum_{\omega' \in \Omega} |X_m(\omega') - X_l(\omega')| P(\{\omega'\}) \geq |X_m(\omega) - X_l(\omega)| P(\{\omega\}). \quad (3.19)$$

Usando (3.19), (3.17) e il fatto che $P(\{\omega\}) > 0$, ne segue che la successione di numeri reali $(X_n(\omega))_{n \geq 1}$ è una successione di Cauchy, e quindi converge, essendo \mathbb{R} uno spazio completo. È dunque lecito definire, per ogni $\omega \in \Omega$,

$$X(\omega) := \lim_{n \rightarrow +\infty} X_n(\omega).$$

Dimostriamo ora che $X \in L^1(\Omega, P)$, e che vale (3.18). Cominciamo con ricordare che una semplice conseguenza della disuguaglianza triangolare è il fatto che

$$\| \|X_m\|_1 - \|X_l\|_1 \| \leq \|X_m - X_l\|_1,$$

da cui segue che la successione di numeri reali $(\|X_n\|_1)_{n \geq 1}$ è di Cauchy. Perciò converge e, in particolare, è limitata. Quindi

$$M := \sup_{n \geq 1} \|X_n\|_1 < +\infty.$$

Sia ora A un sottoinsieme finito di Ω . Poichè il limite conserva le somme finite:

$$\sum_{\omega \in A} |X(\omega)| P(\{\omega\}) = \lim_{n \rightarrow +\infty} \sum_{\omega \in A} |X_n(\omega)| P(\{\omega\}) \leq \lim_{n \rightarrow +\infty} \|X_n\|_1 \leq M.$$

Ricordando la definizione di somma infinita:

$$\sum_{\omega \in \Omega} |X(\omega)| P(\{\omega\}) = \sup_{A \subseteq \Omega: |A| < +\infty} \sum_{\omega \in A} |X(\omega)| P(\{\omega\}) \leq M,$$

segue che $X \in L^1(\Omega, P)$. Resta da dimostrare (3.18). Usando nuovamente il fatto che il limite conserva le somme finite, per $A \subseteq \Omega$ finito si ha

$$\begin{aligned} \sum_{\omega \in A} |X(\omega) - X_n(\omega)|P(\{\omega\}) &= \lim_{m \rightarrow +\infty} \sum_{\omega \in A} |X_m(\omega) - X_n(\omega)|P(\{\omega\}) \\ &\leq \limsup_{m \rightarrow +\infty} \|X_m - X_n\|_1 \leq \sup_{l, m \geq n} \|X_m - X_l\|_1. \end{aligned}$$

Prendendo l'estremo superiore sui sottoinsiemi finiti di Ω nella precedente, si trova

$$\|X - X_n\|_1 \leq \sup_{l, m \geq n} \|X_m - X_l\|_1$$

che, assieme a (3.17), implica (3.18). \square

La definizione di valor medio che abbiamo data risulta poco operativa, in quanto in genere di una variabile casuale è nota solo la distribuzione. Per questo motivo è utile esprimere il valor medio di una variabile casuale in termini della sua densità. Nella seguente proposizione faremo, in realtà, qualcosa in più, ovvero esprimeremo in termini della densità di X la media di ogni "funzione" della variabile X . Useremo, di qui in avanti, la seguente notazione. Sia X una variabile casuale discreta a valori in un insieme generico E , e sia $f : E \rightarrow K$, con $K = \mathbb{R}$ o \mathbb{C} . Allora $f \circ X$ è una variabile casuale a valori in K , che denoteremo con $f(X)$.

Proposizione 3.26. *Sia X una variabile casuale discreta a valori in un insieme generico E , sia p_X la sua densità e sia $f : E \rightarrow K$, con $K = \mathbb{R}$ o \mathbb{C} . $f(X)$ ammette valor medio se e solo se*

$$\sum_{x \in E} |f(x)|p_X(x) < +\infty, \quad (3.20)$$

In questo caso

$$E(f(X)) = \sum_{x \in E} f(x)p_X(x). \quad (3.21)$$

In particolare, se $E = K$, X ammette valor medio se e solo se

$$\sum_{x \in K} |x|p_X(x) < +\infty,$$

e in questo caso

$$E(X) = \sum_{x \in K} xp_X(x).$$

Dimostrazione. Dalla definizione di valor medio, abbiamo che $f(X)$ ammette valor medio se e solo se

$$\sum_{\omega \in \Omega} |f(X(\omega))|P(\{\omega\}) < +\infty. \quad (3.22)$$

Cominciamo col mostrare che la somma in (3.20) coincide con quella in (3.22). Si noti anzitutto, che la famiglia di eventi $\{\omega \in \Omega : X(\omega) = x\}$, al variare di $x \in X(\Omega)$, costituisce

una partizione di Ω . Ma allora

$$\begin{aligned} \sum_{\omega \in \Omega} |f(X(\omega))| P(\{\omega\}) &= \sum_{x \in X(\Omega)} \sum_{\omega: X(\omega)=x} |f(X(\omega))| P(\{\omega\}) \\ &= \sum_{x \in X(\Omega)} |f(x)| P(X=x) \\ &= \sum_{x \in E} |f(x)| p_X(x). \end{aligned}$$

Ciò mostra che (3.20) equivale al fatto che $f(X)$ ammetta valor medio. Per concludere la dimostrazione, occorre mostrare che, se $f(X)$ ammette valor medio, allora la somma in (3.21) coincide con

$$E(f(X)) = \sum_{\omega \in \Omega} f(X(\omega)) P(\{\omega\}).$$

Ma per questo basta ripetere l'argomento appena usato per mostrare che la somma in (3.20) coincide con quella in (3.22). \square

3.5 Spazi L^p , momenti, varianza

Sia $p \geq 1$ un numero reale, e (Ω, P) uno spazio di probabilità discreto. Definiamo

$$L^p(\Omega, P) := \{X : \Omega \rightarrow \mathbb{R} \text{ tali che } E(|X|^p) < +\infty\}.$$

È chiaro che per $p = 1$ ritroviamo la definizione data nel paragrafo precedente. Un primo, semplice, risultato è il seguente.

Proposizione 3.27. *L'insieme $L^p(\Omega, P)$ è uno spazio vettoriale su \mathbb{R} .*

Dimostrazione. Il fatto che $X \in L^p(\Omega, P)$, $\lambda \in \mathbb{R}$ implichi $\lambda X \in L^p(\Omega, P)$ è immediato, e lo lasciamo verificare al lettore. Resta da dimostrare che se $X, Y \in L^p(\Omega, P)$ allora $X + Y \in L^p(\Omega, P)$, cioè

$$E(|X + Y|^p) < +\infty.$$

Essendo $E(|X + Y|^p) \leq E[(|X| + |Y|)^p]$, basta dimostrare la finitezza di quest'ultimo. Consideriamo la funzione $\varphi : [0, +\infty) \rightarrow \mathbb{R}$ data da $\varphi(x) = x^p$. Poiché $p \geq 1$, la funzione φ è convessa. Pertanto, se $x, y \geq 0$

$$\left(\frac{x+y}{2}\right)^p = \varphi\left(\frac{x+y}{2}\right) \leq \frac{\varphi(x) + \varphi(y)}{2} = \frac{x^p + y^p}{2},$$

o, equivalentemente,

$$(x+y)^p \leq 2^{p-1}(x^p + y^p).$$

Usando quest'ultima disuguaglianza:

$$E[(|X| + |Y|)^p] \leq 2^{p-1} [E(|X|^p) + E(|Y|^p)] < +\infty.$$

\square

Per $X \in L^p(\Omega, P)$, definiamo

$$\|X\|_p := [E(|X|^p)]^{1/p}. \quad (3.23)$$

Il teorema seguente generalizza il Teorema 3.25. Per $p > 1$ la dimostrazione è un po' più delicata, e la omettiamo.

Teorema 3.28. *Supponiamo che lo spazio di probabilità discreto (Ω, P) sia tale che*

$$P(\{\omega\}) > 0 \quad \text{per ogni } \omega \in \Omega. \quad (3.24)$$

Allora la funzione $\|\cdot\|_p : L^p(\Omega, P) \rightarrow \mathbb{R}^+$ è una norma. Inoltre lo spazio normato $(L^p(\Omega, P), \|\cdot\|_p)$ è completo, cioè è uno spazio di Banach.

Proposizione 3.29. *Se $1 \leq p < q$, allora $L^q(\Omega, P) \subseteq L^p(\Omega, P)$.*

Dimostrazione. Si osservi che, per ogni $x \in \mathbb{R}$, vale la disuguaglianza

$$|x|^p \leq 1 + |x|^q$$

(per verificarlo è sufficiente osservare che $|x|^p \leq 1$ se $|x| \leq 1$ e $|x|^p \leq |x|^q$ se $|x| > 1$). Pertanto, se $X \in L^q(\Omega, P)$,

$$E(|X|^p) \leq 1 + E(|X|^q) < +\infty,$$

da cui segue la tesi. □

Osservazione 3.30. Dimostreremo nel prossimo paragrafo una versione più forte della proposizione precedente, cioè che se $X \in L^q(\Omega, P)$ allora

$$\|X\|_p \leq \|X\|_q.$$

Quest'ultima disuguaglianza implica, in particolare, che l'inclusione $L^q(\Omega, P) \rightarrow L^p(\Omega, P)$ è continua rispetto alle corrispondenti norme.

Definizione 3.31. Sia $X \in L^k(\Omega, P)$, dove $k \geq 1$ è un intero. Diremo allora che X ammette *momento di ordine k* e la quantità $E(X^k)$ si dice *momento di ordine k* .

Poiché le costanti ammettono momento di ogni ordine, X ammette momento di ordine k se e solo se $X - c$ ammette momento di ordine k per ogni $c \in \mathbb{R}$. Se X ammette momento di ordine $k \geq 1$, per la Proposizione 3.29, X ammette valor medio. Allora ha senso considerare il momento di ordine k di $X - E(X)$, cioè $E[(X - E(X))^k]$, che viene detto *momento centrato di ordine k* . Si vede subito che il momento centrato di ordine 1 vale zero. Il momento centrato di ordine 2 si dice *varianza*, ed è denotato con $Var(\cdot)$:

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

Si osservi che la Varianza non è un operatore lineare. Infatti, per $a, b \in \mathbb{R}$

$$Var(aX) = a^2 Var(X),$$

e

$$Var(X + b) = Var(X).$$

La verifica di tali identità è semplice, ed è lasciata al lettore.

Notiamo ora che $Var(X) = \|X - E(X)\|_2^2$, ossia la Varianza è il quadrato della distanza in $L^2(\Omega, P)$ di X dalla costante $E(X)$: si può dunque interpretare come una misura della "dispersione" dei valori della variabile X attorno alla sua media. In particolare valgono i seguenti risultati.

Proposizione 3.32. Sia $X \in L^2(\Omega, P)$.

- (i) $Var(X) = 0$ se e solo se X è quasi certamente costante, cioè se esiste una costante $c \in \mathbb{R}$ tale che $P(X = c) = 1$.
- (ii) Per ogni $c \in \mathbb{R}$, $Var(X) \leq \|X - c\|_2^2$.

Dimostrazione. (i) La dimostrazione è semplice, ed è basata sull'osservazione che se $Y \in L^1(\Omega, P)$ e $Y \geq 0$, allora $E(Y) = 0$ se e solo se $P(Y = 0) = 1$. I dettagli sono lasciati come esercizio.

- (ii) Si consideri il polinomio in c

$$p(c) = \|X - c\|_2^2 = E[(X - c)^2] = c^2 - 2E(X)c + E(X^2).$$

Si vede subito che $p(c)$ ha minimo assoluto per $c = E(X)$, e $p(E(X)) = Var(X)$. □

Nella precedente proposizione, la (i) afferma che le variabili casuali con varianza zero sono costanti a meno di insiemi di probabilità zero, mentre la (ii) fornisce una caratterizzazione “variazionale” del valor medio, affermando che esso è la costante che realizza la distanza minima da X nel senso di L^2 . Va notato che, se assumiamo l'ipotesi (3.24), allora la Proposizione 3.32 (i) si potrebbe enunciare come segue: $Var(X) = 0$ se e solo se X è costante. L'ipotesi (3.24), sostanzialmente non restrittiva per spazi di probabilità discreti, è invece non ragionevole, come vedremo, in spazi di probabilità generali. Scegliamo pertanto di non fare tale assunzione, in modo da avere enunciati che resteranno validi nell'ambito della teoria generale.

3.6 Funzione generatrice dei momenti

Definizione 3.33. . Sia X una variabile casuale scalare. La funzione $\gamma_X : \mathbb{R} \rightarrow (0, +\infty]$ definita da

$$\gamma_X(t) = E(e^{tX})$$

è detta *funzione generatrice dei momenti* della variabile casuale X .

Osservazione 3.34. Notare che la funzione generatrice dei momenti di una variabile casuale può assumere il valore $+\infty$ per qualche valore di t , nel caso in cui la variabile casuale e^{tX} non ammetta valor medio. Si osservi, inoltre, che necessariamente $\gamma_X(0) = 1$

La funzione generatrice dei momenti gioca un ruolo importante nel calcolo delle probabilità, tuttavia in applicazioni che vanno al di là del contenuto di questo libro. Essa è utile, in alcuni casi, anche per il calcolo dei momenti di una variabile casuale, ove sia possibile far uso del seguente risultato.

Teorema 3.35. Sia X una variabile casuale scalare e γ_X la sua funzione generatrice. Supponiamo esista $a > 0$ tale che $\gamma_X(t) < +\infty$ per ogni $t \in (-a, a)$. Allora

- (i) La variabile casuale X ammette momenti di ogni ordine.
- (ii) γ_X è infinitamente derivabile in $(-a, a)$, e

$$\gamma_X^{(n)}(0) = E(X^n).$$

(iii) Esiste $\epsilon > 0$ tale che, per $t \in (-\epsilon, \epsilon)$, $\gamma_X(t)$ è dato dalla serie di Taylor

$$\gamma_X(t) = \sum_{n=0}^{+\infty} E(X^n) \frac{t^n}{n!}. \quad (3.25)$$

Dimostrazione. (i) Cominciamo con l'osservare che anche $\gamma_{|X|}(t) < +\infty$ per $t \in (-a, a)$. Per $t < 0$ la cosa è ovvia, visto che

$$e^{t|X|} \leq 1 \Rightarrow \gamma_{|X|}(t) \leq 1.$$

Per $0 < t < a$, poiché per ogni $x \in \mathbb{R}$ si ha $e^{|x|} \leq e^x + e^{-x}$, abbiamo che

$$\gamma_{|X|}(t) = E\left(e^{t|X|}\right) \leq E\left(e^{tX}\right) + E\left(e^{-tX}\right) = \gamma_X(t) + \gamma_X(-t).$$

Ora si noti che, per $x \in \mathbb{R}$, $n \geq 0$

$$e^{|x|} = \sum_{k=0}^{+\infty} \frac{|x|^k}{k!} \geq \frac{|x|^n}{n!}. \quad (3.26)$$

Sia ora $0 < b < a$. Usando (3.26) abbiamo

$$|X(\omega)|^n \leq \frac{n!}{b^n} e^{b|X(\omega)|}.$$

Perciò

$$E(|X|^n) \leq \frac{n!}{b^n} \gamma_{|X|}(b) < +\infty,$$

e quindi X ammette momenti di ogni ordine.

(ii) Osserviamo anzitutto che, per ogni $t \in (-a, a)$ e $n \geq 0$, la variabile casuale $|X|^n e^{tX}$ ammette valor medio (e perciò lo ammette anche $X^n e^{tX}$, essendo $|X|^n e^{tX} \leq |X|^n e^{|tX|}$). Scegliamo $h > 0$ tale che $|t| + h < a$. Usando (3.26) come al punto precedente, abbiamo che

$$|X|^n \leq \frac{n!}{h^n} e^{h|X|}.$$

Perciò,

$$|X^n| e^{tX} \leq \frac{n!}{h^n} e^{(|t|+h)|X|}. \quad (3.27)$$

La variabile casuale al membro destro della precedente uguaglianza ammette valor medio, per quanto mostrato all'inizio del punto precedente. Pertanto anche $|X|^n e^{tX}$ ammette valor medio.

A questo punto, mostriamo per induzione su $n \geq 0$ che

$$\gamma_X^{(n)}(t) = E\left(X^n e^{tX}\right) \quad (3.28)$$

per $t \in (-a, a)$, dove si noti che, per quanto appena dimostrato, il valor medio in (3.28) esiste. Per $n = 0$ non c'è nulla da dimostrare. Per il passo induttivo, si noti che, se $t \pm h \in (-a, a)$, usando l'ipotesi induttiva (3.28)

$$\frac{\gamma_X^{(n)}(t+h) - \gamma_X^{(n)}(t)}{h} = E\left(X^n e^{tX} \left[\frac{e^{hX} - 1}{h}\right]\right).$$

Pertanto

$$\left| \frac{\gamma_X^{(n)}(t+h) - \gamma_X^{(n)}(t)}{h} - E(X^{n+1}e^{tX}) \right| \leq E \left(|X|^n e^{tX} \left| \frac{e^{hX} - 1}{h} - X \right| \right). \quad (3.29)$$

A questo punto usiamo il fatto che, per ogni $x \in \mathbb{R}$,

$$|e^x - 1 - x| \leq \frac{x^2}{2} e^{|x|} \quad (3.30)$$

(verificarlo!). Ma allora, usando (3.30) in (3.29) si ha

$$\begin{aligned} \left| \frac{\gamma_X^{(n)}(t+h) - \gamma_X^{(n)}(t)}{h} - E(X^{n+1}e^{tX}) \right| &\leq hE \left(|X|^{n+2} e^{(|t|+|h|)|X|} \right) \\ &\leq hE \left(|X|^{n+2} e^{b|X|} \right), \end{aligned} \quad (3.31)$$

dove b è scelto in modo tale che $|t|+|h| < b < a$. Per quanto visto sopra, $E(|X|^{n+2}e^{b|X|}) < +\infty$, e quindi

$$\lim_{h \rightarrow 0} \left| \frac{\gamma_X^{(n)}(t+h) - \gamma_X^{(n)}(t)}{h} - E(X^{n+1}e^{tX}) \right| = 0,$$

che conclude la dimostrazione del passo induttivo.

- (iii) Usiamo una generalizzazione della disuguaglianza in (3.30), che consiste nello stimare il resto dell'espansione di Taylor per la funzione e^x :

$$\left| e^x - \sum_{k=0}^{n-1} \frac{x^k}{k!} \right| \leq \frac{|x|^n}{n!} e^{|x|}. \quad (3.32)$$

Poniamo ora $\epsilon = \frac{a}{2}$. Per $t \in (-\epsilon, \epsilon)$, usando (3.32), abbiamo che

$$\left| e^{tX} - \sum_{k=0}^{n-1} X^k \frac{t^k}{k!} \right| \leq \frac{|t|^n}{n!} |X|^n e^{|tX|}. \quad (3.33)$$

Perciò, usando la disuguaglianza $|E(Y)| \leq E(|Y|)$, otteniamo

$$\left| E(e^{tX}) - \sum_{k=0}^{n-1} E(X^k) \frac{t^k}{k!} \right| \leq E \left(\left| e^{tX} - \sum_{k=0}^{n-1} X^k \frac{t^k}{k!} \right| \right) \leq \frac{|t|^n}{n!} E(|X|^n e^{|tX|}). \quad (3.34)$$

A questo punto, usando (3.27) con $h = a/2$ abbiamo

$$\left| E(e^{tX}) - \sum_{k=0}^{n-1} E(X^k) \frac{t^k}{k!} \right| \leq \frac{|t|^n}{n!} \frac{n!}{(a/2)^n} E \left(e^{(|t|+\frac{a}{2})|X|} \right) = \left(\frac{t}{a/2} \right)^n E \left(e^{(|t|+\frac{a}{2})|X|} \right). \quad (3.35)$$

Poiché quest'ultimo valor medio è finito e $\left(\frac{t}{a/2}\right)^n$ tende a zero per $n \rightarrow +\infty$, si conclude che

$$\lim_{n \rightarrow +\infty} \left| E(e^{tX}) - \sum_{k=0}^{n-1} E(X^k) \frac{t^k}{k!} \right| = 0,$$

che implica

$$E(e^{tX}) = \lim_{n \rightarrow +\infty} \sum_{k=0}^{n-1} E(X^k) \frac{t^k}{k!} = \sum_{n=0}^{+\infty} E(X^n) \frac{t^n}{n!},$$

e la dimostrazione è conclusa. □

Osservazione 3.36. Con un po' di fatica in più su può dimostrare che la serie in (3.25) converge a $\gamma_X(t)$ per ogni $t \in (-a, a)$.

Osservazione 3.37. Nella dimostrazione del Teorema 3.35 non abbiamo mai usato la seguente espressione per la funzione generatrice dei momenti, che è quella che più useremo nei calcoli espliciti:

$$\gamma_X(t) = \sum_{x \in \mathbb{R}} e^{tx} p_X(x). \quad (3.36)$$

Nel caso in cui la variabile casuale X assuma un numero finito di valori, cioè $|X(\Omega)| < +\infty$, la somma in (3.36) è una somma finita. In questo caso, tutte le affermazioni del Teorema 3.35 si dimostrano facilmente da (3.36). Se invece $|X(\Omega)|$ è numerabile, non c'è alcun vantaggio ad usare (3.36) per dimostrare il Teorema 3.35. La dimostrazione qui data ha il vantaggio di rimanere inalterata nel contesto più generale, che vedremo più avanti, di variabili casuali definite in spazi di probabilità non discreti.

Più avanti in questo capitolo vedremo numerosi esempi di calcolo della funzione generatrice.

3.7 Disuguaglianze

Come in molti altri settori della matematica, in probabilità le disuguaglianze giocano un ruolo fondamentale. La prima che vediamo chiarisce il significato della varianza come indice della dispersione dei valori di una variabile casuale: la probabilità di una deviazione dalla media di una variabile casuale X si può stimare dall'alto con la sua varianza.

Proposizione 3.38. (*Disuguaglianza di Chebichev*)

(i) *Sia X una variabile casuale a valori reali positivi, che ammette valor medio. Allora, per ogni $\epsilon > 0$*

$$P(X \geq \epsilon) \leq \frac{E(X)}{\epsilon}.$$

(ii) *Sia $X \in L^2(\Omega, P)$. Allora, per ogni $\epsilon > 0$*

$$P(|X - E(X)| > \epsilon) \leq \frac{\text{Var}(X)}{\epsilon^2}.$$

Dimostrazione. (i) Tenendo conto che $X(\omega) \geq 0$ per ogni ω , si ha la disuguaglianza

$$X \geq \epsilon 1_{X \geq \epsilon}.$$

Per monotonia del valor medio:

$$E(X) \geq \epsilon E(1_{X \geq \epsilon}) = \epsilon P(X \geq \epsilon),$$

da cui la tesi segue subito.

(ii) Poichè

$$P(|X - E(X)| > \epsilon) = P((X - E(X))^2 > \epsilon^2),$$

è sufficiente applicare quanto dimostrato in (i) a $(X - E(X))^2$ e ϵ^2 . □

La seguente disuguaglianza è utile quando si voglia confrontare il valor medio di una variabile casuale con quelli di sue opportune funzioni.

Proposizione 3.39. (*Disuguaglianza di Jensen*) Sia X una variabile casuale scalare e $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ una funzione convessa. Si assuma che le variabili casuali X e $\varphi(X)$ ammettano valor medio. Allora

$$\varphi(E(X)) \leq E(\varphi(X)).$$

Dimostrazione. Il fatto che φ sia convessa è equivalente ad affermare che, per ogni $x_0 \in \mathbb{R}$, esiste $\lambda(x_0) \in \mathbb{R}$ tale che per ogni $x \in \mathbb{R}$

$$\varphi(x) \geq \varphi(x_0) + \lambda(x_0)(x - x_0) \tag{3.37}$$

(il lettore verifichi quest'ultima affermazione). Posto $x = X(\omega)$ e $x_0 = E(X)$ in (3.37) otteniamo, per tutti gli $\omega \in \Omega$

$$\varphi(X(\omega)) \geq \varphi(E(X)) + \lambda(E(X))(X - E(X)). \tag{3.38}$$

Prendendo il valor medio dei due membri di (3.38) la conclusione segue subito. □

Un esempio di applicazione della disuguaglianza di Jensen è il seguente confronto tra norme L^p di una variabile casuale.

Corollario 3.40. Sia $X \in L^q(\Omega, P)$ e sia $1 \leq p \leq q$. Allora

$$\|X\|_p \leq \|X\|_q.$$

Dimostrazione. È sufficiente applicare la disuguaglianza di Jensen alla variabile casuale $|X|$ e alla funzione convessa $\varphi(x) = |x|^{q/p}$. □

L'ultima disuguaglianza di questo paragrafo è utile ogni qual volta si voglia stimare la media di un prodotto di variabili casuali. Un'applicazione importante la vedremo nel prossimo paragrafo.

Proposizione 3.41. (*Disuguaglianza di Cauchy-Schwartz*) Siano $X, Y \in L^2(\Omega, P)$. Allora XY ammette valor medio, e

$$|E(XY)| \leq \sqrt{E(X^2)E(Y^2)}. \tag{3.39}$$

L'uguaglianza in (3.39) vale se e solo se esiste $c \in \mathbb{R}$ tale che $P(X = cY) = 1$.

Dimostrazione. Per mostrare che XY ammette valor medio, si osservi che per ogni $x, y \in \mathbb{R}$

$$|xy| \leq \frac{1}{2}x^2 + \frac{1}{2}y^2,$$

come si vede dal fatto che

$$\frac{1}{2}x^2 + \frac{1}{2}y^2 - |xy| = \frac{1}{2}(|x| - |y|)^2 \geq 0.$$

Ma allora

$$|XY| \leq \frac{1}{2}X^2 + \frac{1}{2}Y^2,$$

da cui

$$E(|XY|) \leq \frac{1}{2}E(X^2) + \frac{1}{2}E(Y^2) < +\infty.$$

Per mostrare la disuguaglianza (3.39) è necessario un argomento più fine. La disuguaglianza (3.39) fa parte della classe di disuguaglianze di Cauchy-Schwartz, che si dimostrano tutte in modo analogo.

Anzitutto, non è restrittivo supporre $P(X = 0) < 1$, e $P(Y = 0) < 1$. Equivalentemente $E(X^2) > 0$, $E(Y^2) > 0$. In caso contrario la (3.39) è banalmente verificata ($0 = 0!$). Inoltre possiamo assumere $E(XY) \geq 0$. In caso contrario è sufficiente rimpiazzare X con $-X$, e notare che l'intero enunciato non viene modificato da tale rimpiazzamento. Poniamo allora

$$X_* = \frac{X}{\sqrt{E(X^2)}}, \quad Y_* = \frac{Y}{\sqrt{E(Y^2)}}.$$

Si noti che

$$\begin{aligned} 0 \leq E[(X_* - Y_*)^2] &= E(X_*^2) + E(Y_*^2) - 2E(X_*Y_*) \\ &= 2 - 2\frac{E(XY)}{\sqrt{E(X^2)E(Y^2)}}, \end{aligned}$$

da cui (3.39) segue.

Supponiamo ora che (3.39) valga come uguaglianza. Allora, per quanto appena visto, $E[(X_* - Y_*)^2] = 0$ che, come osservato nella dimostrazione della Proposizione 3.32, equivale a

$$P(X_* = Y_*) = 1 \iff P\left(X = \frac{\sqrt{E(X^2)}}{\sqrt{E(Y^2)}}Y\right) = 1,$$

cioè $P(X = cY) = 1$ con $c = \frac{\sqrt{E(X^2)}}{\sqrt{E(Y^2)}}$.

Viceversa, se $P(X = cY) = 1$, allora (perché?) $E(XY) = cE(Y^2)$, $E(X^2) = c^2E(Y^2)$, da cui si verifica che la (3.39) vale come uguaglianza. \square

3.8 Covarianza e coefficiente di correlazione

Media, varianza e momenti sono quantità che si riferiscono ad una singola variabile casuale. L'oggetto che ora definiamo è un indice della relazione che sussiste tra due variabili casuali.

Definizione 3.42. Siano X, Y due variabili casuali scalari definite sullo stesso spazio di probabilità (Ω, P) , che ammettono valor medio. Se la variabile casuale $(X - E(X))(Y - E(Y))$ ammette media, la quantità

$$Cov(X, Y) = E[(X - E(X))(Y - E(Y))]$$

si dice *covarianza* tra X e Y .

Si noti che ogni qual volta X, Y e XY ammettono valor medio, allora la covarianza è ben definita, e

$$\text{Cov}(X, Y) = E(XY) - E(X)E(Y).$$

Proposizione 3.43. *Siano $X, Y \in L^2(\Omega, P)$. Allora $\text{Cov}(X, Y)$ è ben definita. Inoltre*

$$|\text{Cov}(X, Y)| \leq \sqrt{\text{Var}(X)\text{Var}Y}. \quad (3.40)$$

Infine, assumendo $\text{Var}(X) > 0, \text{Var}(Y) > 0, \text{Cov}(X, Y) = \sqrt{\text{Var}(X)\text{Var}Y}$ (risp. $\text{Cov}(X, Y) = -\sqrt{\text{Var}(X)\text{Var}Y}$) se e solo se esistono costanti $a > 0$ (risp. $a < 0$) e $b \in \mathbb{R}$ tali che

$$P(X = aY + b) = 1. \quad (3.41)$$

Dimostrazione. Si applica la Proposizione 3.41 alle variabili casuali $\bar{X} = X - E(X)$ e $\bar{Y} = Y - E(Y)$. Si ottiene immediatamente l'esistenza della covarianza, e la disuguaglianza (3.40) segue da (3.39). Inoltre, se (3.40) vale come uguaglianza, allora (3.39) vale come uguaglianza per \bar{X} e \bar{Y} , e quindi esiste $c \in \mathbb{R}$ per cui $P(\bar{X} = c\bar{Y}) = 1$. Notare che dev'essere $c \neq 0$, essendo $E(\bar{X}^2) = \text{Var}(X) > 0$. Ma allora

$$P(X = cY + E(X) - cE(Y)) = 1,$$

cioè (3.41) vale con $a = c$ e $b = E(X) - cE(Y)$. Infine, un calcolo diretto simile a quello fatto nella Proposizione 3.41, mostra che se vale (3.41) con $a \neq 0$ allora

$$\text{Cov}(X, Y) = \frac{a}{|a|} \sqrt{\text{Var}(X)\text{Var}Y},$$

e la dimostrazione è conclusa. □

Definizione 3.44. Siano $X, Y \in L^2(\Omega, P)$, tali che $\text{Var}(X) > 0, \text{Var}(Y) > 0$. La quantità

$$\rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}Y}}$$

Si dice *coefficiente di correlazione* tra X e Y . Se $\rho_{X,Y} = 0$ (equivalentemente $\text{Cov}(X, Y) = 0$) diremo che le variabili casuali X e Y sono *scorrelate*

Le seguenti proprietà del coefficiente di correlazione discendono immediatamente dalla Proposizione 3.43

Proposizione 3.45. *Siano $X, Y \in L^2(\Omega, P)$, tali che $\text{Var}(X) > 0, \text{Var}(Y) > 0$. Allora $|\rho(X, Y)| \leq 1$. Inoltre $\rho(X, Y) = 1$ (risp. $\rho(X, Y) = -1$) se e solo se esistono $a > 0$ (risp. $a < 0$) e $b \in \mathbb{R}$ tali che*

$$P(X = aY + b) = 1.$$

Il coefficiente di correlazione è un'indice del grado di correlazione *lineare* tra le variabili X e Y , ossia di quanto "bene" X possa essere approssimata da funzioni affini di Y . Tale affermazione è di ovvio significato quando $|\rho(X, Y)| = 1$, vista la Proposizione 3.45. Per capire il caso generale, formuliamo il seguente problema. Siano $X, Y \in L^2(\Omega, P)$, con $\text{Var}(X) > 0, \text{Var}(Y) > 0$, due variabili casuali fissate, e cerchiamo di determinare la funzione affine di Y

che meglio approssima X del senso della distanza $\| \cdot - \cdot \|_2$. In altre parole, cerchiamo le costanti $a, b \in \mathbb{R}$ tali che la distanza

$$\|X - (aY + b)\|_2$$

sia la minima possibile. Questo equivale a minimizzare la funzione, nelle variabili a, b

$$\varphi(a, b) = \|X - (aY + b)\|_2^2 = E[(X - aY - b)^2] = E(Y^2 a^2 + b^2 + 2E(Y)ab - 2E(XY)a - 2E(X)b + E(X^2)).$$

Con le tecniche standard del calcolo differenziale in più variabili, si vede che $\varphi(a, b)$ ammette minimo assoluto in corrispondenza delle soluzioni del sistema

$$\begin{aligned} \frac{\partial}{\partial a} \varphi(a, b) &= 0 \\ \frac{\partial}{\partial b} \varphi(a, b) &= 0 \end{aligned}$$

la cui unica soluzione è

$$\begin{aligned} a^* &= \frac{Cov(X, Y)}{Var(Y)} \\ b^* &= E(X) - \frac{E(Y)Cov(X, Y)}{Var(Y)}. \end{aligned}$$

Il valore del minimo assoluto è

$$\varphi(a^*, b^*) = Var(X)[1 - \rho^2(X, Y)].$$

Dunque X è tanto meglio approssimabile da funzioni affini di Y quanto più vicino ad uno è $\rho^2(X, Y)$. Se, viceversa, le variabili X e Y sono scorrelate, allora $a^* = 0$, cioè la migliore approssimazione di X con funzioni affini di Y *non* dipende da Y , ossia è una costante.

3.9 Valor medio e indipendenza

Nella proposizione che segue, mostriamo che l'indipendenza implica una proprietà moltiplicativa del valor medio.

Proposizione 3.46. *Siano X, Y variabili casuali scalari indipendenti, definite nello stesso spazio di probabilità e che ammettono valor medio. Allora XY ammette valor medio e*

$$E(XY) = E(X)E(Y).$$

Dimostrazione. Cominciamo col mostrare che XY ammette valor medio. Usando il fatto che $p_{X,Y}(x, y) = p_X(x)p_Y(y)$ si ha:

$$\begin{aligned} E(|XY|) &= \sum_{x,y} |x||y|p_{X,Y}(x, y) = \sum_x |x|p_X(x) \sum_y |y|p_Y(y) \\ &= E(|X|)E(|Y|) < +\infty. \end{aligned}$$

La dimostrazione che $E(XY) = E(X)E(Y)$ è del tutto analoga. □

Corollario 3.47. *Siano X, Y due variabili casuali scalari indipendenti, che ammettono valor medio. Allora $Cov(X, Y)$ è ben definita, e vale zero.*

Il Corollario 3.47 segue immediatamente dalla Proposizione 3.46. In particolare, ne segue che due variabili casuali scalari indipendenti che ammettono media, sono scorrelate. Il viceversa non è necessariamente vero, come mostra l'esempio che segue.

Esempio 3.48. Sia Z una variabile casuale tale che

$$p_Z(0) = p_Z(\pi/2) = p_Z(\pi) = \frac{1}{3}.$$

Posto $X = \sin(Z)$ e $Y = \cos(Z)$, si vede subito che $XY \equiv 0$ e $E(Y) = 0$, da cui

$$Cov(X, Y) = 0$$

segue immediatamente. Inoltre

$$P(X = 1, Y = 1) = 0 \neq \frac{1}{9} = P(X = 1)P(Y = 1).$$

Chiudiamo questo paragrafo con un risultato semplice ma utile. È facile vedere (provare!) in generale che se due variabili casuali X, Y ammettono momento secondo, allora

$$Var(X + Y) = Var(X) + Var(Y) + 2Cov(X, Y).$$

Una semplice dimostrazione per induzione, dimostra allora quanto segue.

Corollario 3.49. *Siano X_1, X_2, \dots, X_n variabili casuali che amettono momento secondo, tali che $Cov(X_i, X_j) = 0$ se $i \neq j$. Allora*

$$Var(X_1 + \dots + X_n) = \sum_{i=1}^n Var(X_i).$$

Per il Corollario 3.47, il risultato del Corollario 3.49 vale, in particolare, se le variabili X_1, X_2, \dots, X_n sono indipendenti.

Corollario 3.50. *Siano X_1, X_2, \dots, X_n variabili casuali scalari indipendenti, e siano γ_{X_i} le relative funzioni generatrici. Allora, per ogni $t \in \mathbb{R}$ per cui $\gamma_{X_i}(t) < +\infty$ per ogni $i = 1, 2, \dots, n$, si ha*

$$\gamma_{X_1+X_2+\dots+X_n}(t) = \prod_{i=1}^n \gamma_{X_i}(t).$$

Dimostrazione. Per la Proposizione 3.18, fissato $t \in \mathbb{R}$, le variabili casuali $e^{tX_1}, e^{tX_2}, \dots, e^{tX_n}$ sono indipendenti. È perciò sufficiente applicare ricorsivamente la Proposizione 3.46. \square

3.10 Classi notevoli di variabili casuali discrete

3.10.1 Variabili casuali di Bernoulli

Una variabile casuale X è detta *variabile casuale di Bernoulli* se prende valori nell'insieme $\{0, 1\}$. Posto

$$p = p_X(1),$$

ne segue che $p_X(0) = 1 - p$. La distribuzione di X è dunque completamente determinata dal parametro $p \in [0, 1]$. Diremo allora che X è una variabile casuale di Bernoulli di parametro p , e scriveremo

$$X \sim Be(p).$$

È immediato vedere che

$$E(X) = E(X^2) = p,$$

e dunque

$$Var(X) = E(X^2) - E^2(X) = p(1 - p).$$

3.10.2 Variabili casuali binomiali

Una variabile casuale X si dice *variabile casuale Binomiale* di parametri $n \geq 1$ e $p \in [0, 1]$, e si scrive

$$X \sim B(n, p),$$

se X ha la medesima distribuzione di una variabile casuale del tipo $X_1 + \cdots + X_n$, dove $X_i \sim Be(p)$, e le X_i sono tra loro indipendenti. È facile calcolare la densità di X . Se le X_i rappresentano gli esiti di n prove ripetute indipendenti per ognuna delle quali l'esito positivo ($X_i = 1$) ha probabilità p , l'evento $\{X_1 + \cdots + X_n = k\}$ rappresenta la probabilità di avere ottenuto k esiti positivi. Tale probabilità è stata calcolata nell'Esempio 1.37:

$$p_X(k) = P(X_1 + \cdots + X_n = k) = \binom{n}{k} p^k (1 - p)^{n-k}.$$

Inoltre

$$E(X) = E(X_1 + \cdots + X_n) = np,$$

e, usando il Corollario 3.49

$$Var(X) = \sum_{i=1}^n Var(X_i) = np(1 - p).$$

Un'ultima, rilevante proprietà delle variabili casuali binomiali è la seguente.

Proposizione 3.51. *Siano $X \sim B(n, p)$ e $Y \sim B(m, p)$ due variabili casuali indipendenti. Allora $X + Y \sim B(n + m, p)$.*

Dimostrazione. Siano $X_1, X_2, \dots, X_n, X_{n+1}, \dots, X_{n+m}$ variabili casuali di Bernoulli di parametro p indipendenti. La distribuzione congiunta di (X, Y) è uguale (perché?) alla distribuzione congiunta di $(X_1 + \cdots + X_n, X_{n+1} + \cdots + X_{n+m})$. Ne segue che $X + Y$ ha la stessa distribuzione di $X_1 + \cdots + X_n + X_{n+1} + \cdots + X_{n+m}$ che, per definizione, è una variabile casuale binomiale di parametri $n + m$ e p . \square

3.10.3 Variabili casuali Geometriche

Abbiamo visto nell'Esempio 1.37 che, se si eseguono N prove ripetute indipendenti per ognuna delle quali l'esito positivo ha probabilità p , allora la probabilità di ottenere il primo esito positivo al tentativo $n + 1$ -esimo, con $n < N$, è $p(1 - p)^n$. Per liberarsi della restrizione $n < N$, immaginiamo di eseguire una successione infinita di prove ripetute $N = +\infty$. C'è un problema, però: l'insieme che descrive gli esiti di una successione infinita di prove, $\Omega = \{0, 1\}^{\mathbb{N} \setminus \{0\}}$, non è numerabile, e dunque non può essere lo spazio campionario di un spazio di probabilità discreto. Come vedremo nel prossimo capitolo, introducendo una nozione più generale di spazio di probabilità, tale problema può essere superato. Tralasciando i dettagli tecnici è possibile definire una variabile casuale X tale che l'evento $\{X = n\}$ corrisponde a "il primo esito positivo è stato ottenuto al tentativo $n + 1$ -esimo", ed ha probabilità $p(1 - p)^n$. In altre parole X è il *numero di tentativi falliti prima del primo successo*.

In generale, diremo che una variabile casuale X a valori in \mathbb{N} è una variabile casuale Geometrica di parametro $p \in (0, 1)$, e scriveremo

$$X \sim Ge(p),$$

se

$$p_X(n) = p(1 - p)^n,$$

per ogni $n \geq 0$. Le variabili casuali Geometriche godono di una proprietà rilevante, detta *perdita di memoria*, che si può esprimere con l'affermazione: *se nei primi n tentativi non è stato ottenuto alcun successo, la probabilità di dover attendere altri m tentativi prima del primo successo non dipende da n* . La conseguente inutilità di puntare, ad esempio, sui numeri ritardatari nel gioco del Lotto è, spero, evidente!

Proposizione 3.52. *Sia $X \sim Ge(p)$. Allora, per ogni $n, m \geq 0$,*

$$P(X \geq n + m | X \geq n) = P(X \geq m). \quad (3.42)$$

Dimostrazione. Anzitutto si osservi che, poichè $\{X \geq n + m\} \subseteq \{X \geq n\}$

$$\begin{aligned} P(X \geq n + m | X \geq n) &= \frac{P(\{X \geq n + m\} \cap \{X \geq n\})}{P(\{X \geq n\})} \\ &= \frac{P(\{X \geq n + m\})}{P(\{X \geq n\})}. \end{aligned}$$

Dobbiamo dunque calcolare probabilità del tipo $P(X \geq n)$. Si ottiene:

$$P(X \geq n) = \sum_{k=n}^{+\infty} p(1 - p)^k = p(1 - p)^n \sum_{h=0}^{+\infty} (1 - p)^h = (1 - p)^n,$$

dove si è usato il fatto di saper calcolare la somma di una serie geometrica:

$$\sum_{h=0}^{+\infty} (1 - p)^h = \frac{1}{1 - (1 - p)} = \frac{1}{p}.$$

Ma allora:

$$\begin{aligned}
 P(X \geq n+m | X \geq n) &= \frac{P(\{X \geq n+m\})}{P(\{X \geq n\})} \\
 &= \frac{(1-p)^{n+m}}{(1-p)^n} \\
 &= (1-p)^m \\
 &= P(X \geq m).
 \end{aligned}$$

□

La proprietà di perdita di memoria, caratterizza le variabili casuali Geometriche nella classe delle variabili casuali a valori naturali, come mostriamo nella seguente Proposizione.

Proposizione 3.53. *Sia X una variabile casuale a valori in \mathbb{N} , tale che (3.42) è verificata per ogni $n, m \geq 0$. Allora X è una variabile casuale Geometrica oppure $P(X=0) = 1$.*

Dimostrazione. Posto $m = 1$ in (3.42), si ottiene

$$P(X \geq n+1) = P(X \geq n)P(X \geq 1)$$

che, per induzione, implica

$$P(X \geq n) = P(X \geq 1)^n,$$

per ogni $n \geq 1$. Ma allora, osservando che $\{X = n\} = \{X \geq n\} \setminus \{X \geq n+1\}$,

$$p_X(n) = P(X = n) = P(X \geq n) - P(X \geq n+1) = P(X \geq 1)^n - P(X \geq 1)^{n+1} = (1 - P(X \geq 1))P(X \geq 1)^n,$$

per ogni $n \geq 1$. Inoltre

$$p_X(0) = P(X = 0) = 1 - P(X \geq 1).$$

Dunque, posto $p = 1 - P(X \geq 1)$,

$$p_X(n) = p(1-p)^n$$

per ogni $n \geq 0$, che conclude la dimostrazione. □

Calcoliamo ora media e varianza di una variabile casuale Geometrica. A questo scopo calcoliamo la funzione generatrice dei momenti di una variabile casuale geometrica.

$$\gamma_X(t) = \sum_{n=0}^{+\infty} e^{tn} p(1-p)^n = p \sum_{n=0}^{+\infty} ((1-p)e^t)^n.$$

quest'ultima è una serie geometrica di ragione $(1-p)e^t$ cioè $t < -\log(1-p)$. Perciò

$$\gamma_X(t) = \begin{cases} \frac{p}{1-(1-p)e^t} & \text{se } t < -\log(1-p) \\ +\infty & \text{altrimenti.} \end{cases}$$

Per il Teorema 3.35, $E(X^n) = \gamma_X^{(n)}(0)$ per ogni $n \geq 1$. Essendo

$$\gamma_X'(t) = \frac{p(1-p)e^t}{(1-(1-p)e^t)^2},$$

$$\gamma_X''(t) = \frac{p(1-p)e^t \left[(1 - (1-p)e^t)^2 + 2(1 - (1-p)e^t)(1-p)e^t \right]}{(1 - (1-p)e^t)^4},$$

si trova

$$E(X) = \gamma'(0) = \frac{1-p}{p}, \quad E(X^2) = \gamma''(0) = \frac{(1-p)(2-p)}{p^2},$$

da cui segue

$$\text{Var}(X) = E(X^2) - E^2(X) = \frac{1-p}{p^2}.$$

3.10.4 Variabili casuali di Poisson

In numerose situazioni concrete, ci si trova a considerare delle variabili casuali binomiali i cui parametri, n e p , sono tali che n è molto grande e p è molto piccolo. In altre parole, si eseguono numerose ripetizioni di una prova che ha esito positivo con probabilità piccola. Supponiamo, ad esempio, di considerare il numero di accessi, in un fissato intervallo di tempo, ad un certo servizio (lo sportello di un ufficio pubblico, un pagina web, un centralino...). Vi sarà un numero grande di utenti, diciamo n , che ha accesso a tale servizio. Si osserva però che, tipicamente, un numero di utenti molto minore di n accede effettivamente al servizio in un intervallo di tempo della lunghezza di quello fissato. Un ragionevole modello matematico per tale situazione, è di supporre che ogni utente abbia una probabilità piccola, $p \ll 1$, di accedere al servizio. In una approssimazione che è assai buona in molti casi, assumiamo che il valore di p sia uguale per tutti gli utenti, e che ogni utente si comporti in modo indipendente dagli altri. Sotto queste ipotesi, il numero X di utenti che effettivamente accede al servizio è una variabile casuale binomiale di parametri n e p , cioè

$$p_X(k) = \binom{n}{k} p^k (1-p)^{n-k}.$$

Se n è molto grande, calcoli espliciti con questa densità risultano estremamente pesanti, se non impraticabili, data la difficoltà di calcolare i coefficienti binomiali per grandi valori di n . È allora interessante analizzare il comportamento asintotico di $p_X(k)$ quando $n \rightarrow +\infty$ e $p \rightarrow 0$. Per ottenere un comportamento limite non banale, è necessario che p vada a zero "proporzionalmente" a $\frac{1}{n}$. Per semplicità poniamo semplicemente $p = \frac{\lambda}{n}$ con $\lambda > 0$, ma il calcolo che segue si può facilmente modificare per coprire il caso $p = p_n$ con $\lim_{n \rightarrow +\infty} np_n = \lambda > 0$. Si ha

$$\begin{aligned} p_X(k) &= \frac{n!}{k!(n-k)!} \frac{\lambda^k}{n^k} \left(1 - \frac{\lambda}{n}\right)^{n-k} \\ &= \frac{\lambda^k n(n-1) \cdots (n-k+1)}{k! n^k} \frac{1}{\left(1 - \frac{\lambda}{n}\right)^k} \left(1 - \frac{\lambda}{n}\right)^n. \end{aligned}$$

D'altra parte

$$\begin{aligned} \lim_{n \rightarrow +\infty} \frac{n(n-1) \cdots (n-k+1)}{n^k} &= 1, \\ \lim_{n \rightarrow +\infty} \left(1 - \frac{\lambda}{n}\right)^k &= 1 \end{aligned}$$

e

$$\lim_{n \rightarrow +\infty} \left(1 - \frac{\lambda}{n}\right)^n = e^{-\lambda}.$$

Ne segue che

$$\lim_{n \rightarrow +\infty} p_X(k) = e^{-\lambda} \frac{\lambda^k}{k!}.$$

Si noti che l'espressione appena ottenuta è la densità di una variabile casuale a valori in \mathbb{N} , in quanto

$$\sum_{n=0}^{\infty} e^{-\lambda} \frac{\lambda^n}{n!} = 1.$$

Tale procedura di approssimazione giustifica la definizione di una nuova classe di variabili casuali. Una variabile casuale X a valori in \mathbb{N} si dice variabile casuale di Poisson di parametro $\lambda > 0$, e si scrive

$$X \sim Po(\lambda),$$

se

$$p_X(k) = e^{-\lambda} \frac{\lambda^k}{k!}$$

per ogni $k \in \mathbb{N}$. Tali variabili casuali vengono comunemente usate per modellare, tra le altre, quantità del tipo “numero di accessi ad un servizio”, come abbiamo sopra giustificato.

È agevole calcolare la funzione generatrice dei momenti di $X \sim Po(\lambda)$:

$$\gamma_X(t) = e^{-\lambda} \sum_{n=0}^{+\infty} \frac{(e^t n)^n}{n!} = e^{\lambda(e^t - 1)}.$$

Ne segue che

$$\gamma'_X(t) = \lambda e^t \gamma_X(t), \quad \gamma''_X(t) = \lambda e^t [\gamma_X(t) + \gamma'_X(t)],$$

da cui $E(X) = \gamma'_X(0) = \lambda$, $E(X^2) = \gamma''_X(0) = \lambda + \lambda^2$ e perciò

$$Var(X) = \lambda.$$

Il risultato che ora illustriamo afferma che la somma di due variabili casuali di Poisson indipendenti è ancora una variabile casuale di Poisson.

Proposizione 3.54. *Siano $X \sim Po(\lambda)$ e $Y \sim Po(\mu)$ variabili casuali indipendenti. Allora $X + Y \sim Po(\lambda + \mu)$.*

Prima di dimostrare la Proposizione 3.54, mostriamo un risultato del tutto generale.

Proposizione 3.55. *Siano X e Y due variabili casuali discrete scalari, e sia $p_{X,Y}$ la loro densità congiunta. Allora*

$$p_{X+Y}(z) = P(X + Y = z) = \sum_x p_{X,Y}(x, z - x) = \sum_y p_{X,Y}(z - y, y). \quad (3.43)$$

Dimostrazione. Usando la σ -additività della probabilità, si ha

$$\begin{aligned} P(X + Y = z) &= P\left(\bigcup_{x \in X(\Omega)} \{X = x, Y = z - x\}\right) \\ &= \sum_{x \in X(\Omega)} P(X = x, Y = z - x) \\ &= \sum_x p_{X,Y}(x, z - x), \end{aligned}$$

ove si è osservato che $P(X = x, Y = z - x) = 0$ se $x \notin X(\Omega)$. L'ultima uguaglianza in (3.43) si ottiene scambiando i ruoli di X e Y . \square

Nel caso in cui le variabili casuali X e Y nella Proposizione 3.55 siano indipendenti, la (3.43) assume la forma

$$p_{X+Y}(z) = \sum_x p_X(x)p_Y(z - x) = \sum_y p_X(z - y)p_Y(y), \quad (3.44)$$

che ora useremo.

Dimostrazione della Proposizione 3.54. Per la (3.44) si ha

$$\begin{aligned} p_{X+Y}(n) &= \sum_{k=0}^n e^{-\lambda} e^{-\mu} \frac{\lambda^k}{k!} \frac{\mu^{n-k}}{(n-k)!} \\ &= e^{-(\lambda+\mu)} \frac{1}{n!} \sum_{k=0}^n \binom{n}{k} \lambda^k \mu^{n-k} \\ &= e^{-(\lambda+\mu)} \frac{(\lambda + \mu)^n}{n!}, \end{aligned}$$

dove nell'ultima uguaglianza abbiamo applicato la formula di Newton per la potenza di un binomio. \square

3.10.5 Funzione di ripartizione

Definizione 3.56. Sia X una variabile casuale discreta scalare. La funzione $F_X : \mathbb{R} \rightarrow [0, 1]$ definita da

$$F_X(x) = P(X \leq x)$$

si dice *funzione di ripartizione* della variabile casuale X .

Alcune proprietà elementari delle funzione di ripartizione sono mostrate nella seguente proposizione.

Proposizione 3.57. Sia F_X la funzione di ripartizione di una variabile casuale discreta X . Allora

- (i) F_X è non decrescente
- (ii) F_X è continua da destra.

(iii)

$$\lim_{x \rightarrow -\infty} F_X(x) = 0.$$

(iv)

$$\lim_{x \rightarrow +\infty} F_X(x) = 1.$$

Dimostrazione. (i) Se $x < y$ allora $\{X \leq x\} \subseteq \{X \leq y\}$. Perciò

$$F_X(x) = P(X \leq x) \leq P(X \leq y) = F_X(y).$$

(ii) Sia $x \in \mathbb{R}$. Basta dimostrare che se (x_n) è una successione decrescente tale che $x_n \downarrow x$ allora

$$\lim_{n \rightarrow +\infty} F_X(x_n) = F(x). \quad (3.45)$$

Si osservi che

$$\{X \leq x\} = \bigcap_n \{X \leq x_n\},$$

e che la successione di eventi $\{X \leq x_n\}$ è decrescente. Per la Proposizione 1.10 si ha

$$F_X(x) = P(X \leq x) = \lim_{n \rightarrow +\infty} P(X \leq x_n) = \lim_{n \rightarrow +\infty} F_X(x_n).$$

(iii) È sufficiente applicare la Proposizione 1.10 alla famiglia decrescente di eventi $\{X \leq x_n\}$ dove $x_n \downarrow -\infty$, osservando che

$$\bigcap_n \{X \leq x_n\} = \emptyset.$$

(iv) È sufficiente applicare la Proposizione 1.10 alla famiglia crescente di eventi $\{X \leq x_n\}$ dove $x_n \uparrow +\infty$, osservando che

$$\bigcup_n \{X \leq x_n\} = \Omega.$$

□

Per quanto visto nell'Osservazione 3.7, la funzione di ripartizione si può esprimere in termini della densità come segue:

$$F_X(x) = \sum_{y \leq x} p_X(y).$$

Viceversa, mostriamo che la densità è esprimibile in termini della funzione di ripartizione. In particolare, la funzione di ripartizione determina completamente la distribuzione di una variabile casuale.

Proposizione 3.58. *Se X è una variabile casuale discreta scalare, per ogni $x \in \mathbb{R}$ vale la relazione*

$$p_X(x) = F_X(x) - F_X(x^-),$$

dove

$$F_X(x^-) = \lim_{y \rightarrow x^-} F_X(y).$$

Dimostrazione. Se $y < x$, si ha che

$$P(X \in (y, x]) = P(X \leq x) - P(X \leq y) = F_X(x) - F_X(y).$$

Allora, se (y_n) è una successione di elementi di $(-\infty, x)$ tali che $y_n \uparrow x$, poichè gli eventi $\{X \in (y_n, x]\}$ formano una successione decrescente per cui $\bigcap_n \{X \in (y_n, x]\} = \{X = x\}$, si ha

$$\begin{aligned} p_X(x) = P(X = x) &= \lim_{n \rightarrow +\infty} P(X \in (y_n, x]) = \lim_{n \rightarrow +\infty} [F_X(x) - F_X(y_n)] = F_X(x) - \lim_{n \rightarrow +\infty} F_X(y_n) \\ &= F_X(x) - F_X(x^-). \end{aligned}$$

□

Osservazione 3.59. Dalla proposizione precedente segue che F_X è discontinua in x se e solo se $p_X(x) > 0$, e il valore di $p_X(x)$ rappresenta l'entità del "salto" di F_X in x . Nel caso in cui $X(\Omega)$ non abbia punti di accumulazione, dall'identità

$$F_X(x) = \sum_{y \leq x} p_X(y).$$

segue che F_X è una funzione costante a tratti.

La funzione di ripartizione è spesso utile nei calcoli con le distribuzioni di variabili casuali. In particolare, come ora vedremo, è usata quando si tratta di calcolare la distribuzione del massimo o del minimo di variabili casuali indipendenti.

Proposizione 3.60. *Siano X_1, X_2, \dots, X_n variabili casuali scalari indipendenti. Definiamo:*

$$Z = \max(X_1, \dots, X_n), \quad W = \min(X_1, \dots, X_n).$$

Allora

$$F_Z(x) = \prod_{k=1}^n F_{X_k}(x), \tag{3.46}$$

e

$$F_W(x) = 1 - \prod_{k=1}^n [1 - F_{X_k}(x)]. \tag{3.47}$$

Dimostrazione. Cominciamo col dimostrare (3.46). Si osservi che

$$\{Z \leq x\} = \bigcap_{k=1}^n \{X_k \leq x\},$$

usando l'indipendenza delle X_k si ha

$$F_Z(x) = P(Z \leq x) = \prod_{k=1}^n P(X_k \leq x) = \prod_{k=1}^n F_{X_k}(x).$$

Per quanto riguarda (3.47), si vede che

$$\{W > x\} = \bigcap_{k=1}^n \{X_k > x\}.$$

e perciò

$$F_W(x) = 1 - P(W > x) = 1 - \prod_{k=1}^n P(X_k > x) = 1 - \prod_{k=1}^n [1 - P(X_k \leq x)] = 1 - \prod_{k=1}^n [1 - F_{X_k}(x)].$$

□

Esempio 3.61. Siano $X_1, \dots, X_n \sim Ge(p)$ indipendenti. Determiniamo la densità di $Z = \max(X_1, \dots, X_n)$ e $W = \min(X_1, \dots, X_n)$.

Come si vede dai calcoli fatti nella Proposizione 3.52,

$$F_{X_i}(k) = 1 - (1 - p)^{k+1}.$$

Dunque

$$F_Z(k) = \left[1 - (1 - p)^{k+1}\right]^n.$$

Essendo, per $k \in \mathbb{N}$, $F_Z(k^-) = F_Z(k - 1)$, si ha, usando la Proposizione 3.58

$$p_Z(k) = F_Z(k) - F_Z(k - 1) = \left[1 - (1 - p)^{k+1}\right]^n - \left[1 - (1 - p)^k\right]^n.$$

Analogamente,

$$F_W(k) = 1 - \left[(1 - p)^{k+1}\right]^n = 1 - [(1 - p)^n]^{k+1},$$

che coincide con la funzione di ripartizione di una variabile casuale Geometrica di parametro $1 - (1 - p)^n$. Poichè la funzione di ripartizione individua completamente la distribuzione, possiamo concludere che

$$W \sim Ge(1 - (1 - p)^n).$$

Capitolo 4

Spazi di probabilità generali (cenni)

4.1 σ -algebre e misure di probabilità

Con gli spazi di probabilità discreti, abbiamo visto molti aspetti importanti del Calcolo delle Probabilità, senza dover ricorrere a strumenti di Analisi Matematica troppo sofisticati. La teoria fin qui sviluppata non ci permette, però, di affrontare due argomenti fondamentali, sia dal punto di vista teorico che applicativo. Anzitutto la definizione di variabili casuali l'insieme dei cui valori sia non numerabile. Molte grandezze che trattiamo quotidianamente (tempi, masse, lunghezze,...) possono assumere qualunque valore di un'intervallo di \mathbb{R} . È ovviamente impossibile, in uno spazio di probabilità discreto, definire una variabile casuale X per cui $X(\Omega)$ sia un intervallo di \mathbb{R} , dato che $X(\Omega)$ è sempre al più numerabile. L'altra questione riguarda lo studio delle *successioni* di variabili casuali. Le prime applicazioni “moderne” del calcolo delle probabilità (de Moivre, Laplace), riguardarono il calcolo “approssimato” di certe probabilità. Una formulazione rigorosa di tali approssimazioni conduce a diverse nozioni di *convergenza* di successioni di variabili casuali. Uno spazio di probabilità discreto è troppo “povero” perchè in esso si possano definire successioni interessanti di variabili casuali, ad esempio successioni di variabili casuali indipendenti con la stessa distribuzione.

Risulta quindi naturale cercare di definire spazi di probabilità in cui lo spazio campionario sia non numerabile. Per avere un'idea del tipo di problemi che si affrontano, vediamo due esempi significativi.

Esempio 4.1. Il concetto di probabilità uniforme è chiaro e naturale se lo spazio campionario è finito. È possibile estendere tale nozione ad uno spazio campionario continuo, ad esempio un intervallo limitato di \mathbb{R} ? In altre parole, dato un intervallo I di \mathbb{R} , si può formalizzare l'idea di “scegliere a caso” un punto di I ?

Per fissare le idee, sia $I = \Omega = [0, 1]$. Se P è la probabilità “uniforme” che stiamo cercando di definire, è naturale assumere che, se $0 \leq a \leq b \leq 1$ allora

$$P([a, b]) = b - a. \quad (4.1)$$

In tal modo, oltre a verificarsi il fatto che $P(\Omega) = 1$, si ha che la probabilità di un intervallo dipende solo dalla sua lunghezza geometrica, e non dalla sua “posizione” in $[0, 1]$. Si noti che, da (4.1), segue che $P(\{x\}) = P([x, x]) = 0$ per ogni $x \in [0, 1]$. Il problema è: è possibile estendere la P definita in (4.1) a tutti i sottoinsiemi di $[0, 1]$, in modo che l'estensione verifichi la proprietà di σ -additività? Ricordiamo che quasi tutti i risultati concernenti gli spazi di probabilità discreti sono basati sulla σ -additività.

Esempio 4.2. Nell'esempio 1.37 abbiamo costruito un modello probabilistico per N prove ripetute indipendenti per le quali $p \in (0, 1)$ è la probabilità di successo. Pensiamo ora di effettuare una successione infinita di prove ripetute, cioè $N = +\infty$. La scelta naturale per lo spazio campionario è allora

$$\Omega = \{0, 1\}^{\mathbb{N} \setminus \{0\}},$$

cioè, se $\omega \in \Omega$, allora $\omega = (\omega_1, \omega_2, \dots)$ dove $\omega_i \in \{0, 1\}$. È ben noto che Ω non è numerabile. La probabilità P che vogliamo costruire sui sottoinsiemi di Ω dovrà soddisfare un requisito del tutto naturale: se consideriamo un evento che dipende solo dagli esiti delle prime N prove, con $N < +\infty$, allora la sua probabilità dovrà essere uguale a quella calcolata, nell'Esempio 1.37, con N fissato. In altre parole, se $x_1, x_2, \dots, x_N \in \{0, 1\}$, dovrà essere

$$P(\{\omega \in \Omega : \omega_1 = x_1, \dots, \omega_N = x_N\}) = p^{\sum_{i=1}^N x_i} (1-p)^{N-\sum_{i=1}^N x_i}. \quad (4.2)$$

Come nell'esempio precedente, il problema è di stabilire se è possibile estendere P a tutti i sottoinsiemi di Ω . Si noti che, se tale estensione (σ -additiva) esiste, allora $P(\{\eta\}) = 0$ per ogni $\eta \in \Omega$. Infatti

$$\{\eta\} = \bigcap_{N \in \mathbb{N} \setminus \{0\}} \{\omega : \omega_1 = \eta_1, \dots, \omega_N = \eta_N\}.$$

Quest'ultima è l'intersezione di una famiglia decrescente di eventi. Poiché abbiamo assunto che P sia σ -additiva e osservando che la Proposizione 1.10 non utilizza la numerabilità dello spazio campionario, abbiamo

$$P(\{\eta\}) = \lim_{N \rightarrow +\infty} p^{\sum_{i=1}^N \eta_i} (1-p)^{N-\sum_{i=1}^N \eta_i}.$$

Osserviamo ora che se $(\sum_{i=1}^N \eta_i) \geq N/2$ si ha $p^{\sum_{i=1}^N \eta_i} \leq p^{N/2}$, mentre se $(\sum_{i=1}^N \eta_i) < N/2$ si ha $(1-p)^{N-\sum_{i=1}^N \eta_i} \leq (1-p)^{N/2}$; in ogni caso

$$p^{\sum_{i=1}^N \eta_i} (1-p)^{N-\sum_{i=1}^N \eta_i} \leq (\max\{p, (1-p)\})^{N/2} \xrightarrow{N \rightarrow \infty} 0,$$

poiché $p \in (0, 1)$.

In entrambi gli esempi appena visti, la funzione $P(\cdot)$ viene definita dapprima in una famiglia di insiemi "semplici". Si può dimostrare, in entrambi i casi, che P non si può estendere a tutto $\mathcal{P}(\Omega)$ in modo che la P estesa risulti σ -additiva. Dobbiamo dunque ridimensionare l'obbiettivo iniziale di estendere P a tutto $\mathcal{P}(\Omega)$. Questo conduce alla seguente definizione.

Definizione 4.3. Sia Ω un insieme, e $\mathcal{A} \subseteq \mathcal{P}(\Omega)$. Diciamo che \mathcal{A} è una σ -algebra se

- (i) $\emptyset \in \mathcal{A}$.
- (ii) Se $A \in \mathcal{A}$ allora $A^c \in \mathcal{A}$.
- (iii) Se $(A_n)_{n \geq 0}$ è una successione di elementi di \mathcal{A} , allora $\bigcup_n A_n \in \mathcal{A}$.

Si noti che, per (i) e (ii), $\Omega \in \mathcal{A}$. Inoltre \mathcal{A} è chiuso per unione finita (ogni famiglia finita $\{A_1, \dots, A_n\}$ di elementi di \mathcal{A} può essere completata in una successione ponendo $A_k = \emptyset$ per $k > n$, senza modificarne l'unione). Infine usando l'identità $\bigcap_n A_n = (\bigcup_n A_n^c)^c$, si vede che una σ algebra è chiusa per intersezione, sia di una famiglia finita sia di una successione.

Nel seguito, una coppia (Ω, \mathcal{A}) formata da un insieme e da una σ -algebra di suoi sottoinsiemi verrà chiamata *spazio misurabile*.

Definizione 4.4. Sia (Ω, \mathcal{A}) uno spazio misurabile. Una funzione

$$P : \mathcal{A} \rightarrow [0, 1]$$

si dice *probabilità* (o *misura di probabilità*) se valgono le seguenti proprietà:

(P1)

$$P(\Omega) = 1.$$

(P2) (σ -additività) Per ogni successione $(A_n)_{n \in \mathbb{N}}$ di elementi di \mathcal{A} a due a due disgiunti, si ha

$$P\left(\bigcup_{n=0}^{+\infty} A_n\right) = \sum_{n=0}^{+\infty} P(A_n).$$

In altre parole, la nozione di probabilità su un insieme generale Ω è definita dallo stesso sistema di assiomi della probabilità su spazi discreti, con la differenza che la probabilità è definita su una σ algebra di sottoinsiemi di Ω che non necessariamente coincide con $\mathcal{P}(\Omega)$. La terna (Ω, \mathcal{A}, P) , dove (Ω, \mathcal{A}) è uno spazio misurabile e P è una probabilità su (Ω, \mathcal{A}) , verrà chiamata *spazio di probabilità*. In analogia a quanto fatto nel caso discreto, Ω sarà chiamato *spazio campionario* e gli elementi di \mathcal{A} saranno chiamati *eventi*. Esattamente come nel caso discreto, si mostra che in uno spazio di probabilità vale l'additività finita: per ogni $A_1, A_2, \dots, A_N \in \mathcal{A}$ a due a due disgiunti,

$$P\left(\bigcup_{n=0}^N A_n\right) = \sum_{n=0}^N P(A_n). \quad (4.3)$$

I risultati contenuti nel paragrafo 1.2 continuano a valere nel contesto più generale ora introdotto, a patto di restringere i risultati alla σ -algebra degli eventi. Per completezza, rinunciando ai risultati principali, le dimostrazioni dei quali sono identiche a quelle date per spazi di probabilità discreti.

Proposizione 4.5. Sia (Ω, \mathcal{A}, P) uno spazio di probabilità e siano $A, B \in \mathcal{A}$. Allora valgono le seguenti proprietà:

(i)

$$P(A^c) = 1 - P(A).$$

(ii) Se $A \subseteq B$ allora

$$P(B \setminus A) = P(B) - P(A).$$

In particolare

$$P(A) \leq P(B).$$

(iii)

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

In particolare

$$P(A \cup B) \leq P(A) + P(B).$$

Proposizione 4.6. Sia (Ω, \mathcal{A}) uno spazio misurabile, e $P : \mathcal{A} \rightarrow [0, 1]$ una funzione che soddisfa (P1) e l'additività in (4.3). Allora le seguenti proprietà sono equivalenti:

(a) P è σ -additiva.

(b) Per ogni successione crescente di eventi $(A_n)_{n \geq 1}$ (tale cioè che $A_n \subseteq A_{n+1}$ per ogni $n \geq 1$) si ha

$$P\left(\bigcup_{n \geq 1} A_n\right) = \lim_{n \rightarrow +\infty} P(A_n).$$

(c) Per ogni successione decrescente di eventi $(A_n)_{n \geq 1}$ (tale cioè che $A_n \supseteq A_{n+1}$ per ogni $n \geq 1$) si ha

$$P\left(\bigcap_{n \geq 1} A_n\right) = \lim_{n \rightarrow +\infty} P(A_n).$$

Torniamo ora al problema enunciato negli esempi 4.1 e 4.2. In entrambi i casi avevamo definito una funzione $P : \mathcal{I} \rightarrow [0, 1]$, dove $\mathcal{I} =$ famiglia degli intervalli in $[0, 1]$ nell'Esempio 4.1, e $\mathcal{I} =$ famiglia degli insiemi del tipo $\{\omega : \omega_1 = x_1, \dots, \omega_N = x_N\}$ nell'Esempio 4.2. È facile vedere che in entrambi i casi \mathcal{I} non è una σ -algebra. Il problema è allora di trovare una σ -algebra \mathcal{A} contenente \mathcal{I} , e una probabilità su (Ω, \mathcal{A}) che estenda la P originaria.

La scelta della σ -algebra \mathcal{A} si può fare in modo “canonico”, grazie al seguente risultato.

Proposizione 4.7. *Sia Ω un insieme arbitrario.*

- (i) Se $\{\mathcal{A}_\alpha : \alpha \in I\}$ è una famiglia di σ -algebre di sottoinsiemi di Ω indicizzata da un insieme arbitrario I , allora $\bigcap_{\alpha \in I} \mathcal{A}_\alpha$ è una σ -algebra di sottoinsiemi di Ω .
- (ii) Sia $\mathcal{I} \subseteq \mathcal{P}(\Omega)$. Allora esiste una minima σ -algebra \mathcal{A} contenente \mathcal{I} , ossia $\mathcal{I} \subseteq \mathcal{A}$, e se \mathcal{A}' è una σ -algebra contenente \mathcal{I} allora $\mathcal{A} \subseteq \mathcal{A}'$. Tale σ -algebra è denotata con $\sigma(\mathcal{I})$, e chiamata la σ -algebra generata da \mathcal{I} .

Dimostrazione. (i) La dimostrazione è semplice, ed è lasciata al lettore come esercizio.

(ii) Sia

$$\Xi = \{\mathcal{A} \subseteq \mathcal{P}(\Omega) : \mathcal{I} \subseteq \mathcal{A}, \mathcal{A} \text{ è una } \sigma\text{-algebra}\}.$$

Notare che $\Xi \neq \emptyset$, essendo $\mathcal{P}(\Omega) \in \Xi$. Per (i),

$$\mathcal{A} = \bigcap_{\mathcal{A}' \in \Xi} \mathcal{A}'$$

è una σ -algebra contenente \mathcal{I} , e, per definizione, $\mathcal{A} \subseteq \mathcal{A}'$ per ogni σ -algebra \mathcal{A}' contenente \mathcal{I} . □

Tornando agli Esempi 4.1 e 4.2, si vuole mostrare che esiste una probabilità che estende P definita almeno in $\sigma(\mathcal{I})$. Tale problema di estensione è altamente non banale, e al di là degli scopi di questo corso. È possibile dimostrare che, per entrambi gli esempi, esiste effettivamente un'unica probabilità che estende P a $\sigma(\mathcal{I})$. Inoltre, è possibile estendere P ad una σ -algebra più grande di $\sigma(\mathcal{I})$, ma non a tutto $\mathcal{P}(\Omega)$.

Osservazione 4.8. Se (Ω, P) è uno spazio di probabilità discreto, allora si può identificare con lo spazio di probabilità $(\Omega, \mathcal{P}(\Omega), P)$.

4.2 Variabili casuali

Definizione 4.9. Sia (Ω, \mathcal{A}, P) uno spazio di probabilità, e sia (E, \mathcal{E}) uno spazio misurabile. Una funzione

$$X : \Omega \rightarrow E$$

si dice *variabile casuale* a valori in (E, \mathcal{E}) se per ogni $C \in \mathcal{E}$ si ha

$$X^{-1}(C) \in \mathcal{A}.$$

Dunque, se X è una variabile casuale e $C \in \mathcal{E}$, allora $P(X \in C) := P(X^{-1}(C))$ è ben definita.

Definizione 4.10. Se X è una variabile casuale a valori in (E, \mathcal{E}) , la mappa

$$\begin{aligned} \mu_X : \mathcal{E} &\rightarrow [0, 1] \\ C &\mapsto P(X \in C) \end{aligned}$$

si dice *distribuzione* della variabile casuale X .

La dimostrazione del seguente risultato è lasciata per esercizio al lettore.

Proposizione 4.11. *La distribuzione μ_X di una variabile casuale X a valori in (E, \mathcal{E}) è una probabilità su (E, \mathcal{E}) .*

La nozione di variabile casuale dipende dalla scelta della σ -algebra sull'insieme E . Nel caso in cui $E = \mathbb{R}, \mathbb{R}^d$ o \mathbb{C} sceglieremo come σ -algebra la cosiddetta σ -algebra di Borel $\mathcal{B}(\mathbb{R}), \mathcal{B}(\mathbb{R}^d), \mathcal{B}(\mathbb{C})$, definita come la σ -algebra generata dai sottoinsiemi aperti. Tale scelta verrà sempre sottintesa nel seguito. Come nel caso discreto, chiameremo *scalari, vettoriali e complesse* le variabili casuali a valori, rispettivamente, in $\mathbb{R}, \mathbb{R}^d, \mathbb{C}$.

In completa analogia col caso discreto, diamo la seguente definizione.

Definizione 4.12. Siano X_1, X_2, \dots, X_n variabili casuali definite sullo stesso spazio di probabilità (Ω, \mathcal{A}, P) a valori rispettivamente negli spazi misurabili $(E_1, \mathcal{E}_1), (E_2, \mathcal{E}_2), \dots, (E_n, \mathcal{E}_n)$. Esse si dicono indipendenti se per ogni scelta di $A_1 \in \mathcal{E}_1, A_2 \in \mathcal{E}_2, \dots, A_n \in \mathcal{E}_n$ si ha

$$P(X_1 \in A_1, X_2 \in A_2, \dots, X_n \in A_n) = \prod_{i=1}^n P(X_i \in A_i). \quad (4.4)$$

Infine, per una variabile casuale scalare X , definiamo la funzione di ripartizione

$$F_X(x) = P(X \leq x).$$

Le proprietà della funzione di ripartizione contenute nella Proposizione 3.57 continuano a valere nel caso generale, come si vede dal fatto che la dimostrazione della Proposizione 3.57 è corretta anche in spazi di probabilità generali. Inoltre, la stessa dimostrazione usata nella Proposizione 3.58 mostra che

$$F_X(x) - F_X(x^-) = P(X = x).$$

Si noti come l'identità precedente, nel caso di variabili casuali discrete, permetta di mostrare che se due variabili casuali hanno la stessa funzione di ripartizione, allora hanno la stessa distribuzione. Tale affermazione è vera anche nel caso generale, ma omettiamo la dimostrazione di questo fatto.

Proposizione 4.13. *Siano X e Y due variabili casuali a valori in \mathbb{R} , tali che $F_X = F_Y$. Allora $\mu_X = \mu_Y$.*

4.3 Valor medio

La nozione generale di valor medio esula dagli scopi di questo corso. Pertanto, in questo paragrafo, verranno date le definizioni e i risultati principali, senza dettagli e dimostrazioni.

Per garantire la coerenza di quanto vedremo ora con quanto visto per spazi di probabilità discreti, introduciamo ora la nozione di variabile casuale discreta nel contesto più generale.

Definizione 4.14. Sia (Ω, \mathcal{A}, P) uno spazio di probabilità, (E, \mathcal{E}) uno spazio misurabile tale che, per ogni $x \in E$ si abbia $\{x\} \in \mathcal{E}$, e sia X una variabile casuale a valori in (E, \mathcal{E}) . Diremo che X è una *variabile casuale discreta* se $X(\Omega)$ è finito o numerabile.

Per una variabile casuale discreta è possibile definire la densità

$$p_X(x) = P(X = x) = P(X^{-1}(\{x\})).$$

Se $E = \mathbb{R}$, non ha però alcun senso definire il valor medio di X tramite la formula $E(X) = \sum_{\omega \in \Omega} X(\omega)P(\{\omega\})$: negli esempi 4.1 e 4.2 abbiamo infatti visto che $P(\{\omega\}) = 0$ per ogni $\omega \in \Omega$, e quindi ogni valor medio sarebbe nullo. Tuttavia, avendo a disposizione la densità, è naturale definire $E(X)$ tramite la relazione

$$E(X) := \sum_{x \in X(\omega)} x p_X(x), \quad (4.5)$$

se la serie in (4.5) converge assolutamente; in caso contrario diremo che la variabile casuale X non ammette valor medio.

Vediamo ora come si estenda la nozione di valor medio a variabili casuali generali. Nel seguito, se X è una variabile casuale scalare, definiamo $X^+ = \max(X, 0)$ e $X^- = -\min(X, 0)$. Si noti che $X^+, X^- \geq 0$ e $X = X^+ - X^-$.

Definizione 4.15. Sia X una variabile casuale scalare. Se $X \geq 0$, definiamo

$$E(X) = \sup\{E(Y) : 0 \leq Y \leq X, Y \text{ è una variabile casuale discreta}\} \in [0, +\infty].$$

(in particolare, $E(X)$ è dato da (4.5) se X è discreta). Per una X generale, diremo che X ammette valor medio se $E(X^+) < +\infty$ e $E(X^-) < +\infty$, e in tal caso

$$E(X) = E(X^+) - E(X^-).$$

Infine, se X è una variabile casuale complessa, posto $X = \operatorname{Re}(X) + i\operatorname{Im}(X)$, diciamo che X ammette valor medio se $\operatorname{Re}(X)$ e $\operatorname{Im}(X)$ ammettono entrambe valor medio, e in questo caso poniamo

$$E(X) = E(\operatorname{Re}(X)) + iE(\operatorname{Im}(X)).$$

Contrariamente al caso discreto, le dimostrazioni di alcune delle proprietà fondamentali del valor medio sono non banali, e non verranno trattate in questo corso.

Proposizione 4.16. *Siano X, Y due variabili casuali scalari o complesse, definite nello stesso spazio di probabilità (Ω, \mathcal{A}, P) . Allora valgono le seguenti proprietà:*

- (i) (*Monotonia*) *Se X, Y sono a valori reali, entrambe ammettono valor medio e $X(\omega) \leq Y(\omega)$, allora $E(X) \leq E(Y)$.*

(ii) X ammette valor medio se e solo se $|X|$ ammette valor medio, e in tal caso

$$|E(X)| \leq E(|X|).$$

(iii) (*Linearità*) Se X e Y ammettono valor medio e $a, b \in \mathbb{C}$, allora la variabile casuale $aX + bY$ definita da

$$(aX + bY)(\omega) = aX(\omega) + bY(\omega),$$

ammette valor medio e

$$E(aX + bY) = aE(X) + bE(Y).$$

Sia ora $g : \mathbb{R} \rightarrow \mathbb{R}$, e X una variabile casuale a valori in \mathbb{R} . In generale non è detto che $g(X) := g \circ X$ sia una variabile casuale. Per averne la garanzia, occorre assumere che g sia una *funzione misurabile*, cioè

- per ogni $A \in \mathcal{B}(\mathbb{R})$ si ha che $g^{-1}(A) \in \mathcal{B}(\mathbb{R})$.

La misurabilità è una proprietà piuttosto debole; ad esempio tutte le funzioni continue a tratti sono misurabili. Si rimanda ad un testo più avanzato per una discussione più approfondita sull'argomento.

Se g è misurabile, ha senso chiedersi se $g(X)$ ammette valor medio. È possibile dimostrare che tanto il fatto che $g(X)$ ammetta valor medio, quanto eventualmente il valore di tale valor medio, dipendono solo dalla distribuzione di X . In altre parole, *se X e Y hanno la stessa distribuzione e g è misurabile, allora $g(X)$ ammette valor medio se e solo se $g(Y)$ ammette valor medio, e in caso affermativo i due valori medi sono uguali.*

Poichè i polinomi sono funzioni continue, e perciò misurabili, analogamente al caso discreto, si possono definire Varianza, Covarianza e momenti e funzione generatrice dei momenti di una variabile casuale. I risultati contenuti nei paragrafi 3.5 e 3.8 sono validi in questo contesto più generale.

Capitolo 5

Variabili casuali assolutamente continue

5.1 Richiami sull'integrale di Riemann

Cominciamo con l'introdurre alcune notazioni, e alcune precisazioni tecniche. In quanto segue daremo per nota la nozione di *integrale di Riemann*

$$\int_a^b f(x) dx,$$

dove $f : [a, b] \rightarrow \mathbb{R}$. Vale tuttavia la pena di ricordare alcuni fatti fondamentali.

- (1) L'integrale "proprio" di Riemann viene definito per una classe di funzioni limitate, dette *Riemann-integrabili* (talvolta diremo, semplicemente, integrabili). Tutte le funzioni $f : [a, b] \rightarrow \mathbb{R}$ limitate tali che per ogni $y \in [a, b]$ i limiti destro e sinistro

$$\lim_{x \downarrow y} f(x), \quad \lim_{x \uparrow y} f(x),$$

esistono finiti (solo il primo se $y = a$ e solo il secondo se $y = b$), sono Riemann-integrabili. In particolare sono integrabili le funzioni continue e quelle monotone.

- (2) Sia $f : [a, b] \rightarrow \mathbb{R}$ una funzione *continua a tratti*: supponiamo cioè che esista un sottoinsieme finito $N \subseteq [a, b]$ tale che f è continua in ogni punto $x \in [a, b] \setminus N$. Se inoltre i limiti destro e sinistro di f esistono finiti nei punti di N , allora f è Riemann-integrabile e, definita la funzione integrale $F : [a, b] \rightarrow \mathbb{R}$ mediante

$$F(x) := \int_a^x f(t) dt,$$

si ha che F è continua su $[a, b]$ e derivabile in ogni punto di $[a, b] \setminus N$, con

$$F'(x) = f(x), \quad \forall x \in [a, b] \setminus N.$$

Viceversa, sia $F : [a, b] \rightarrow \mathbb{R}$ un funzione \mathcal{C}^1 *a tratti*: supponiamo cioè che F sia continua su $[a, b]$ e che esista un sottoinsieme finito $N \subseteq [a, b]$ tale che F sia derivabile in ogni

punto di $[a, b] \setminus N$, e che F' sia continua in ogni punto di $[a, b] \setminus N$. Assumiamo inoltre che per ogni $y \in N$ esistano finiti i limiti

$$\lim_{x \downarrow y} F'(x), \quad \lim_{x \uparrow y} F'(x).$$

Allora se definiamo

$$f(x) := \begin{cases} F'(x) & \text{se } x \in [a, b] \setminus N \\ \text{un valore arbitrario} & \text{altrimenti,} \end{cases}$$

si ha che f è Riemann-integrabile e vale inoltre la relazione

$$F(x) = \int_a^x f(t)dt, \quad \forall x \in [a, b].$$

I risultati appena enunciati costituiscono il *teorema fondamentale del calcolo integrale*.

- (3) Si possono definire varie forme “generalizzate” di integrale di Riemann. Ad esempio se $f : (a, b] \rightarrow \mathbb{R}$ è Riemann integrabile su ogni intervallo $[a + \epsilon, b]$ con $\epsilon \in (0, b - a)$, e il limite

$$\lim_{\epsilon \downarrow 0} \int_{a+\epsilon}^b f(x)dx$$

esiste finito, diciamo che f è *integrabile in senso generalizzato* su $[a, b]$, e il suo integrale

$$\int_a^b f(x)dx$$

è dato dal limite precedente. In modo analogo definiamo l’integrale generalizzato di $f : [a, b) \rightarrow \mathbb{R}$. Se f è integrabile in senso generalizzato su $[a, c]$ e su $[c, b]$ diremo che è integrabile in senso generalizzato su $[a, b]$ e poniamo

$$\int_a^b f := \int_a^c f + \int_c^b f.$$

In altre parole, con questa definizione otteniamo, quando possibile, l’integrale di una funzione $f : [a, b] \setminus \{c\} \rightarrow \mathbb{R}$ che, ad esempio, non ammetta limite finito in c . Questo procedimento si estende in modo evidente a funzioni $f : [a, b] \setminus N \rightarrow \mathbb{R}$, dove N è un insieme finito.

Sia ora $f : [a, +\infty) \rightarrow \mathbb{R}$ una funzione che assumiamo Riemann-integrabile (anche nel senso generalizzato appena visto) su ogni intervallo $[a, c]$, e per cui il limite

$$\lim_{c \rightarrow +\infty} \int_a^c f(x)dx$$

esiste finito. Allora diciamo che f è integrabile in senso generalizzato su $[a, +\infty)$ e poniamo

$$\int_a^{+\infty} f(x)dx := \lim_{c \rightarrow +\infty} \int_a^c f(x)dx.$$

In modo analogo definiamo l’integrabilità su semirette del tipo $(-\infty, b]$. Infine, se $f : \mathbb{R} \rightarrow \mathbb{R}$ è integrabile su $(-\infty, a]$ e su $[a, +\infty)$ allora diciamo che f è integrabile in senso generalizzato su \mathbb{R} e poniamo

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

- (4) Non è difficile fornire generalizzazioni del teorema fondamentale del calcolo integrale a funzioni integrabili nel senso generalizzato appena descritto. Ad esempio, se $f : \mathbb{R} \rightarrow \mathbb{R}$ è *continua a tratti* ed è integrabile in senso generalizzato su \mathbb{R} , allora, posto

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

si ha che $F'(x) = f(x)$ per ogni x in cui f è continua. Viceversa, se $F : \mathbb{R} \rightarrow \mathbb{R}$ è una funzione \mathcal{C}^1 a tratti, allora la funzione F' (definita in modo arbitrario nei punti in cui F non è derivabile) è integrabile su \mathbb{R} in senso generalizzato e, per ogni $x \in \mathbb{R}$, si ha

$$F(x) = \int_{-\infty}^x F'(t)dt.$$

5.2 Variabili casuali scalari assolutamente continue

Definizione 5.1. Sia (Ω, \mathcal{A}, P) uno spazio di probabilità, e $X : \Omega \rightarrow \mathbb{R}$ una variabile casuale scalare. Diciamo che X è assolutamente continua se esiste $f_X : \mathbb{R} \in [0, +\infty)$ integrabile su \mathbb{R} in senso generalizzato tale che, se $F_X(x) := P(X \leq x)$ è la funzione di ripartizione di X , si ha

$$F_X(x) = \int_{-\infty}^x f_X(t)dt. \quad (5.1)$$

Una tale funzione f_X viene detta *densità* di X .

Osservazione 5.2. Prendendo il limite $x \rightarrow \infty$ nella relazione (5.1) e usando la continuità dal basso della probabilità, si ottiene che la densità f_X soddisfa la seguente relazione:

$$\int_{-\infty}^{+\infty} f_X(x) dx = 1. \quad (5.2)$$

Osservazione 5.3. Se X è una variabile casuale assolutamente continua, la definizione 5.1 non identifica la sua densità f_X in modo unico. Infatti, se f_X è una densità di X , ogni funzione g per cui

$$\int_{-\infty}^x g(t)dt = \int_{-\infty}^x f_X(t)dt \quad (5.3)$$

per ogni $x \in \mathbb{R}$, è una densità di X . Ad esempio, se g è ottenuta da f_X modificandone il valore in un numero finito di punti, allora (5.3) vale. Questa “ambiguità” nella nozione di densità di una variabile casuale assolutamente continua non porterà tuttavia alcun problema. Spesso diremo, impropriamente, che una certa funzione f è la densità di X .

Dalla definizione 5.1 seguono facilmente alcune proprietà delle variabili casuali assolutamente continue, che raccogliamo nella seguente

Osservazione 5.4. (1) Come abbiamo visto, la relazione

$$P(X = x) = F_X(x) - F_X(x^-)$$

vale per tutte le variabili casuali a valori in \mathbb{R} . Se X è assolutamente continua, la sua funzione di ripartizione

$$F_X(x) = \int_{-\infty}^x f_X(t)dt$$

è, per il teorema fondamentale del calcolo integrale, una funzione continua, e pertanto $F_X(x) = F_X(x^-)$ per ogni $x \in \mathbb{R}$. Quindi, se X è assolutamente continua,

$$P(X = x) = 0$$

per ogni $x \in \mathbb{R}$.

(2) Per ogni $a, b \in [-\infty, +\infty]$, $a \leq b$, si ha

$$P(X \in (a, b]) = P(X \leq b) - P(X \leq a) = \int_a^b f_X(t) dt.$$

Si noti che

$$P(X \in [a, b]) = P(X \in (a, b]) + P(X = a) = P(X \in (a, b]),$$

per quanto abbiamo appena visto. In modo analogo, si mostra che

$$P(X \in [a, b]) = P(X \in (a, b]) = P(X \in [a, b)) = P(X \in (a, b)) = \int_a^b f_X(t) dt.$$

(3) Sia X una variabile casuale a valori in \mathbb{R} tale che la funzione di ripartizione F_X sia \mathcal{C}^1 a tratti. Dal teorema fondamentale del calcolo integrale, richiamato in precedenza, segue che X è una variabile casuale assolutamente continua, e ogni funzione f tale che $f(x) = F'_X(x)$ per ogni x in cui F'_X è continua è una densità di X .

Enunciamo ora il seguente risultato fondamentale, omettendo la dimostrazione, che richiede tecniche più avanzate.

Teorema 5.5. *Sia X una variabile casuale assolutamente continua con densità f_X . Sia inoltre $g : \mathbb{R} \rightarrow \mathbb{R}$ una funzione misurabile tale che $g(x)f_X(x)$ e $|g(x)|f_X(x)$ siano integrabili in senso generalizzato su \mathbb{R} . Allora $g(X)$ ammette valor medio e*

$$E[g(X)] = \int_{-\infty}^{+\infty} g(x)f_X(x) dx.$$

Enunciamo ora senza dimostrazione un altro risultato, che rappresenta l'analogo della formula (3.44), dimostrata per le variabili casuali discrete.

Teorema 5.6. *In uno stesso spazio di probabilità (Ω, \mathcal{A}, P) , siano definite X e Y due variabili casuali assolutamente continue, indipendenti, con densità rispettivamente f_X e f_Y . Allora, per ogni $z \in \mathbb{R}$, le funzioni $x \mapsto f_X(x)f_Y(z-x)$ e $x \mapsto f_X(z-x)f_Y(x)$ sono integrabili in senso generalizzato su \mathbb{R} . Inoltre la variabile casuale $X+Y$ è assolutamente continua, e una sua densità è data da*

$$f_{X+Y}(z) = \int_{-\infty}^{+\infty} f_X(x)f_Y(z-x) dx = \int_{-\infty}^{+\infty} f_X(z-x)f_Y(x) dx.$$

5.3 Classi notevoli di variabili casuali assolutamente continue

5.3.1 Variabili casuali uniformi

Siano $a, b \in \mathbb{R}$ con $a < b$. Una variabile casuale assolutamente continua X a valori in \mathbb{R} si dice *uniforme* su $[a, b]$, e scriviamo

$$X \sim U(a, b),$$

se

$$f_X(x) = \frac{1}{b-a} 1_{[a,b]}(x).$$

È naturale interpretare una variabile casuale $X \sim U(a, b)$ come “un punto scelto a caso su $[a, b]$ ”. Le variabili uniformi possono essere “simulate” al calcolatore, da speciali programmi chiamati *generatori di numeri casuali*. Essi sono alla base di numerosi algoritmi, detti *algoritmi stocastici*.

Calcoliamo ora media e varianza di $X \sim U(a, b)$.

$$E(X) = \frac{1}{b-a} \int_a^b x dx = \frac{a+b}{2}.$$

$$E(X^2) = \frac{1}{b-a} \int_a^b x^2 dx = \frac{a^2 + ab + b^2}{3},$$

da cui

$$\text{Var}(X) = E(X^2) - E^2(X) = \frac{(b-a)^2}{12}.$$

La funzione generatrice dei momenti vale

$$\gamma_X(t) = E(e^{tX}) = \frac{1}{b-a} \int_a^b e^{tx} dx = \frac{e^{tb} - e^{ta}}{t(b-a)}.$$

Le variabili casuali uniformi possono essere usate come ingredienti per “costruire” variabili casuali a valori in \mathbb{R} con distribuzione arbitraria. Infatti, come mostra il seguente risultato, data una qualunque variabile casuale X a valori in \mathbb{R} , esiste una funzione misurabile $f: [0, 1] \rightarrow \mathbb{R}$ tale che, se $Y \sim U(0, 1)$, X e $f(Y)$ hanno la stessa distribuzione.

Proposizione 5.7. *Sia X una variabile casuale a valori in \mathbb{R} , e sia F_X la sua funzione di ripartizione. Consideriamo la funzione $f: \mathbb{R} \rightarrow \mathbb{R}$ così definita*

$$f(x) := \begin{cases} 0 & \text{se } x \leq 0 \text{ oppure } x \geq 1 \\ \inf\{z : F_X(z) \geq x\} & \text{se } x \in (0, 1) \end{cases}.$$

Allora f è una funzione misurabile e, se $Y \sim U(0, 1)$, le variabili casuali X e $f(Y)$ hanno la stessa distribuzione.

Dimostrazione. Omettiamo la dimostrazione della misurabilità di f . Sia $Z := f(Y)$. Per mostrare che X e Z hanno la stessa distribuzione, è sufficiente mostrare che $F_X = F_Z$. Sia allora $x \in \mathbb{R}$ arbitrario, e mostriamo che

$$F_X(x) = F_Z(x), \tag{5.4}$$

cioè

$$P(X \leq x) = P(Z \leq x).$$

Per dimostrarlo, consideriamo $y \in (0, 1)$. Notare che

$$f(y) \leq x \iff \inf\{z : F_X(z) \geq y\} \leq x.$$

Poiché F_X è crescente

$$\inf\{z : F_X(z) \geq y\} \leq x \iff F_X(x) \geq y.$$

Poiché $P(Y \in (0, 1)) = 1$, abbiamo allora che

$$\begin{aligned} F_Z(x) &= P(f(Y) \leq x) = P(\{f(Y) \leq x\} \cap \{Y \in (0, 1)\}) = P(\{F_X(x) \geq Y\} \cap \{Y \in (0, 1)\}) \\ &= P(F_X(x) \geq Y) = F_X(x), \end{aligned}$$

dove abbiamo usato il semplice fatto che, se $Y \sim U(0, 1)$ e $c \in [0, 1]$, allora $P(Y \leq c) = c$. \square

5.3.2 Variabili casuali Gamma ed esponenziali

Cominciamo col ricordare la definizione della funzione Gamma di Eulero:

$$\Gamma(\alpha) = \int_0^{+\infty} x^{\alpha-1} e^{-x} dx,$$

ben definita per $\alpha > 0$ (perché?). Il valore di $\Gamma(\alpha)$ è noto esplicitamente per vari valori di α . Infatti, osservando che

$$\Gamma(1) = \int_0^{+\infty} e^{-x} dx = 1$$

e che, integrando per parti

$$\Gamma(\alpha + 1) = \int_0^{+\infty} x^\alpha e^{-x} dx = -x^\alpha e^{-x} \Big|_0^{+\infty} + \alpha \int_0^{+\infty} x^{\alpha-1} e^{-x} dx = \alpha \Gamma(\alpha),$$

si ha

$$\Gamma(n) = (n-1)!$$

per $n \in \mathbb{N} \setminus \{0\}$. Inoltre, se $\alpha = \frac{1}{2}$, usando il cambiamento di variabile $x = y^2$,

$$\Gamma(1/2) = \int_0^{+\infty} x^{-1/2} e^{-x} dx = 2 \int_0^{+\infty} e^{-y^2} dy = \int_{-\infty}^{+\infty} e^{-y^2} dy = \sqrt{\pi},$$

ove si è usato il noto valore dell'integrale di Gauss

$$\int_{-\infty}^{+\infty} e^{-y^2} dy = \sqrt{\pi}.$$

Dunque, usando ancora la relazione $\Gamma(\alpha + 1) = \alpha \Gamma(\alpha)$, si trova, per $n \in \mathbb{N} \setminus \{0\}$

$$\Gamma\left(n + \frac{1}{2}\right) = \sqrt{\pi} \prod_{k=0}^{n-1} \left(k + \frac{1}{2}\right).$$

Notiamo anche che se nella definizione di Γ operiamo il cambio di variabili $x = \lambda y$, con $\lambda > 0$, otteniamo

$$\Gamma(\alpha) = \int_0^{+\infty} \lambda^\alpha y^{\alpha-1} e^{-\lambda y} dy,$$

ossia

$$\int_0^{+\infty} \frac{1}{\Gamma(\alpha)} \lambda^\alpha x^{\alpha-1} e^{-\lambda x} dx = 1.$$

Ciò garantisce la bontà della seguente definizione. Una variabile casuale scalare assolutamente continua X è detta variabile casuale Gamma di parametri $\alpha, \lambda > 0$, se

$$f_X(x) = \frac{1}{\Gamma(\alpha)} \lambda^\alpha x^{\alpha-1} e^{-\lambda x} 1_{[0,+\infty)}(x).$$

In tal caso scriveremo $X \sim \Gamma(\alpha, \lambda)$. Si ha, se $X \sim \Gamma(\alpha, \lambda)$,

$$E(X) = \frac{\lambda^\alpha}{\Gamma(\alpha)} \int_0^{+\infty} x^\alpha e^{-\lambda x} dx = \frac{1}{\lambda \Gamma(\alpha)} \int_0^{+\infty} \lambda^{\alpha+1} x^\alpha e^{-\lambda x} dx = \frac{1}{\lambda \Gamma(\alpha)} \Gamma(\alpha+1) = \frac{\alpha}{\lambda},$$

ricordando che $\Gamma(\alpha+1) = \alpha \Gamma(\alpha)$. Analogamente

$$E(X^2) = \frac{\lambda^\alpha}{\Gamma(\alpha)} \int_0^{+\infty} x^{\alpha+1} e^{-\lambda x} dx = \frac{1}{\lambda^2 \Gamma(\alpha)} \int_0^{+\infty} \lambda^{\alpha+2} x^{\alpha+1} e^{-\lambda x} dx = \frac{1}{\lambda^2 \Gamma(\alpha)} \Gamma(\alpha+2) = \frac{\alpha(\alpha+1)}{\lambda^2},$$

da cui, facilmente,

$$\text{Var}(X) = \frac{\alpha}{\lambda^2}.$$

Tutti questi calcoli si sarebbero potuti fare facilmente calcolando la funzione generatrice dei momenti. Infatti

$$\gamma_X(t) = \int_{-\infty}^{+\infty} e^{tx} f_X(x) dx = \frac{\lambda^\alpha}{\Gamma(\alpha)} \int_0^{+\infty} x^{\alpha-1} e^{-(\lambda-t)x} dx$$

Quest'ultimo integrale è finito solo per $t < \lambda$, e in tal caso

$$\int_0^{+\infty} x^{\alpha-1} e^{-(\lambda-t)x} dx = \frac{\Gamma(\alpha)}{(\lambda-t)^\alpha},$$

e quindi

$$\gamma_X(t) = \begin{cases} \left(\frac{\lambda}{\lambda-t}\right)^\alpha & \text{per } t < \lambda \\ +\infty & \text{altrimenti.} \end{cases}$$

Il seguente risultato illustra una proprietà molto importante delle variabili casuali Gamma.

Proposizione 5.8. *Siano $X \sim \Gamma(\alpha, \lambda)$ e $Y \sim \Gamma(\beta, \lambda)$ indipendenti. Allora $X + Y \sim \Gamma(\alpha + \beta, \lambda)$.*

Dimostrazione. Usando il Teorema 5.6, per $z > 0$ abbiamo:

$$\begin{aligned} f_{X+Y}(z) &= \int f_X(z-x) f_Y(x) dx \\ &= \frac{\lambda^\alpha \lambda^\beta}{\Gamma(\alpha) \Gamma(\beta)} \int 1_{(0,+\infty)}(z-x) (z-x)^{\alpha-1} e^{-\lambda(z-x)} 1_{(0,+\infty)}(x) x^{\beta-1} e^{-\lambda x} dx \\ &= \frac{\lambda^\alpha \lambda^\beta}{\Gamma(\alpha) \Gamma(\beta)} e^{-\lambda z} \int_0^z (z-x)^{\alpha-1} x^{\beta-1} dx = \frac{\lambda^\alpha \lambda^\beta}{\Gamma(\alpha) \Gamma(\beta)} e^{-\lambda z} z^{\alpha+\beta-1} \int_0^1 (1-t)^{\alpha-1} t^{\beta-1} dt, \end{aligned}$$

dove nell'ultimo passaggio abbiamo usato il cambio di variabile $x = zt$. Ne segue che $f_{X+Y}(z)$ è proporzionale a $z^{\alpha+\beta-1}e^{-\lambda z}$, e quindi è necessariamente la densità di una $\Gamma(\alpha + \beta, \lambda)$. \square

Osservazione 5.9. Si noti che dalla dimostrazione precedente segue la non ovvia identità

$$\Gamma(\alpha + \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\int_0^1 (1-t)^{\alpha-1}t^{\beta-1}dt}.$$

Di grande interesse applicativo sono alcuni casi particolari di variabili casuali Gamma. Se $X \sim \Gamma(1, \lambda)$, diciamo che X è una variabile casuale esponenziale di parametro λ , e scriveremo

$$X \sim Exp(\lambda).$$

Si noti che in questo caso

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

Tali variabili casuali si possono interpretare come l'analogo continuo delle variabili casuali Geometriche. Esse, infatti, soddisfano ad una proprietà di perdita di memoria del tutto analoga a quella delle variabili casuali Geometriche.

Proposizione 5.10. *Sia $X \sim Exp(\lambda)$. Allora per ogni $s, t > 0$*

$$P(X \geq s + t | X > s) = P(X > t).$$

Dimostrazione. Osservando che

$$P(X \geq s + t | X > s) = \frac{P(X \geq s + t)}{P(X \geq s)},$$

e che

$$P(X \geq s) = \int_s^{+\infty} e^{-\lambda x} dx = e^{-\lambda s},$$

il risultato desiderato segue immediatamente. \square

Come per le variabili casuali Geometriche, si può dimostrare che le variabili casuali esponenziali sono le uniche variabili casuali assolutamente continue a valori reali positivi per cui vale la precedente proprietà di perdita di memoria. La dimostrazione di tale fatto è omessa. Le variabili casuali esponenziali sono normalmente usate come modelli per “tempi di attesa”: tempo di decadimento di atomi radioattivi, tempo intercorrente tra due terremoti successivi, tempo intercorrente tra l'arrivo di due clienti ad uno sportello, ...

Un'altra classe di variabili Gamma che hanno grande rilevanza soprattutto in statistica sono le variabili χ^2 . Se $n \geq 1$ è un numero naturale, e $X \sim \Gamma\left(\frac{n}{2}, \frac{1}{2}\right)$, diciamo che X è una variabile casuale χ^2 con n gradi di libertà, e scriviamo $X \sim \chi^2(n)$. Vedremo più avanti un esempio in cui appaiono variabili casuali di questo tipo.

5.3.3 Variabili casuali Normali o Gaussiane

Una variabile casuale scalare assolutamente continua X si dice Normale o Gaussiana *standard*, e si scrive $X \sim N(0, 1)$, se

$$f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}.$$

Notare che si tratta di una buona definizione dato che

$$\int_{-\infty}^{+\infty} e^{-\frac{x^2}{2}} dx = \sqrt{2\pi}.$$

È facile vedere che, se $X \sim N(0, 1)$

$$E(X) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} x e^{-\frac{x^2}{2}} dx = 0$$

dato che l'integrando è una funzione dispari (e Riemann-integrabile in senso generalizzato). Inoltre, integrando per parti,

$$E(X^2) = Var(X) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} x^2 e^{-\frac{x^2}{2}} dx = \frac{1}{\sqrt{2\pi}} \left[-x e^{-\frac{x^2}{2}} \Big|_{-\infty}^{+\infty} + \int_{-\infty}^{+\infty} e^{-\frac{x^2}{2}} dx \right] = 1.$$

Sia ora $Y \sim N(0, 1)$, $\mu \in \mathbb{R}$ e $\sigma > 0$, e definiamo

$$X = \sigma Y + \mu.$$

Dalle proprietà elementari di media e varianza, si vede che

$$E(X) = \mu, \quad Var(X) = \sigma^2.$$

Per determinare la densità di X procediamo come segue.

$$F_X(x) = P(X \leq x) = P(\sigma Y + \mu \geq x) = P\left(Y \leq \frac{x - \mu}{\sigma}\right) = F_Y\left(\frac{x - \mu}{\sigma}\right).$$

Essendo Y una variabile casuale assolutamente continua con densità continua, F_Y , e dunque F_X , è di classe \mathcal{C}^1 . Perciò

$$\begin{aligned} f_X(x) = F'_X(x) &= \frac{1}{\sigma} F'_Y\left(\frac{x - \mu}{\sigma}\right) = \frac{1}{\sigma} f_Y\left(\frac{x - \mu}{\sigma}\right) \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}. \end{aligned} \tag{5.5}$$

Una variabile casuale la cui densità è data da (5.5) si dice Normale o Gaussiana di media μ e varianza σ^2 , e si scrive

$$X \sim N(\mu, \sigma^2).$$

Passando per la funzione di ripartizione come nell'argomento precedente, si dimostra (esercizio!) che se $X \sim N(\mu, \sigma^2)$ e $a, b \in \mathbb{R}$, $a \neq 0$, allora

$$aX + b \sim N(a\mu + b, a^2\sigma^2).$$

In particolare

$$\frac{X - \mu}{\sigma} \sim N(0, 1).$$

Le variabili casuali Normali sono senz'altro le più usate nelle applicazioni. Detto in modo grossolano, questo è perchè ogni qual volta una quantità aleatoria è la somma di molte quantità aleatorie indipendenti, allora la sua distribuzione è approssimativamente Gaussiana. Una versione parziale, ma rigorosa, di tale affermazione, verrà data nel prossimo capitolo, con il teorema limite centrale.

Anche per le variabili casuali Gaussiane è facile calcolare la funzione generatrice dei momenti. Sia $X \sim N(0, 1)$.

$$\gamma_X(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{tx - \frac{1}{2}x^2} dx.$$

L'integrale precedente viene calcolato usando il metodo del *completamento dei quadrati*, molto utile per integrali gaussiani. Si tratta di osservare che

$$tx - \frac{1}{2}x^2 = -\frac{1}{2}(x - t)^2 + \frac{t^2}{2}.$$

Pertanto

$$\gamma_X(t) = e^{\frac{t^2}{2}} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{1}{2}(x-t)^2} dx.$$

Si osservi che $\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x-t)^2}$ è la densità di una variabile casuale Gaussiana di media t e varianza 1, e quindi il suo integrale è 1. Ne segue che

$$\gamma_X(t) = e^{\frac{t^2}{2}}.$$

Per calcolare la funzione generatrice di una generica Gaussiana si procede come segue. Sia $Y \sim N(\mu, \sigma^2)$. Da quanto sopra osservato, Y si può scrivere nella forma $Y = \sigma Z + \mu$, con $Z := (Y - \mu)/\sigma \sim N(0, 1)$. Ma allora

$$\gamma_Y(t) = E\left(e^{t(\sigma Z + \mu)}\right) = e^{t\mu} E\left(e^{t\sigma Z}\right) = e^{t\mu} \gamma_Z(t\sigma) = e^{t\mu} e^{\frac{t^2\sigma^2}{2}}.$$

La proprietà enunciata nella seguente proposizione è di estrema importanza.

Proposizione 5.11. *Siano $X \sim N(\mu_1, \sigma_1^2)$ e $Y \sim N(\mu_2, \sigma_2^2)$ indipendenti. Allora*

$$X + Y \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2).$$

Dimostrazione. Iniziamo col supporre $\mu_1 = \mu_2 = 0$. Dal Teorema 5.6, si ha

$$\begin{aligned} f_{X+Y}(z) &= \int f_X(z-x)f_Y(x)dx \\ &= \frac{1}{2\pi\sigma_1\sigma_2} \int \exp\left[-\frac{1}{2\sigma_1^2}(z-x)^2\right] \exp\left[-\frac{1}{2\sigma_2^2}x^2\right] dx \\ &= \frac{1}{2\pi\sigma_1\sigma_2} \exp\left[\frac{1}{2(\sigma_1^2 + \sigma_2^2)}z^2\right] \int \exp\left[-\frac{1}{2}\left(\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}\right)(x-\xi)^2\right] dx, \end{aligned}$$

dove $\xi = \frac{\sigma_2}{\sigma_1\sqrt{\sigma_1^2+\sigma_2^2}}z$. Utilizzando il fatto che

$$\int \exp \left[-\frac{1}{2} \left(\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2} \right) (x - \xi)^2 \right] dx = \sqrt{2\pi \left(\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2} \right)^{-1}},$$

con un po' di calcoli si trova

$$f_{X+Y}(z) = \frac{1}{\sqrt{2\pi(\sigma_1^2 + \sigma_2^2)}} \exp \left[-\frac{1}{2(\sigma_1^2 + \sigma_2^2)} z^2 \right],$$

cioè $X + Y \sim N(0, \sigma_1^2 + \sigma_2^2)$. La proposizione è dunque dimostrata nel caso $\mu_1 = \mu_2 = 0$.

In generale, per quanto appena visto, abbiamo che $(X - \mu_1) + (Y - \mu_2) \sim N(0, \sigma_1^2 + \sigma_2^2)$, da cui segue che $X + Y = (X - \mu_1) + (Y - \mu_2) + \mu_1 + \mu_2 \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$. □

5.4 Calcoli con densità: alcuni esempi

Esempio 5.12. Sia $X \sim N(0, 1)$ e $Y = X^2$. Allora

$$F_Y(x) = 0 \quad \text{se } x < 0.$$

Se invece $x \geq 0$

$$\begin{aligned} F_Y(x) &= P(X^2 \leq x) = P(-\sqrt{x} \leq X \leq \sqrt{x}) = P(X \leq \sqrt{x}) - P(X < -\sqrt{x}) \\ &= F_X(\sqrt{x}) - F_X(-\sqrt{x}), \end{aligned}$$

dove si è usato il fatto che

$$P(X < -\sqrt{x}) = P(X \leq -\sqrt{x}) = F_X(-\sqrt{x})$$

(perchè?). Dunque F_Y è \mathcal{C}^1 su $\mathbb{R} \setminus \{0\}$, continua su \mathbb{R} , e dunque

$$\begin{aligned} f_Y(x) &= F'_Y(x) = \frac{1}{2\sqrt{x}} [F'_X(\sqrt{x}) + F'_X(-\sqrt{x})] \\ &= \frac{1}{2\sqrt{x}} [f_X(\sqrt{x}) + f_X(-\sqrt{x})] 1_{[0,+\infty)}(x) = \frac{1}{\sqrt{2x\pi}} e^{-x/2} 1_{[0,+\infty)}(x), \end{aligned}$$

ossia $Y \sim \chi^2(1) = \Gamma(\frac{1}{2}, \frac{1}{2})$.

Esempio 5.13. Siano $X_1, X_2, \dots, X_n \sim N(0, 1)$ indipendenti. Dall'Esempio 5.12 e dalla Proposizione 5.8, segue che

$$X_1^2 + X_2^2 + \dots + X_n^2 \sim \Gamma\left(\frac{n}{2}, \frac{1}{2}\right) = \chi^2(n).$$

Esempio 5.14. Siano $X, Y \sim \text{Exp}(\lambda)$ indipendenti. Definiamo $Z = \max(X, Y)$ e $W = \min(X, Y)$. Vogliamo determinare la distribuzione di Z e W . Quanto visto nel caso discreto nella Proposizione 3.60 riguardo alla funzione di ripartizione del massimo e del minimo di variabili casuali indipendenti continua a valere per variabili casuali generiche, in quanto non veniva usata la discretezza dello spazio di probabilità. Perciò

$$F_Z(x) = F_X(x)F_Y(x) = F_X^2(x), \quad F_W(x) = 1 - (1 - F_X(x))(1 - F_Y(x)) = 1 - (1 - F_X(x))^2.$$

Essendo

$$F_X(x) = F_Y(x) = [1 - e^{-\lambda x}]1_{[0, +\infty)}(x),$$

abbiamo che F_Z e F_W sono di classe \mathcal{C}^1 a tratti, e dunque

$$f_Z(x) = 2F_X(x)f_X(x) = 2\lambda(1 - e^{-\lambda x})e^{-\lambda x}1_{[0, +\infty)}(x),$$

e

$$f_W(x) = 2(1 - F_X(x))f_X(x) = 2\lambda e^{-2\lambda x}1_{[0, +\infty)}(x).$$

In particolare, $W \sim \text{Exp}(2\lambda)$.

Capitolo 6

Teoremi limite classici

I Teoremi limite classici, la *legge dei grandi numeri* e il *teorema limite centrale*, costituiscono il nucleo del Calcolo delle Probabilità, per la loro portata sia teorica che applicativa. La legge dei grandi numeri è, tra l'altro, alla base della molti algoritmi che utilizzano metodi probabilistici (metodi di Monte Carlo). Il teorema limite centrale giustifica il ruolo centrale che le variabili casuali Gaussiane hanno nella modellistica e nella statistica. Inoltre consente di effettuare numerosi calcoli approssimati di probabilità di interesse applicativo (approssimazione normale).

6.1 La legge dei grandi numeri

Sia data una successione $(X_n)_{n \geq 1}$ di variabili casuali scalari definite nello stesso spazio di probabilità (Ω, \mathcal{A}, P) . Per avere più chiaro il senso di ciò che segue, si può immaginare che le X_n rappresentino misurazioni successive di una grandezza, ad esempio una grandezza fisica, la cui aleatorietà è dovuta all'imprecisione degli strumenti di misura. Se si effettuano n misure successive, è assai naturale considerare la media aritmetica dei risultati ottenuti, cioè

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

Nel linguaggio del calcolo delle probabilità, \bar{X}_n è detta *media campionaria*. Una parte considerevole dei teoremi limite del calcolo delle probabilità riguarda il comportamento asintotico, per $n \rightarrow +\infty$, della media campionaria.

Definizione 6.1. Si assuma che le X_n ammettano tutte la stessa media $\mu \in (0, \infty)$. Diremo che la successione (X_n) soddisfa alla legge debole dei grandi numeri se per ogni $\epsilon > 0$

$$\lim_{n \rightarrow +\infty} P(|\bar{X}_n - \mu| > \epsilon) = 0. \quad (6.1)$$

In altre parole, la legge dei grandi numeri afferma che per $n \rightarrow \infty$ la media campionaria \bar{X}_n converge (nel senso descritto da (6.1)) verso media probabilistica, fornendo così una giustificazione *a posteriori* della nozione di valor medio.

Resta naturalmente da stabilire sotto quali condizioni sulla successione (X_n) sia valida la legge dei grandi numeri. L'ipotesi più comunemente assunta è quella in cui le variabili X_n sono tra loro indipendenti e che abbiano tutte la stessa distribuzione. Diremo, in tal caso,

che le variabili X_n sono *indipendenti e identicamente distribuite (i.i.d.)* o, più semplicemente, che (X_n) è una successione i.i.d.. Facendo l'ipotesi aggiuntiva che le variabili X_n ammettano momento secondo, è facile dare una dimostrazione elementare della legge dei grandi numeri.

Proposizione 6.2. *Sia (X_n) una successione di variabili i.i.d. che ammettono momento secondo. Allora la successione (X_n) soddisfa alla legge debole dei grandi numeri.*

Dimostrazione. Poniamo $\mu = E(X_n)$, e $\sigma^2 = Var(X_n)$. Dalla linearità del valor medio si vede che

$$E(\bar{X}_n) = \mu. \quad (6.2)$$

Inoltre, usando il Corollario 3.49 si ha

$$Var(\bar{X}_n) = \frac{\sigma^2}{n}. \quad (6.3)$$

Ma allora, applicando a \bar{X}_n la Disuguaglianza di Chebichev, si ha

$$P(|\bar{X}_n - \mu| \geq \epsilon) \leq \frac{\sigma^2}{n\epsilon^2}, \quad (6.4)$$

da cui la tesi segue immediatamente. \square

Osservazione 6.3. Nella Proposizione 6.2, l'ipotesi che le variabili siano i.i.d. può essere notevolmente indebolita. Infatti è sufficiente assumere che le variabili X_n abbiano tutte la stessa media e la stessa varianza, cioè $E(X_n) = \mu$ e $Var(X_n) = \sigma^2$ per ogni $n \in \mathbb{N}$, e che siano scorrelate a due a due, cioè $Cov(X_i, X_j) = 0$ per $i \neq j$, e le relazioni (6.2) e (6.3) continuano a valere (si veda il Corollario 3.49). Di conseguenza anche la relazione (6.4) resta valida e si ottiene la legge debole dei grandi numeri sotto queste ipotesi più generali.

Usando una tecnica più sofisticata, è possibile dimostrare la legge debole dei grandi numeri per successioni i.i.d., senza assumere l'esistenza del momento secondo. Riportiamo di seguito l'enunciato (la dimostrazione è qui omessa).

Teorema 6.4. *Sia (X_n) una successione di variabili i.i.d. che ammettono valor medio. Allora la successione (X_n) soddisfa alla legge debole dei grandi numeri.*

Esempio 6.5. Per avere un'idea della portata applicativa della legge dei grandi numeri, consideriamo il seguente problema. Sia f una funzione di \mathbb{R} in \mathbb{R} Riemann integrabile nell'intervallo $[a, b]$. Le funzioni il cui integrale è calcolabile esattamente con metodi analitici sono, in realtà, abbastanza poche, e perciò sono importanti metodi numerici per il calcolo approssimato di tale integrale. I metodi più comuni consistono nel "discretizzare" l'intervallo $[a, b]$, e, ad esempio, approssimare $\int_a^b f(x)dx$ con la somma di Riemann

$$\frac{1}{N} \sum_{i=1}^N f\left(a + \frac{i}{N}(b-a)\right),$$

dove $1/N$ è il "passo" della discretizzazione. Tale approssimazione "ingenua", basata direttamente sulla definizione di integrale di Riemann, può essere migliorata, in particolare quando siano note ulteriori informazioni su f come, ad esempio, la sua derivata. Per una f sufficientemente "buona" è possibile, una volta fissato il massimo errore di approssimazione "tollerabile",

determinare *esattamente* quanto grande dev'essere N per garantire di non superare tale soglia di errore.

Gli algoritmi *stocastici*, ossia quelli che utilizzano generazione di numeri casuali, sono basati su un principio diverso. Vengono fissati due numeri: la soglia di errore e la massima probabilità tollerabile di commettere un errore maggiore della soglia data. In altre parole, non si pretende la *certezza* di commettere un errore piccolo, ma soltanto che sia estremamente improbabile commettere un errore maggiore della soglia fissata.

Tornando ai dettagli del problema in esame, siano $X_1, \dots, X_N \sim U(a, b)$ indipendenti. In altre parole, X_1, \dots, X_N sono N numeri casuali generati con distribuzione uniforme in $[a, b]$. Si noti che, per ogni i ,

$$E[f(X_i)] = \frac{1}{b-a} \int_a^b f(x) dx.$$

La legge dei grandi numeri applicata alle variabili casuali $f(X_1), \dots, f(X_N)$, ci dice che

$$\lim_{N \rightarrow +\infty} P \left[\left| \frac{1}{N} \sum_{i=1}^N f(X_i) - \frac{1}{b-a} \int_a^b f(x) dx \right| > \epsilon \right] = 0,$$

dove $\epsilon > 0$. Dunque, se N è sufficientemente grande, la quantità aleatoria $\frac{1}{N} \sum_{i=1}^N f(X_i)$ è una buona approssimazione dell'integrale "normalizzato" $\frac{1}{b-a} \int_a^b f(x) dx$ con probabilità elevata. Si può dire di più. Sia ϵ la soglia di errore nel calcolo di $\frac{1}{b-a} \int_a^b f(x) dx$, e $\delta > 0$ la probabilità con cui si accetta di compiere un errore maggiore di ϵ . Vogliamo determinare quanti numeri casuali dobbiamo generare affinché

$$P \left[\left| \frac{1}{N} \sum_{i=1}^N f(X_i) - \frac{1}{b-a} \int_a^b f(x) dx \right| > \epsilon \right] \leq \delta. \quad (6.5)$$

Dalla dimostrazione della Proposizione 6.2, sappiamo che la probabilità in (6.5) è minore o uguale a $\frac{Var[f(X_1)]}{\epsilon^2 N}$. Supponiamo sia nota una costante $M > 0$ tale che $|f(x)| \leq M$ per ogni $x \in [a, b]$. Allora

$$Var[f(X_1)] \leq E[f^2(X_1)] \leq M^2.$$

Ne segue che la disuguaglianza (6.5) vale se

$$\frac{M^2}{\epsilon^2 N} \leq \delta \iff N \geq \frac{M^2}{\delta \epsilon^2}. \quad (6.6)$$

Dunque, se generiamo almeno $\frac{M^2}{\delta \epsilon^2}$ numeri casuali, sappiamo che con probabilità maggiore o uguale a $1 - \delta$ la quantità $\frac{1}{N} \sum_{i=1}^N f(X_i)$ dista non più di ϵ da $\frac{1}{b-a} \int_a^b f(x) dx$.

Questo metodo per il calcolo approssimato di integrali definiti ha il vantaggio di essere molto facile da implementare, in quanto richiede solo un generatore di numeri casuali con distribuzione uniforme. Tuttavia, benchè le disuguaglianze in (6.6) possano essere migliorate, per ottenere una precisione accettabile è necessario generare molti numeri casuali, il che rende questo metodo meno efficiente degli algoritmi "deterministici". Questo discorso cambia radicalmente quando si tratta di calcolare integrali di funzioni di molte variabili. In tal caso esistono varianti multidimensionali dell'algoritmo appena descritto, che risultano, in dimensione elevata, assai più efficienti degli algoritmi deterministici.

6.1.1 Il teorema di approssimazione di Weierstrass

Tra le conseguenze più interessanti della legge dei grandi numeri (o, più semplicemente, della disuguaglianza di Chebyshev) è la seguente dimostrazione *costruttiva* del teorema di approssimazione di Weierstrass. In ciò che segue, per $f \in \mathcal{C}([0, 1])$, consideriamo la seguente successione di polinomi, detti i *Polinomi di Bernstein* di f :

$$p_n(x) = \sum_{k=0}^n \binom{n}{k} f\left(\frac{k}{n}\right) x^k (1-x)^{n-k}.$$

Teorema 6.6. *Se $f \in \mathcal{C}([0, 1])$, la successione dei Polinomi di Bernstein di f converge uniformemente a f , cioè*

$$\lim_{n \rightarrow +\infty} \sup_{x \in [0, 1]} |f(x) - p_n(x)| = 0.$$

Dimostrazione. Ricordiamo che una funzione continua su un intervallo chiuso e limitato di \mathbb{R} è limitata e uniformemente continua, cioè $M := \sup_{x \in [0, 1]} |f(x)| < +\infty$, e per ogni $\epsilon > 0$ esiste $\delta > 0$ tale che $|y - x| < \delta \Rightarrow |f(y) - f(x)| < \epsilon$.

Fissiamo dunque $\epsilon > 0$, e sia δ tale che $|y - x| < \delta \Rightarrow |f(y) - f(x)| < \epsilon/2$. Siano inoltre, per $x \in [0, 1]$, $X_1, X_2, \dots, X_n \sim Be(x)$ indipendenti. Sia inoltre

$$A_{n,\epsilon} := \{|\bar{X}_n - x| \geq \epsilon\}.$$

Per lo stesso argomento usato nella dimostrazione della legge dei grandi numeri

$$P(A_{n,\epsilon}) \leq \frac{x(1-x)}{n\epsilon^2} \leq \frac{1}{4n\epsilon^2}, \quad (6.7)$$

dove si è usato il fatto che la funzione $x \mapsto x(1-x)$ ha massimo in $x = 1/2$. A questo punto, osservando che

$$p_n(x) = E(f(\bar{X}_n)),$$

e usando (6.7), si ha

$$\begin{aligned} |f(x) - p_n(x)| &= |f(x) - E(f(\bar{X}_n))| = |E(f(x) - f(\bar{X}_n))| \leq E[|f(x) - f(\bar{X}_n)|] \\ &= E[1_{A_{n,\epsilon}} |f(x) - f(\bar{X}_n)|] + E[1_{A_{n,\epsilon}^c} |f(x) - f(\bar{X}_n)|] \\ &\leq 2MP(A_{n,\epsilon}) + \frac{\epsilon}{2}P(A_{n,\epsilon}^c) \leq \frac{M}{2n\epsilon^2} + \frac{\epsilon}{2} < \epsilon \end{aligned}$$

se $n \geq n_0$, per un opportuno n_0 . Poichè quest'ultima stima è indipendente da x , la conclusione segue facilmente. \square

6.2 Il teorema limite centrale: enunciato e metodo dell'approssimazione normale

La dimostrazione del Teorema 6.2, come abbiamo visto, è basata sul fatto che \bar{X}_n ha varianza che tende a 0 per $n \rightarrow +\infty$. Se la differenza $\bar{X}_n - \mu$, dove $\mu = E(X_i)$, viene *amplificata* di una quantità proporzionale a \sqrt{n} , si ottiene una nuova successione di variabili casuali,

$$Y_n = \sqrt{n}[\bar{X}_n - \mu], \quad (6.8)$$

che hanno media zero e varianza uguale a $\sigma^2 = \text{Var}(X_i)$ (verificarlo!). Il teorema limite centrale fornisce la distribuzione di Y_n nel limite per $n \rightarrow +\infty$.

Prima di enunciare il teorema, osserviamo che le variabili casuali Gaussiane hanno un comportamento peculiare rispetto l'espressione (6.8). Infatti, applicando ricorsivamente la Proposizione 5.11, si ha che se $X_1, X_2, \dots, X_n \sim N(\mu, \sigma^2)$ indipendenti, allora $X_1 + X_2 + \dots + X_n \sim N(n\mu, n\sigma^2)$, e quindi

$$Y_n = \sqrt{n}[\bar{X}_n - \mu] \sim N(0, \sigma^2).$$

Dunque, se le X_i hanno distribuzione Gaussiana, Y_n ha anch'essa distribuzione Gaussiana, e tale distribuzione è indipendente da n . Il teorema limite centrale afferma che, anche se le X_i non hanno distribuzione Gaussiana, la distribuzione di Y_n è "vicina", in un senso opportuno, ad una Gaussiana. Seguendo la tradizione più diffusa, enunciamo il teorema limite centrale usando la successione *normalizzata* (con varianza 1) $Z_n = \frac{1}{\sigma} Y_n$, anzichè Y_n .

Teorema 6.7. (*teorema limite centrale*). Sia (X_n) una successione *i.i.d.* di variabili casuali che ammettono momento secondo e con varianza non nulla. Posto $\mu = E(X_n)$, $\sigma^2 = \text{Var}(X_n)$ e

$$Z_n \equiv \frac{\bar{X}_n - \mu}{\sigma} \sqrt{n},$$

allora per ogni $x \in \mathbb{R}$

$$\lim_{n \rightarrow +\infty} P(Z_n \leq x) = P(Z \leq x),$$

dove $Z \sim N(0, 1)$.

Il Teorema 6.7, la cui dimostrazione verrà data nel paragrafo successivo, può essere usato per effettuare calcoli approssimati di probabilità legate alla media campionaria. Lo schema, che chiameremo di *approssimazione normale*, è quello descritto nell'esempio seguente.

Esempio 6.8. Si lancia n volte un dado equilibrato.

- Se $n = 1000$, qual è la probabilità che il punteggio totale sia minore o uguale di 3400?
- Quanto grande deve essere n affinché con probabilità maggiore o uguale a 0.99 il punteggio totale sia almeno $3.3n$?
- Quanto grande deve essere n affinché con probabilità maggiore o uguale a 0.99 il punteggio totale sia almeno 500?

La strategia che si segue è la seguente, in tutti e tre i casi. Sia X_i il punteggio ottenuto all' i -esimo lancio. Si esprime la probabilità in esame in termini di

$$Z_n \equiv \frac{\bar{X}_n - \mu}{\sigma} \sqrt{n}.$$

Successivamente, assumendo n sufficientemente grande, si sostituisce a Z_n il "limite" Z , ottenendo un valore approssimato, ma esplicitamente calcolabile.

- Vogliamo calcolare

$$P(X_1 + \dots + X_{1000} \leq 3400)$$

o, equivalentemente,

$$P(\bar{X}_{1000} \leq 3.4). \tag{6.9}$$

Si noti anzitutto che le variabili casuali X_i assumono i valori 1, 2, 3, 4, 5, 6 ognuno con probabilità 1/6. Da ciò si trova facilmente che

$$E(X_i) = \mu = 3.5$$

e

$$\text{Var}(X_i) = \sigma^2 \simeq 2.917.$$

La probabilità in (6.9) si può riscrivere nella forma, con $n = 1000$,

$$P\left(Z_n \leq \frac{3.4 - \mu}{\sigma} \sqrt{n}\right) \simeq P(Z_n \leq -1.85).$$

Se n è “sufficientemente grande”, la probabilità precedente è approssimativamente uguale a

$$P(Z \leq -1.85) = \Phi(-1.85),$$

dove con Φ denotiamo la funzione di ripartizione di una normale standard. I valori di Φ si possono trovare in apposite tavole, che tipicamente forniscono i valori di $\Phi(x)$ per $x > 0$. I rimanenti valori di Φ si ottengono osservando che, essendo la densità di Z una funzione pari,

$$\Phi(-x) = P(Z \leq -x) = P(Z \geq x) = 1 - \Phi(x).$$

Concludiamo allora che

$$P(X_1 + \cdots + X_{1000} \leq 3400) \simeq 1 - \Phi(1.85) \simeq 0.032.$$

(b) Procediamo come sopra, ma lasciando incognito il valore di n . Vogliamo che sia

$$P(X_1 + \cdots + X_n \geq 3.3n) \geq 0.99$$

o, equivalentemente,

$$0.99 \leq P(\bar{X}_n \geq 3.3) = P\left(Z_n \geq \frac{3.3 - \mu}{\sigma} \sqrt{n}\right) \simeq P(Z_n \geq -0.117\sqrt{n})$$

$$\simeq P(Z \geq -0.117\sqrt{n}) = \Phi(0.117\sqrt{n}).$$

In altre parole vogliamo trovare per quali valori di n

$$\Phi(0.117\sqrt{n}) \geq 0.99. \tag{6.10}$$

Dalle tavole per Φ si vede che

$$\Phi^{-1}(0.99) \simeq 2.326.$$

Essendo Φ strettamente crescente, (6.10) è equivalente a

$$0.117\sqrt{n} \geq 2.326 \Leftrightarrow n \geq 395,23 \Leftrightarrow n \geq 396.$$

(c) Vogliamo che sia

$$P(X_1 + \dots + X_n \geq 500) \geq 0.99 \Leftrightarrow P\left(Z_n \geq \frac{500}{n} - \mu \sqrt{n}\right) \geq 0.99.$$

Come prima, approssimiamo la precedente probabilità con quella corrispondente rimpiazzando Z_n con Z , ottenendo

$$\Phi\left(-\frac{500}{n} - \mu \sqrt{n}\right) \geq 0.99,$$

che equivale a

$$-\frac{500}{n} - \mu \sqrt{n} \geq 2.326,$$

che riscriviamo nella forma

$$3.5n - 3.973\sqrt{n} - 500 \geq 0.$$

Risolvendo la precedente come disequazione di secondo grado in \sqrt{n} , si trova

$$\sqrt{n} \geq 12.53 \Leftrightarrow n \geq 158.$$

Naturalmente, i risultati ottenuti nell'esempio precedente si possono considerare affidabili se effettivamente Z_n è "molto vicino" al suo limite Z . Esistono risultati che forniscono stime esplicite per l'errore in tale approssimazione. Ci basta qui rimarcare che se la distribuzione di X_i non è troppo asimmetrica rispetto alla media, le probabilità del tipo $P(Z_n \in I)$, con I intervallo di \mathbb{R} , sono in termini pratici indistinguibili da $P(Z \in I)$ quando n è dell'ordine di alcune decine. Nella gran parte dei casi $n \geq 30$ è sufficiente ad avere un'ottima approssimazione. Dunque, se nell'Esempio 6.8, con riferimento ai quesiti b. e c., avessimo ottenuto dei valori di n minori di 30, tali valori si sarebbero dovuti scartare, in quanto per essi il procedimento di approssimazione usato non è affidabile.

Una stima più precisa di quanto dev'essere grande n per poter usare l'approssimazione normale è nota nel caso in cui le variabili casuali $X_i \sim Be(p)$. In tal caso l'approssimazione è buona se $np \geq 5$. Se $p = 1/2$, per cui la distribuzione di X_i è esattamente simmetrica rispetto alla media, è sufficiente che sia $n \geq 10$. Nel caso di p molto vicino a 0 o a 1, per cui la distribuzione è altamente asimmetrica, sono necessari valori più grandi di n . Ad esempio, per $p = 0.01$, occorre avere $n \geq 500$!

Osservazione 6.9. Nelle parti a. e c. dell'esempio 6.8 abbiamo visto istanze del seguente problema: date n variabili casuali i.i.d. a valori interi X_1, X_2, \dots, X_n e $m \in \mathbb{N}$ calcolare, usando l'approssimazione normale,

$$P(X_1 + \dots + X_n \leq m). \tag{6.11}$$

Posto $\mu = E(X_i)$ e $\sigma^2 = Var(X_i)$, la probabilità in (6.11) è uguale a

$$\Phi\left(\frac{m - \mu n}{\sigma \sqrt{n}}\right). \tag{6.12}$$

Tuttavia, usando il fatto che le X_i sono a valori interi, la probabilità in (6.11) è uguale a

$$P(X_1 + \dots + X_n < m + 1)$$

che, usando di nuovo l'approssimazione normale e la continuità di Φ , è approssimativamente uguale a

$$\Phi\left(\frac{\frac{m+1}{n} - \mu}{\sigma} \sqrt{n}\right). \quad (6.13)$$

Nell'esempio 6.8 la differenza tra (6.11) e (6.13) è pressoché irrilevante, ma non è sempre così. Supponiamo, ad esempio, $n = 25$, $X_i = Be(1/2)$ e $m = 15$. In questo caso $\mu = 1/2$, $\sigma = 1/2$. Pertanto

$$\Phi\left(\frac{\frac{m}{n} - \mu}{\sigma} \sqrt{n}\right) = \Phi(1) \simeq 0.841,$$

mentre

$$\Phi\left(\frac{\frac{m+1}{n} - \mu}{\sigma} \sqrt{n}\right) = \Phi(1.4) \simeq 0.919.$$

La differenza è considerevole! Per avere comunque una buona approssimazione è opportuno usare la cosiddetta *correzione di continuità*, che consiste nel rimpiazzare m in (6.11) con $m + \frac{1}{2}$. Tale “mediazione” tra (6.11) e (6.13), nel caso delle distribuzioni “usuali” (Binomiale, Geometrica, Poisson, che hanno un andamento sufficientemente regolare) migliora talvolta considerevolmente la precisione nell'approssimazione. Nell'esempio appena considerato

$$\Phi\left(\frac{\frac{m+1/2}{n} - \mu}{\sigma} \sqrt{n}\right) = \Phi(1.2) \simeq 0.8849.$$

Il valore esatto della probabilità stimata è

$$\frac{1}{2^{25}} \sum_{k=0}^{15} \binom{25}{k} \simeq 0.8852,$$

da cui si vede l'estrema accuratezza dell'approssimazione.

6.3 Teorema limite centrale: dimostrazione

La dimostrazione del Teorema 6.7 verrà divisa in più passi. Inoltre, dimostreremo prima il teorema sotto l'ipotesi aggiuntiva che le variabili casuali X_i ammettano momento terzo. Il caso generale necessita di un po' più di attenzione, e verrà trattato alla fine di questo paragrafo. Iniziamo con alcuni lemmi ausiliari.

Lemma 6.10. *Denotiamo con \mathcal{C}_b^3 l'insieme delle funzioni di \mathbb{R} in \mathbb{R} le cui prime tre derivate esistono continue e limitate. Sia $(Z_n)_{n \geq 1}$ una successione di variabili casuali, e Z una variabile casuale la cui funzione di ripartizione F_Z è continua. Supponiamo che per ogni $g \in \mathcal{C}_b^3$ si abbia che*

$$\lim_{n \rightarrow +\infty} E[g(Z_n)] = E[g(Z)].$$

Allora per ogni $x \in \mathbb{R}$

$$\lim_{n \rightarrow +\infty} F_{Z_n}(x) = F_Z(x).$$

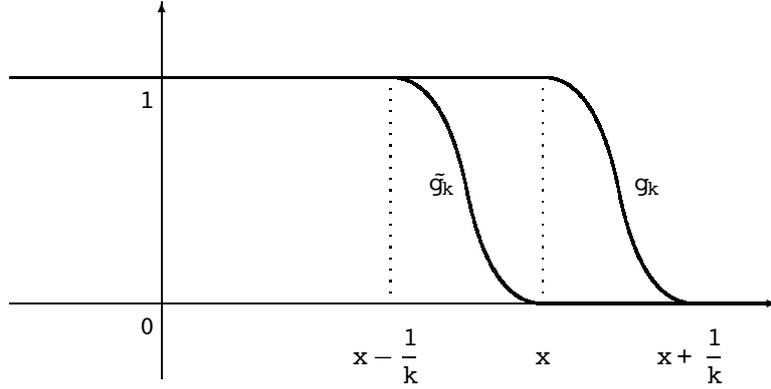


Figura 6.1: Una rappresentazione grafica delle funzioni g_k e \tilde{g}_k , che approssimano dall'alto e dal basso la funzione indicatrice $\mathbf{1}_{(-\infty, x]}(\cdot)$.

Dimostrazione. Sia $x \in \mathbb{R}$, e $k \geq 1$. Denotiamo con g_k e \tilde{g}_k due elementi di \mathcal{C}_b^3 tali che per ogni $z \in \mathbb{R}$

$$\mathbf{1}_{(-\infty, x - \frac{1}{k}]}(z) \leq \tilde{g}_k(z) \leq \mathbf{1}_{(-\infty, x]}(z) \leq g_k(z) \leq \mathbf{1}_{(-\infty, x + \frac{1}{k}]}(z).$$

(vedi Figura 6.1).

Ne segue che, per ogni variabile casuale W

$$\begin{aligned} F_W\left(x - \frac{1}{k}\right) &= E\left[\mathbf{1}_{(-\infty, x - \frac{1}{k}]}(W)\right] \leq E[\tilde{g}_k(W)] \leq E\left[\mathbf{1}_{(-\infty, x]}(W)\right] \\ &= F_W(x) \leq E[g_k(W)] \leq E\left[\mathbf{1}_{(-\infty, x + \frac{1}{k}]}(W)\right] = F_W\left(x + \frac{1}{k}\right). \end{aligned} \quad (6.14)$$

Usando le disuguaglianze in (6.14) e l'ipotesi, abbiamo

$$\limsup_{n \rightarrow +\infty} F_{Z_n}(x) \leq \lim_{n \rightarrow +\infty} E[g_k(Z_n)] = E[g(Z_n)] \leq F_Z\left(x + \frac{1}{k}\right)$$

e

$$\liminf_{n \rightarrow +\infty} F_{Z_n}(x) \geq \lim_{n \rightarrow +\infty} E[\tilde{g}_k(Z_n)] = E[\tilde{g}(Z_n)] \geq F_Z\left(x - \frac{1}{k}\right).$$

Pertanto

$$F_Z\left(x - \frac{1}{k}\right) \leq \liminf_{n \rightarrow +\infty} F_{Z_n}(x) \leq \limsup_{n \rightarrow +\infty} F_{Z_n}(x) \leq F_Z\left(x + \frac{1}{k}\right).$$

Prendendo il limite per $k \rightarrow +\infty$ nelle disuguaglianze precedenti e usando il fatto che, essendo F_Z continua,

$$\lim_{k \rightarrow +\infty} F_Z\left(x - \frac{1}{k}\right) = \lim_{k \rightarrow +\infty} F_Z\left(x + \frac{1}{k}\right) = F_Z(x),$$

troviamo

$$\liminf_{n \rightarrow +\infty} F_{Z_n}(x) = \limsup_{n \rightarrow +\infty} F_{Z_n}(x) = F_Z(x),$$

e si conclude. \square

Lemma 6.11. *Siano V, Y, Z tre variabili casuali indipendenti tali che Y, Z ammettono momento terzo, $E(Y) = E(Z)$, $E(Y^2) = E(Z^2)$. Sia $g \in \mathcal{C}_b^3$ e $C = \sup_{x \in \mathbb{R}} |g^{(3)}(x)|$. Allora*

$$|E[g(V + Y)] - E[g(V + Z)]| \leq \frac{C}{6} [E(|Y|^3) + E(|Z|^3)].$$

Dimostrazione. La formula di Taylor per funzioni di classe \mathcal{C}^3 con resto integrale ci dà, per ogni $x, h \in \mathbb{R}$

$$g(x + h) = g(x) + g'(x)h + \frac{1}{2}g''(x)h^2 + R_2(x, h),$$

dove

$$R_2(x, h) = \frac{1}{2} \int_x^{x+h} (x + h - t)^2 g^{(3)}(t) dt.$$

In particolare

$$|R_2(x, h)| \leq \frac{C}{6} |h|^3. \quad (6.15)$$

Si ricava facilmente che

$$g(x + h) - g(x + k) = g'(x)[h - k] + \frac{1}{2}g''(x)[h^2 - k^2] + R_2(x, h) - R_2(x, k). \quad (6.16)$$

A questo punto poniamo, in (6.15), $x = V, h = Y, k = Z$ e prendiamo la media:

$$E[g(V + Y)] - E[g(V + Z)] = E[g'(V)(Y - Z)] + \frac{1}{2}E[g''(V)(Y^2 - Z^2)] + E[R_2(V, Y) - R_2(V, Z)].$$

Ma, essendo V, Y, Z indipendenti e $E(Y) = E(Z)$, $E(Y^2) = E(Z^2)$, abbiamo

$$E[g'(V)(Y - Z)] = E[g'(V)]E[(Y - Z)] = 0 \quad E[g''(V)(Y^2 - Z^2)] = E[g''(V)]E[(Y^2 - Z^2)] = 0.$$

Ne segue, anche usando (6.15),

$$\begin{aligned} |E[g(V + Y)] - E[g(V + Z)]| &= |E[R_2(V, Y) - R_2(V, Z)]| \leq E[|R_2(V, Y)|] + E[|R_2(V, Z)|] \\ &\leq \frac{C}{6} [E(|Y|^3) + E(|Z|^3)], \end{aligned}$$

che completa la dimostrazione. \square

La seguente proposizione rappresenta il “cuore” della dimostrazione del Teorema 6.7.

Proposizione 6.12. *Siano Y_1, Y_2, \dots, Y_n variabili casuali i.i.d., che ammettono momento terzo, e tali che $E(Y_1) = 0$, $E(Y_1^2) = 1$. Siano inoltre W_1, W_2, \dots, W_n variabili casuale con le medesime proprietà, cioè sono i.i.d., $E(W_1) = 0$, $E(W_1^2) = 1$. Sia inoltre $g \in \mathcal{C}_b^3$ e $C := \sup_{x \in \mathbb{R}} |g^{(3)}(x)|$. Allora*

$$\left| E \left[g \left(\frac{Y_1 + \dots + Y_n}{\sqrt{n}} \right) \right] - E \left[g \left(\frac{W_1 + \dots + W_n}{\sqrt{n}} \right) \right] \right| \leq \frac{C}{6} \frac{E(|Y_1|^3) + E(|W_1|^3)}{\sqrt{n}}.$$

Dimostrazione. Sia $Y := (Y_1, Y_2, \dots, Y_n)$ e $W := (W_1, W_2, \dots, W_n)$. Il risultato da dimostrare dipende solo dalle distribuzioni marginali di Y e W , ma non dalla distribuzione congiunta di Y e W . Non è perciò restrittivo assumere che Y e W siano indipendenti, cioè le variabili casuali $Y_1, Y_2, \dots, Y_n, W_1, W_2, \dots, W_n$ sono tutte indipendenti tra loro. L'idea chiave consiste nello scrivere la seguente somma telescopica:

$$\begin{aligned} & E \left[g \left(\frac{Y_1 + \dots + Y_n}{\sqrt{n}} \right) \right] - E \left[g \left(\frac{W_1 + \dots + W_n}{\sqrt{n}} \right) \right] \\ &= \sum_{k=0}^{n-1} \left\{ E \left[g \left(\frac{Y_1 + \dots + Y_k + Y_{k+1} + W_{k+2} + \dots + W_n}{\sqrt{n}} \right) \right] \right. \\ &\quad \left. - E \left[g \left(\frac{Y_1 + \dots + Y_k + W_{k+1} + W_{k+2} + \dots + W_n}{\sqrt{n}} \right) \right] \right\}. \\ &= \sum_{k=0}^{n-1} \left\{ E \left[g \left(V_k + \frac{Y_{k+1}}{\sqrt{n}} \right) \right] - E \left[g \left(V_k + \frac{W_{k+1}}{\sqrt{n}} \right) \right] \right\}, \end{aligned}$$

dove abbiamo posto

$$V_k := \frac{Y_1 + \dots + Y_k + W_{k+2} + \dots + W_n}{\sqrt{n}}.$$

Per il Lemma 6.11

$$\left| E \left[g \left(V_k + \frac{Y_{k+1}}{\sqrt{n}} \right) \right] - E \left[g \left(V_k + \frac{W_{k+1}}{\sqrt{n}} \right) \right] \right| \leq \frac{C E(|Y_1|^3) + E(|W_1|^3)}{6 n\sqrt{n}}.$$

Pertanto

$$\begin{aligned} & \left| E \left[g \left(\frac{Y_1 + \dots + Y_n}{\sqrt{n}} \right) \right] - E \left[g \left(\frac{W_1 + \dots + W_n}{\sqrt{n}} \right) \right] \right| \\ & \leq \sum_{k=0}^{n-1} \left| E \left[g \left(V_k + \frac{Y_{k+1}}{\sqrt{n}} \right) \right] - E \left[g \left(V_k + \frac{W_{k+1}}{\sqrt{n}} \right) \right] \right| \\ & \leq n \frac{C E(|Y_1|^3) + E(|W_1|^3)}{6 n\sqrt{n}} = \frac{C E(|Y_1|^3) + E(|W_1|^3)}{6 \sqrt{n}}, \end{aligned}$$

che è quanto volevamo dimostrare. □

Dalla proposizione precedente segue il fatto, assolutamente non banale e non intuitivo, che

$$\lim_{n \rightarrow +\infty} \left\{ E \left[g \left(\frac{Y_1 + \dots + Y_n}{\sqrt{n}} \right) \right] - E \left[g \left(\frac{W_1 + \dots + W_n}{\sqrt{n}} \right) \right] \right\} = 0 \quad (6.17)$$

indipendentemente dalle distribuzioni delle Y_i e delle W_i .

Dimostrazione del Teorema 6.7 con l'ipotesi aggiuntiva $E(|X_1|^3) < +\infty$. Usando le notazioni nell'enunciato del Teorema 6.7, poniamo

$$Y_i := \frac{X_i - \mu}{\sigma}.$$

Sia inoltre (W_n) una successione di variabili casuali i.i.d. con distribuzione $N(0, 1)$. Per quanto visto nella Proposizione 5.11,

$$\frac{W_1 + \cdots + W_n}{\sqrt{n}} \sim N(0, 1).$$

Quindi, se $Z \sim N(0, 1)$ e $g \in \mathcal{C}_b^3$,

$$E \left[g \left(\frac{W_1 + \cdots + W_n}{\sqrt{n}} \right) \right] = E[g(Z)]$$

ed è quindi indipendente da n . Quindi possiamo applicare (6.17) e otteniamo

$$\lim_{n \rightarrow +\infty} E \left[g \left(\frac{Y_1 + \cdots + Y_n}{\sqrt{n}} \right) \right] = E[g(Z)].$$

A questo punto, per completare la dimostrazione e dimostrare la convergenza delle funzioni di ripartizione, basta applicare il Lemma 6.10. \square

Nella parte restante di questo paragrafo vedremo come modificare la precedente dimostrazione al fine di dimostrare il Teorema 6.7 *senza* l'ipotesi aggiuntiva $E(|X_1|^3) < +\infty$. La modifica fondamentale è costituita da una versione più raffinata del Lemma 6.11

Lemma 6.13. *Siano V, Y, Z tre variabili casuali indipendenti tali che Z ammette momento terzo, $E(Y) = E(Z)$, $E(Y^2) = E(Z^2)$. Sia $g \in \mathcal{C}_b^3$ e $C_3 = \sup_{x \in \mathbb{R}} |g^{(3)}(x)|$, $C_2 := \sup_{x \in \mathbb{R}} |g''(x)|$. Allora, per ogni $\epsilon > 0$*

$$|E[g(V + Y)] - E[g(V + Z)]| \leq \frac{C_3}{2} E[|Y|^3 1_{\{|Y| \leq \epsilon\}}] + \frac{C_2}{2} E[|Y|^2 1_{\{|Y| > \epsilon\}}] + \frac{C_3}{6} E[|Z|^3]$$

Dimostrazione. La formula di Taylor arrestata al primo ordine ci dà

$$\begin{aligned} g(x+h) &= g(x) + g'(x)h + \int_x^{x+h} (x+h-t)g''(t)dt \\ &= g(x) + g'(x)h + \frac{1}{2}g''(x)h^2 + \int_x^{x+h} (x+h-t)[g''(t) - g''(x)]dt \\ &= g(x) + g'(x)h + \frac{1}{2}g''(x)h^2 + \bar{R}_2(x, h), \end{aligned}$$

dove

$$\bar{R}_2(x, h) := \int_x^{x+h} (x+h-t)[g''(t) - g''(x)]dt,$$

e dove abbiamo usato il fatto che

$$\int_x^{x+h} (x+h-t)dt = \frac{h^2}{2}.$$

Per stimare $\bar{R}_2(x, h)$ osserviamo anzitutto che

$$|\bar{R}_2(x, h)| \leq \sup_{t \in [x, x+h]} |g''(t) - g''(x)| \int_x^{x+h} (x+h-t)dt = \frac{h^2}{2} \sup_{t \in [x, x+h]} |g''(t) - g''(x)|.$$

Ora, usiamo due diverse stime della differenza $|g''(t) - g''(x)|$, per $t \in [x, x+h]$. In modo banale

$$|g''(t) - g''(x)| \leq 2C_2 \quad (6.18)$$

Inoltre, per il teorema del valor medio

$$|g''(t) - g''(x)| \leq C_3|t - x| \leq C_3h. \quad (6.19)$$

Ora, usiamo (6.18) se $|h| > \epsilon$ e (6.19) se $|h| \leq \epsilon$. Otteniamo

$$|\bar{R}_2(x, h)| \leq C_2h^2\mathbf{1}_{[-\epsilon, \epsilon]^c}(h) + \frac{C_3}{2}|h|^3\mathbf{1}_{[-\epsilon, \epsilon]}(h). \quad (6.20)$$

Applicando (6.16) esattamente come nel Lemma 6.11, ma usando il resto \bar{R}_2 per $g(V+Y)$ e R_2 per $g(V+Z)$, otteniamo:

$$E[g(V+Y)] - E[g(V+Z)] = E[g'(V)(Y-Z)] + \frac{1}{2}E[g''(V)(Y^2-Z^2)] + E[\bar{R}_2(V, Y) - R_2(V, Z)].$$

Come nel Lemma 6.11, essendo V, Y, Z indipendenti e $E(Y) = E(Z)$, $E(Y^2) = E(Z^2)$, abbiamo

$$E[g'(V)(Y-Z)] = E[g'(V)]E[(Y-Z)] = 0 \quad E[g''(V)(Y^2-Z^2)] = E[g''(V)]E[(Y^2-Z^2)] = 0.$$

Pertanto,

$$\begin{aligned} |E[g(V+Y)] - E[g(V+Z)]| &= |E[\bar{R}_2(V, Y) - R_2(V, Z)]| \leq E[|\bar{R}_2(V, Y)|] + E[|R_2(V, Z)|] \\ &\leq C_2E[Y^2\mathbf{1}_{[-\epsilon, \epsilon]^c}(Y)] + \frac{C_3}{2}E[|Y|^3\mathbf{1}_{[-\epsilon, \epsilon]}(Y)] + \frac{C}{6}E(|Z|^3) \end{aligned}$$

che completa la dimostrazione. \square

Prima di completare la dimostrazione del Teorema 6.7 dimostriamo il seguente Lemma.

Lemma 6.14. *Sia (Ω, \mathcal{A}, P) uno spazio di probabilità, e $X \in L^1(\Omega, \mathcal{A}, P)$. Allora*

$$\lim_{a \rightarrow +\infty} E[|X|\mathbf{1}_{[-a, a]^c}(X)] = 0.$$

Dimostrazione. Dimostriamo la tesi solo nel caso in cui $(\Omega, \mathcal{A}, P) = (\Omega, P)$ è uno spazio di probabilità discreto. Il caso generale richiede argomenti un po' più avanzati, non trattati in questo corso. Usando la definizione di somma infinita

$$E[|X|] = \sum_{\omega \in \Omega} |X(\omega)|P(\{\omega\}) = \sup_{A \subseteq \Omega: |A| < +\infty} \sum_{\omega \in A} |X(\omega)|P(\{\omega\}).$$

Pertanto, per ogni $\epsilon > 0$, esiste un sottoinsieme finito A_ϵ di Ω tale che

$$\sum_{\omega \in A_\epsilon} |X(\omega)|P(\{\omega\}) \geq E[|X|] - \epsilon.$$

Sia $a_\epsilon := \max\{|X(\omega)| : \omega \in A_\epsilon\}$, che ovviamente è finito poichè A_ϵ è un insieme finito. Abbiamo

$$\sum_{\omega: |X(\omega)| \leq a_\epsilon} |X(\omega)|P(\{\omega\}) \geq \sum_{\omega \in A_\epsilon} |X(\omega)|P(\{\omega\}) \geq E[|X|] - \epsilon,$$

quindi

$$E \left[|X| 1_{[-a_\epsilon, a_\epsilon]^c}(X) \right] = E[|X|] - \sum_{\omega: |X(\omega)| \leq a_\epsilon} |X(\omega)| P(\{\omega\}) \leq \epsilon.$$

Da questo la tesi del Lemma segue immediatamente. \square

Dimostrazione del Teorema 6.7. Seguiamo l'argomento e le notazioni usate nella Proposizione 6.12. Poniamo $Y_i := \frac{X_i - \mu}{\sigma}$, $W_1, W_2, \dots, W_n \sim N(0, 1)$ indipendenti. Allora, come nella dimostrazione della Proposizione 6.12, per $g \in \mathcal{C}_b^3$ si ha

$$\begin{aligned} & E \left[g \left(\frac{Y_1 + \dots + Y_n}{\sqrt{n}} \right) \right] - E \left[g \left(\frac{W_1 + \dots + W_n}{\sqrt{n}} \right) \right] \\ &= \sum_{k=0}^{n-1} \left\{ E \left[g \left(\frac{Y_1 + \dots + Y_k + Y_{k+1} + W_{k+2} + \dots + W_n}{\sqrt{n}} \right) \right] \right. \\ &\quad \left. - E \left[g \left(\frac{Y_1 + \dots + Y_k + W_{k+1} + W_{k+2} + \dots + W_n}{\sqrt{n}} \right) \right] \right\} \\ &= \sum_{k=0}^{n-1} \left\{ E \left[g \left(V_k + \frac{Y_{k+1}}{\sqrt{n}} \right) \right] - E \left[g \left(V_k + \frac{W_{k+1}}{\sqrt{n}} \right) \right] \right\}, \end{aligned}$$

con

$$V_k := \frac{Y_1 + \dots + Y_k + W_{k+2} + \dots + W_n}{\sqrt{n}}.$$

Per il Lemma 6.13

$$\begin{aligned} & \left| E \left[g \left(V_k + \frac{Y_{k+1}}{\sqrt{n}} \right) \right] - E \left[g \left(V_k + \frac{W_{k+1}}{\sqrt{n}} \right) \right] \right| \\ & \leq C_2 n^{-1} E \left[Y_1^2 1_{[-\epsilon\sqrt{n}, \sqrt{n}\epsilon]^c}(Y_1) \right] + \frac{C_3}{2n\sqrt{n}} E \left[|Y_1|^3 1_{[-\sqrt{n}\epsilon, \sqrt{n}\epsilon]}(Y) \right] + \frac{C}{6n\sqrt{n}} E(|W_1|^3). \end{aligned}$$

Perciò

$$\begin{aligned} & \left| E \left[g \left(\frac{Y_1 + \dots + Y_n}{\sqrt{n}} \right) \right] - E \left[g \left(\frac{W_1 + \dots + W_n}{\sqrt{n}} \right) \right] \right| \\ & \leq C_2 E \left[Y_1^2 1_{[-\epsilon\sqrt{n}, \sqrt{n}\epsilon]^c}(Y_1) \right] + \frac{C_3}{2\sqrt{n}} E \left[|Y_1|^3 1_{[-\sqrt{n}\epsilon, \sqrt{n}\epsilon]}(Y_1) \right] + \frac{C}{6\sqrt{n}} E(|W_1|^3) \\ & \leq C_2 E \left[Y_1^2 1_{[-\epsilon\sqrt{n}, \sqrt{n}\epsilon]^c}(Y_1) \right] + \frac{C_3}{2} \epsilon + \frac{C}{6\sqrt{n}} E(|W_1|^3), \end{aligned}$$

dove abbiamo usato il fatto che $|Y_1|^3 1_{[-\sqrt{n}\epsilon, \sqrt{n}\epsilon]}(Y_1) \leq \epsilon\sqrt{n}|Y_1|^2$. Mandando $n \rightarrow +\infty$ e usando l'arbitrarietà di ϵ , concludiamo che

$$\lim_{n \rightarrow +\infty} \left\{ E \left[g \left(\frac{Y_1 + \dots + Y_n}{\sqrt{n}} \right) \right] - E \left[g \left(\frac{W_1 + \dots + W_n}{\sqrt{n}} \right) \right] \right\} = 0.$$

Da qui in poi la dimostrazione è identica a quella vista in precedenza con l'ipotesi aggiuntiva. \square

Capitolo 7

Applicazioni alla statistica matematica

In questo capitolo applicheremo alcune delle nozioni fin qui introdotte alla *Statistica Matematica*. Lo scopo del primo paragrafo è quello di fornire qualche nozione fondamentale di Statistica Matematica, senza peraltro alcuna pretesa di sistematicità o completezza. Successivamente, studieremo alcuni rilevanti problemi di stima parametrica.

7.1 Qualche nozione di Statistica Matematica

Il Calcolo delle Probabilità fornisce modelli matematici per fenomeni aleatori; la *Statistica* ha lo scopo di confrontare il modello con “dati sperimentali”, traendone opportune conclusioni. Ad esempio, in *Statistica Parametrica*, viene proposta una classe di modelli dipendenti da un parametro; avendo a disposizione dati sperimentali, ci si chiede per esempio quale sia il valore del parametro per cui si ha il miglior accordo con i dati.

Definizione 7.1. Si dice *Modello Statistico parametrico* una famiglia di spazi di Probabilità $(\Omega, \mathcal{A}, P_\theta) : \theta \in \Theta$, dove Θ viene detto *l'insieme dei parametri*.

Un modello statistico è dunque una classe di modelli per un fenomeno aleatorio. Nell'approccio *classico* alla statistica, si assume che *esista un valore di θ che fornisce il modello corretto*: tale valore va stimato sulla base di osservazioni.

Definizione 7.2. Chiameremo *campione* una successione $(X_n)_{n \geq 1}$ di variabili aleatorie a valori reali, definite su Ω , che, per ogni probabilità P_θ , $\theta \in \Theta$, sono indipendenti ed identicamente distribuite. La sequenza finita (X_1, X_2, \dots, X_n) viene chiamata *campione di taglia n* .

Le variabili X_n rappresentano le osservazioni riguardanti il fenomeno aleatorio. La loro distribuzione dipende dal parametro θ .

- Nel caso in cui le X_n siano discrete denoteremo con $p(x; \theta)$ il valore in $x \in \mathbb{R}$ della loro comune densità. Cioè

$$p(x; \theta) := P_\theta(X_1 = x).$$

- In modo analogo, se le X_n sono variabili casuali assolutamente continue denotiamo con $f(x; \theta)$ la loro comune densità.

Nelle definizioni e risultati che seguono, avremo bisogno di varie ipotesi sul modello statistico. I risultati che enunceremo possono essere dimostrati anche con ipotesi più deboli, al costo di usare tecniche che vanno al di là degli scopi di questo testo.

Ipotesi A. L'insieme Θ è un intervallo aperto di \mathbb{R} (che può essere una semiretta aperta o tutto \mathbb{R}).

Ci limitiamo quindi al caso in cui il modello statistico dipende da un solo parametro reale.

Definizione 7.3. Una successione $(Y_n)_{n \geq 1}$ di variabili casuali della forma

$$Y_n = h_n(X_1, X_2, \dots, X_n),$$

dove $h : \mathbb{R}^n \rightarrow \mathbb{R}$, viene detta *statistica campionaria*.

Definizione 7.4. Una statistica campionaria (Y_n) si dice *stimatore corretto* se per ogni $\theta \in \Theta$ e $n \geq 1$

$$E_\theta(Y_n) = \theta,$$

dove E_θ denota il valor medio rispetto alla probabilità P_θ (in questa definizione, come nelle successive, la condizione che Y_n ammetta valor medio rispetto ad ogni probabilità P_θ è implicita).

Esempio 7.5. Consideriamo un modello statistico per il quale $p(x; \theta)$ è la densità di una variabile di Poisson di parametro $\theta \in \Theta := (0, +\infty)$. Poniamo

$$Y_n := \bar{X}_n.$$

Si vede facilmente che (Y_n) è uno stimatore corretto.

Definizione 7.6. Una statistica campionaria (Y_n) si dice *stimatore asintoticamente corretto* se per ogni $\theta \in \Theta$

$$\lim_{n \rightarrow +\infty} E_\theta(Y_n) = \theta,$$

Esempio 7.7. Consideriamo un modello statistico per cui la densità $f(x; \theta)$ è data da

$$f(x; \theta) = \frac{1}{\theta} \mathbf{1}_{(0, \theta)}(x),$$

dove $\theta \in \Theta = (0, +\infty)$. In altre parole, $X_n \sim U(0, \theta)$. Definiamo

$$Y_n := \max(X_1, X_2, \dots, X_n).$$

Anzitutto calcoliamo la funzione di ripartizione di Y_n . Per $x \in (0, \theta)$ abbiamo

$$F_{Y_n}(x; \theta) = P_\theta(Y_n \leq x) = P_\theta(X_1 \leq x, X_2 \leq x, \dots, X_n \leq x) = [P_\theta(X_1 \leq x)]^n = \frac{x^n}{\theta^n}.$$

Derivando, si ottiene

$$f_{Y_n}(x; \theta) = \frac{nx^{n-1}}{\theta^n} \mathbf{1}_{(0, \theta)}(x).$$

Quindi

$$E_\theta(Y_n) = \frac{n}{\theta^n} \int_0^\theta x^n dx = \frac{n}{n+1} \theta,$$

da cui si vede che Y_n è asintoticamente corretto.

Definizione 7.8. Una statistica campionaria (Y_n) si dice *stimatore consistente* se per ogni $\theta \in \Theta$ e ogni $\epsilon > 0$

$$\lim_{n \rightarrow +\infty} P_\theta(|Y_n - \theta| > \epsilon) = 0.$$

In termini intuitivi, possiamo dire che se (Y_n) è una statistica campionaria, allora Y_n , per n grande, è una funzione delle osservazioni che, con probabilità quasi 1, assume valori *vicini* al valore *vero* del parametro θ .

Esempio 7.9. Consideriamo la statistica dell'Esempio 7.5. La consistenza dello stimatore deriva immediatamente dalla legge dei grandi numeri.

Esempio 7.10. Consideriamo la statistica dell'Esempio 7.7. Notare che $P_\theta(Y_n \in (0, \theta)) = 1$. Quindi

$$P_\theta(|Y_n - \theta| > \epsilon) = P_\theta(Y_n < \theta - \epsilon).$$

Se $\epsilon \geq \theta$ queste probabilità valgono zero, e non c'è nulla da dimostrare. Se invece $0 < \epsilon < \theta$,

$$P_\theta(|Y_n - \theta| > \epsilon) = P_\theta(Y_n < \theta - \epsilon) = F_{Y_n}(\theta - \epsilon; \theta) = \left(\frac{\theta - \epsilon}{\theta}\right)^n,$$

da cui la consistenza segue immediatamente.

In quanto segue, se $(Z_n)_{n \geq 1}$ è una successione di variabili casuali su $(\Omega, \mathcal{A}, P_\theta)$ a valori reali tale che, per ogni $z \in \mathbb{R}$

$$\lim_{n \rightarrow +\infty} P_\theta(Z_n \leq z) = P(Z \leq z)$$

dove $Z \sim N(0, 1)$, scriveremo

$$Z_n \xrightarrow{D} N(0, 1).$$

Definizione 7.11. Una statistica campionaria (Y_n) si dice *stimatore asintoticamente normale* se per ogni $\theta \in \Theta$ esiste una costante positiva $\sigma(\theta)$, che può dipendere da θ , tale che

$$\sqrt{n} \frac{Y_n - \theta}{\sigma(\theta)} \xrightarrow{D} N(0, 1). \quad (7.1)$$

La normalità asintotica di uno stimatore permette di quantificare l'*errore di stima* che si commette nello stimare θ con Y_n . Sia $0 < \alpha \ll 1$ e definiamo

$$z_\alpha := \Phi^{-1}(1 - \alpha).$$

Si verifica facilmente che, se $Z \sim N(0, 1)$,

$$P(|Z| \leq z_{\alpha/2}) = 1 - \alpha.$$

Usando (7.1), otteniamo

$$\lim_{n \rightarrow +\infty} P_\theta \left(\left| \sqrt{n} \frac{Y_n - \theta}{\sigma(\theta)} \right| \leq z_{\alpha/2} \right) = 1 - \alpha, \quad (7.2)$$

o, equivalentemente,

$$\lim_{n \rightarrow +\infty} P_\theta \left(\theta \in \left[Y_n - \frac{\sigma(\theta)}{\sqrt{n}} z_{\alpha/2}, Y_n + \frac{\sigma(\theta)}{\sqrt{n}} z_{\alpha/2} \right] \right) = 1 - \alpha.$$

Dunque, con probabilità che tende a $1 - \alpha$ per $n \rightarrow +\infty$, Y_n dista dal valore ignoto del parametro θ al più $\frac{\sigma(\theta)}{\sqrt{n}} z_{\alpha/2}$. Quest'ultima quantità, tuttavia, dipende dal parametro incognito θ . Nei casi in cui

$$\bar{\sigma} := \sup_{t \in \Theta} \sigma(t) < +\infty,$$

si ha che

$$P_\theta \left(\theta \in \left[Y_n - \frac{\bar{\sigma}}{\sqrt{n}} z_{\alpha/2}, Y_n + \frac{\bar{\sigma}}{\sqrt{n}} z_{\alpha/2} \right] \right) \geq P_\theta \left(\theta \in \left[Y_n - \frac{\sigma(\theta)}{\sqrt{n}} z_{\alpha/2}, Y_n + \frac{\sigma(\theta)}{\sqrt{n}} z_{\alpha/2} \right] \right),$$

e quindi

$$\liminf_{n \rightarrow +\infty} P_\theta \left(\theta \in \left[Y_n - \frac{\bar{\sigma}}{\sqrt{n}} z_{\alpha/2}, Y_n + \frac{\bar{\sigma}}{\sqrt{n}} z_{\alpha/2} \right] \right) \geq 1 - \alpha.$$

L'intervallo

$$\left[Y_n - \frac{\bar{\sigma}}{\sqrt{n}} z_{\alpha/2}, Y_n + \frac{\bar{\sigma}}{\sqrt{n}} z_{\alpha/2} \right]$$

viene chiamato *intervallo di confidenza asintotico* per θ di livello di confidenza $1 - \alpha$.

Osservazione 7.12. Si noti che la normalità asintotica implica la consistenza dello stimatore. Questo si può, ad esempio, vedere da (7.2). Infatti, fissiamo $\epsilon > 0$ arbitrario. Per n sufficientemente grande, $\epsilon > \frac{z_{\alpha/2} \sigma(\theta)}{\sqrt{n}}$. Pertanto

$$\liminf_{n \rightarrow +\infty} P_\theta (|Y_n - \theta| \leq \epsilon) \geq \liminf_{n \rightarrow +\infty} P_\theta \left(|Y_n - \theta| \leq \frac{z_{\alpha/2} \sigma(\theta)}{\sqrt{n}} \right) = 1 - \alpha$$

che, per l'arbitrarietà di α , implica

$$\lim_{n \rightarrow +\infty} P_\theta (|Y_n - \theta| \leq \epsilon) = 1,$$

cioè la consistenza.

Esempio 7.13. Lo stimatore ottenuto nell'esempio 7.5 è asintoticamente normale, con $\sigma(\theta) = \sqrt{\theta}$. Ciò deriva direttamente dal teorema limite centrale.

Esempio 7.14. Consideriamo lo stimatore (Y_n) dell'esempio 7.7. Abbiamo visto che $P(Y_n \leq \theta) = 1$. pertanto, per ogni scelta di $\sigma(\theta)$, la variabile casuale

$$\sqrt{n} \frac{Y_n - \theta}{\sigma(\theta)}$$

assume solo valori negativi, e quindi non può essere

$$\sqrt{n} \frac{Y_n - \theta}{\sigma(\theta)} \xrightarrow{D} N(0, 1).$$

Quindi lo stimatore *non* è asintoticamente normale.

Dato un modello statistico ed un associato campione, vi sono varie tecniche per determinare stimatori con "buone" proprietà. Uno degli approcci più generali, e l'unico che tratteremo in questo testo, è descritto nel prossimo paragrafo.

7.2 Stimatori di Massima Verosimiglianza

Consideriamo un modello statistico $(\Omega, \mathcal{A}, P_\theta)$ con $\theta \in \Theta$, dove Θ è un intervallo aperto di \mathbb{R} , e sia $(X_n)_{n \geq 1}$ un campione. Per il momento assumiamo che le X_n siano variabili casuali discrete, con densità $p(x; \theta)$. Consideriamo n numeri reali x_1, x_2, \dots, x_n , che interpretiamo come i *valori osservati* delle variabili X_1, X_2, \dots, X_n . È ragionevole affermare che il valore di θ che *miglior si accorda* ai valori osservati è quel valore di θ che massimizza, rispetto a θ , la probabilità

$$P_\theta(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n) = p(x_1; \theta)p(x_2, \theta) \cdots p(x_n; \theta).$$

Assumendo che tale massimo esista e che sia unico, esso dipende dalla scelta di x_1, x_2, \dots, x_n : sarà cioè una *funzione* di x_1, x_2, \dots, x_n , che denoteremo con $\hat{\theta}_n(x_1, x_2, \dots, x_n)$. Possiamo dunque porre

$$Y_n := \hat{\theta}_n(X_1, X_2, \dots, X_n).$$

La successione $(Y_n)_{n \geq 1}$ si dice *stimatore di massima verosimiglianza*, o *stimatore MV*.

Per rendere più precisa la nozione appena introdotta, e per dimostrare alcune proprietà degli stimatori MV, introduciamo alcune ipotesi ulteriori.

Ipotesi B. Le variabili casuali del campione $(X_n)_{n \geq 1}$ sono discrete, con densità $p(x; \theta)$, o assolutamente continue, con densità $f(x; \theta)$. Inoltre esiste $I \in \mathcal{B}(\mathbb{R})$ tale che

- per ogni $\theta \in \Theta$, $P_\theta(X_1 \in I) = 1$;
- per ogni $x \in I$ e $\theta \in \Theta$, si ha che $p(x; \theta) > 0$ (risp. $f(x; \theta) > 0$);
- per ogni $x \in I$, la funzione $\theta \mapsto p(x; \theta)$ (risp. $\theta \mapsto f(x; \theta)$) è continua.

Ipotesi C. Sia $n \geq 1$ arbitrario. Per $\underline{x} \in I^n$ definiamo

$$L_n(\underline{x}, \theta) = \frac{1}{n} \sum_{i=1}^n \log p(x_i, \theta),$$

per il caso discreto, e analogamente per il caso assolutamente continuo con $f(x_i, \theta)$ al posto di $p(x_i, \theta)$. Per ogni $\underline{x} \in I^n$, la funzione

$$\begin{aligned} \Theta &\rightarrow \mathbb{R} \\ \theta &\mapsto L_n(\underline{x}, \theta) \end{aligned}$$

ha un unico massimo locale in $\hat{\theta}(x_1, x_2, \dots, x_n)$. Si noti che, essendo il logaritmo una funzione strettamente crescente, massimizzare $L_n(\underline{x}, \theta)$ equivale a massimizzare $p(x_1; \theta)p(x_2, \theta) \cdots p(x_n; \theta)$, come suggerito all'inizio di questo paragrafo.

Ponendo $\hat{\theta}_n(x_1, x_2, \dots, x_n) = 0$ per $(x_1, x_2, \dots, x_n) \notin I^n$, possiamo definire

$$Y_n = \hat{\theta}_n(X_1, X_2, \dots, X_n).$$

La statistica campionaria $(Y_n)_{n \geq 1}$ appena definita, si dice *stimatore di massima verosimiglianza*, o MV.

Ipotesi D. Per ogni $\theta, t \in \Theta$, con $\theta \neq t$, le distribuzioni di X_1 rispetto a P_θ e P_t sono diverse.

Ipotesi E. Per ogni $\theta, t \in \Theta$

$$E_\theta \left[(\log p(X_1, t))^2 \right] < +\infty,$$

(o $f(X_1, t)$ al posto di $p(X_1, t)$ nel caso assolutamente continuo).

Dalle Ipotesi A-E ricaveremo la consistenza dello stimatore MV. Da qui in avanti limitiamo enunciati e dimostrazioni al caso discreto; gli enunciati si modificano in modo ovvio nel caso assolutamente continuo; le dimostrazioni richiedono invece qualche attenzione in più, e saranno omesse.

Lemma 7.15. *Supponiamo che valgano le ipotesi A-E. Allora per ogni $\theta, t \in \Theta$ la somma*

$$\sum_x p(x; \theta) \log p(x; t)$$

è finita, e se $\theta \neq t$

$$\sum_x p(x; \theta) \log p(x; t) < \sum_x p(x; \theta) \log p(x; \theta).$$

Dimostrazione. Per l'Ipotesi E, la variabile casuale $\log p(X_1, t)$ ammette momento secondo rispetto a P_θ , e quindi ammette valor medio che, per la Proposizione 3.26, è dato da $\sum_x p(x; \theta) \log p(x; t)$.

Notamo ora che

$$\sum_x p(x; \theta) \log p(x; \theta) - \sum_x p(x; \theta) \log p(x; t) = \sum_{x \in I} p(x; t) \frac{p(x; \theta)}{p(x; t)} \log \frac{p(x; \theta)}{p(x; t)}.$$

Ora usiamo il seguente fatto elementare, che deriva dalla stretta convessità della funzione $h(z) = z \log z - z + 1$: per ogni $z > 0$

$$z \log z - z + 1 \geq 0,$$

e $z \log z - z + 1 = 0$ solo per $z = 1$. Pertanto, per ogni $x \in I$

$$\frac{p(x; \theta)}{p(x; t)} \log \frac{p(x; \theta)}{p(x; t)} - \frac{p(x; \theta)}{p(x; t)} + 1 \geq 0,$$

dove la disuguaglianza è stretta ogni qual volta $p(x; \theta) \neq p(x; t)$. Per Ipotesi D, tale disuguaglianza stretta vale per almeno un $x \in I$. Perciò

$$\sum_x p(x, t) \left[\frac{p(x; \theta)}{p(x; t)} \log \frac{p(x; \theta)}{p(x; t)} - \frac{p(x; \theta)}{p(x; t)} + 1 \right] > 0,$$

da cui la tesi segue facilmente. □

Teorema 7.16. *Supponiamo che valgano le ipotesi A-E. Allora la successione (Y_n) data da $Y_n = \hat{\theta}_n(X_1, X_2, \dots, X_n)$ è consistente.*

Dimostrazione. Siano $\theta, t \in \Theta$. Le variabili casuali $\log p(X_n; t)$ sono i.i.d. rispetto a P_θ e, per Ipotesi E, ammettono momento secondo. Posto

$$l(\theta, t) := E_\theta [\log p(X_n; t)] = \sum_x p(x; \theta) \log p(x; t),$$

per la legge debole dei grandi numeri, per ogni $\epsilon > 0$

$$\lim_{n \rightarrow +\infty} P_\theta (|L_n(\underline{X}, t) - l(\theta, t)| > \epsilon) = 0. \quad (7.3)$$

Ma allora

$$\begin{aligned} & P_\theta (|(L_n(\underline{X}, t) - L_n(\underline{X}, \theta)) - (l(\theta, t) - l(\theta, \theta))| > \epsilon) \\ & \leq P_\theta \left(\{|L_n(\underline{X}, t) - l(\theta, t)| > \frac{\epsilon}{2}\} \cup \{|L_n(\underline{X}, \theta) - l(\theta, \theta)| > \frac{\epsilon}{2}\} \right) \\ & \leq P_\theta \left(\{|L_n(\underline{X}, t) - l(\theta, t)| > \frac{\epsilon}{2}\} \right) + P_\theta \left(\{|L_n(\underline{X}, \theta) - l(\theta, \theta)| > \frac{\epsilon}{2}\} \right). \end{aligned}$$

Perciò, per (7.3),

$$\lim_{n \rightarrow +\infty} P_\theta (|(L_n(\underline{X}, t) - L_n(\underline{X}, \theta)) - (l(\theta, t) - l(\theta, \theta))| > \epsilon) = 0. \quad (7.4)$$

Per il Lemma 7.15, se $t \neq \theta$, si ha $l(\theta, t) - l(\theta, \theta) < 0$. Pertanto, usando (7.4) con $\epsilon < l(\theta, \theta) - l(\theta, t)$, abbiamo che, per $t \neq \theta$,

$$\lim_{n \rightarrow +\infty} P_\theta (L_n(\underline{X}, t) - L_n(\underline{X}, \theta) < 0) = 1. \quad (7.5)$$

Sia allora, per $\delta > 0$ fissato ma arbitrario,

$$A_n := \{L_n(\underline{X}, \theta - \delta) - L_n(\underline{X}, \theta) < 0\} \cap \{L_n(\underline{X}, \theta + \delta) - L_n(\underline{X}, \theta) < 0\}.$$

Per (7.5)

$$\lim_{n \rightarrow +\infty} P_\theta(A_n) = 1.$$

Se $\omega \in A_n$, si ha che $L_n(\underline{X}(\omega), \theta) > L_n(\underline{X}(\omega), \theta - \delta)$, e $L_n(\underline{X}(\omega), \theta) > L_n(\underline{X}(\omega), \theta + \delta)$. Poiché, per Ipotesi B, $t \mapsto L_n(\underline{X}(\omega), t)$ è continua in t , si ha che $L_n(\underline{X}(\omega), t)$ ha un massimo locale in $(\theta - \delta, \theta + \delta)$, per ogni $\omega \in A_n$. Per Ipotesi C, abbiamo perciò

$$A_n \subseteq \{\hat{\theta}_n(X_1, X_2, \dots, X_n) \in (\theta - \delta, \theta + \delta)\} = \{|Y_n - \theta| < \delta\}.$$

Quindi, per (7.5)

$$\lim_{n \rightarrow +\infty} P_\theta(|Y_n - \theta| < \delta) = 1,$$

cioè

$$\lim_{n \rightarrow +\infty} P_\theta(|Y_n - \theta| > \delta) = 0.$$

Per l'arbitrarietà di δ ciò dimostra la consistenza. \square

Ci occupiamo ora di fornire condizioni sufficienti affinché lo stimatore MV sia asintoticamente normale. Ci serviranno le seguenti ulteriori ipotesi sul modello statistico.

Ipotesi F. Ulteriore regolarità su $p(x; \theta)$ (resp. $f(x; \theta)$).

i) Per ogni $x \in I$ la mappa $\theta \mapsto \log p(x; \theta)$ (risp. $\log f(x; \theta)$) è di classe \mathcal{C}^2 .

ii) La funzione

$$B(x, \theta) := \frac{d^2 \log p(x; \theta)}{d\theta^2}$$

è continua in θ uniformemente per $x \in I$; cioè per ogni $\theta \in \Theta$ e $\epsilon > 0$ esiste $\delta > 0$ tale che

$$|t - \theta| < \delta \Rightarrow |B(x, t) - B(x, \theta)| < \epsilon \text{ per ogni } x \in I.$$

Nel caso assolutamente continuo la condizione è analoga con $f(x; \theta)$ al posto di $p(x; \theta)$.

iii) È lecita la seguente commutazione di somma e derivata: per $k = 1, 2$

$$\sum_x \frac{d^k}{d\theta^k} p(x; \theta) = \frac{d^k}{d\theta^k} \sum_x p(x; \theta) = \frac{d^k}{d\theta^k} 1 = 0.$$

Per il caso assolutamente continuo

$$\int \frac{d^k}{d\theta^k} f(x; \theta) = \frac{d^k}{d\theta^k} \int f(x; \theta) = \frac{d^k}{d\theta^k} 1 = 0.$$

Osservazione 7.17. La condizione iii) dell'ipotesi F è un'ipotesi abbastanza debole. Si può dimostrare che una condizione sufficiente è la seguente: per ogni $\theta \in \Theta$ esiste $\epsilon > 0$ tale che

$$\sum_x \sup_{t \in (\theta - \epsilon, \theta + \epsilon)} \left| \frac{d}{d\theta} p(x; \theta) \right| < +\infty \quad \sum_x \sup_{t \in (\theta - \epsilon, \theta + \epsilon)} \left| \frac{d^2}{d\theta^2} p(x; \theta) \right| < +\infty.$$

Osservazione 7.18. La condizione ii) dell'ipotesi F è invece assai restrittiva e, come vedremo, non è soddisfatta in molti casi interessanti. Essa può essere indebolita in modo considerevole, al prezzo di rendere più complicata la dimostrazione della normalità asintotica. Un caso particolare, ma interessante, in cui la dimostrazione che daremo può essere modificata in modo abbastanza indolore, è quello in cui $B(x, \theta)$ è della forma

$$B(x, \theta) = \varphi(x) + \psi(x)g(\theta),$$

dove g è una funzione continua e $E_\theta[|\psi(X_1)|] < +\infty$ per ogni $\theta \in \Theta$.

Ipotesi G. Per ogni $\theta \in \Theta$, la variabile casuale $B(X_1, \theta)$ ammette momento secondo rispetto a P_θ

Teorema 7.19. *Supponiamo che valgano le ipotesi A-G. Allora la successione (Y_n) data da $Y_n = \hat{\theta}_n(X_1, X_2, \dots, X_n)$ è asintoticamente normale, con $\sigma(\theta)$ dato da*

$$\frac{1}{\sigma^2(\theta)} = -E_\theta[B(X_1, \theta)] = \sum_x p(x; \theta) \left[\frac{d}{d\theta} \log p(x; \theta) \right]^2.$$

Alla dimostrazione del Teorema 7.19 premettiano i seguenti lemmi.

Lemma 7.20. *In uno spazio di probabilità (Ω, \mathcal{A}, P) , si considerino due successioni $(Z_n)_{n \geq 1}$ e $(X_n)_{n \geq 1}$ di variabili casuali a valori reali, tali che*

$$Z_n \xrightarrow{D} N(0, 1)$$

e, per ogni $\epsilon > 0$

$$\lim_{n \rightarrow +\infty} P(|X_n - 1| > \epsilon) = 0.$$

Allora

$$X_n Z_n \xrightarrow{D} N(0, 1).$$

Dimostrazione. Consideriamo le seguenti inclusioni:

$$\{X_n Z_n \leq x\} \subseteq \{X_n Z_n \leq x, |X_n - 1| \leq \epsilon\} \cup \{|X_n - 1| > \epsilon\} \subseteq \{(1 - \epsilon)Z_n \leq x\} \cup \{|X_n - 1| > \epsilon\}.$$

Quindi

$$P(X_n Z_n \leq x) \leq P(1 - \epsilon)Z_n \leq x + P(|X_n - 1| > \epsilon),$$

da cui segue

$$\limsup_{n \rightarrow +\infty} P(X_n Z_n \leq x) \leq \Phi\left(\frac{x}{1 - \epsilon}\right).$$

Facendo tendere ϵ a zero, e usando il fatto che Φ è continua, si ottiene

$$\limsup_{n \rightarrow +\infty} P(X_n Z_n \leq x) \leq \Phi(x). \quad (7.6)$$

In modo analogo

$$\begin{aligned} \{X_n Z_n \leq x\} &\supseteq \{X_n Z_n \leq x, |X_n - 1| \leq \epsilon\} \supseteq \{(1 + \epsilon)Z_n \leq x, |X_n - 1| \leq \epsilon\} \\ &= \{(1 + \epsilon)Z_n \leq x\} \setminus \{(1 + \epsilon)Z_n \leq x, |X_n - 1| > \epsilon\}. \end{aligned}$$

Pertanto

$$\begin{aligned} P(X_n Z_n \leq x) &\geq P((1 + \epsilon)Z_n \leq x) - P((1 + \epsilon)Z_n \leq x, |X_n - 1| > \epsilon) \\ &\geq P((1 + \epsilon)Z_n \leq x) - P(|X_n - 1| > \epsilon). \end{aligned}$$

Si ha allora

$$\liminf_{n \rightarrow +\infty} P(X_n Z_n \leq x) \geq \Phi\left(\frac{x}{1 + \epsilon}\right)$$

che, facendo tendere ϵ a zero, implica

$$\liminf_{n \rightarrow +\infty} P(X_n Z_n \leq x) \geq \Phi(x). \quad (7.7)$$

Mettendo assieme (7.6) e (7.7), otteniamo

$$\lim_{n \rightarrow +\infty} P(X_n Z_n \leq x) \geq \Phi(x),$$

che è quanto volevamo dimostrare. \square

Lemma 7.21. *Supponiamo che valgano le ipotesi A-G. Allora*

$$E_\theta[B(X_1, \theta)] = - \sum_x p(x; \theta) \left[\frac{d}{d\theta} \log p(x; \theta) \right]^2 = - \frac{1}{\sigma^2(\theta)}.$$

Dimostrazione. Abbiamo che

$$E_\theta[B(X_1, \theta)] = \sum_x p(x; \theta) \frac{d^2 \log p(x; \theta)}{d\theta^2}. \quad (7.8)$$

Si noti che

$$\begin{aligned} \frac{d^2 \log p(x; \theta)}{d\theta^2} &= \frac{d}{d\theta} \left[\frac{\frac{d}{d\theta} p(x; \theta)}{p(x; \theta)} \right] = \frac{1}{p(x; \theta)} \frac{d^2}{d\theta^2} p(x; \theta) - \frac{\left[\frac{d}{d\theta} p(x; \theta) \right]^2}{p^2(x; \theta)} \\ &= \frac{1}{p(x; \theta)} \frac{d^2}{d\theta^2} p(x; \theta) - \left[\frac{d}{d\theta} \log p(x; \theta) \right]^2. \end{aligned}$$

Pertanto

$$\begin{aligned} \sum_x p(x; \theta) \frac{d^2 \log p(x; \theta)}{d\theta^2} &= \sum_x \frac{d^2}{d\theta^2} p(x; \theta) - \sum_x p(x; \theta) \left[\frac{d}{d\theta} \log p(x; \theta) \right]^2 \\ &= - \sum_x p(x; \theta) \left[\frac{d}{d\theta} \log p(x; \theta) \right]^2, \end{aligned}$$

in quanto, per l'Ipotesi F,

$$\sum_x \frac{d^2}{d\theta^2} p(x; \theta) = 0.$$

Con ciò la dimostrazione è completa. □

Dimostrazione del Teorema 7.19. Per $\underline{x} = (x_1, x_2, \dots, x_n) \in I^n$ e $\theta \in \Theta$ poniamo

$$H_n(\underline{x}, \theta) := \frac{d}{d\theta} L_n(\underline{x}, \theta) = \frac{1}{n} \sum_{i=1}^n \frac{d}{d\theta} \log p(x_i; \theta).$$

Inoltre definiamo

$$B_n(\underline{x}, t, \theta) := \begin{cases} \frac{H_n(\underline{x}, t) - H_n(\underline{x}, \theta)}{t - \theta} & \text{per } t \neq \theta \\ \frac{d}{d\theta} H_n(\underline{x}, \theta) = \frac{1}{n} \sum_{i=1}^n \frac{d^2}{d\theta^2} \log p(x_i; \theta) & \text{per } t = \theta. \end{cases}$$

Tanto H_n che B_n si possono definire per ogni $\underline{x} \in \mathbb{R}^n$, ponendole identicamente uguali a zero se $\underline{x} \notin I^n$. Per il teorema della media o di Lagrange, per ogni $t, \theta \in \Theta$ e ogni $\underline{x} \in \mathbb{R}^n$, esiste θ^* con $|\theta^* - \theta| \leq |t - \theta|$

$$B_n(\underline{x}, t, \theta) = \frac{d}{d\theta} H_n(\underline{x}, \theta^*) = \frac{1}{n} \sum_{i=1}^n \frac{d^2}{d\theta^2} \log p(x_i; \theta^*).$$

Fissiamo $\epsilon > 0$. Per la parte ii) dell'Ipotesi F, esiste $\delta > 0$ tale che se $|t - \theta| < \delta$ si ha

$$\left| \frac{d^2}{d\theta^2} \log p(x; t) - \frac{d^2}{d\theta^2} \log p(x; \theta) \right| < \epsilon$$

per ogni $x \in I$. Pertanto, per ogni $\underline{x} \in \mathbb{R}^n$, se $|t - \theta| < \delta$ si ha

$$|B_n(\underline{x}, t, \theta) - B_n(\underline{x}, \theta, \theta)| \leq \frac{1}{n} \sum_{i=1}^n \left| \frac{d^2}{d\theta^2} \log p(x_i; \theta^*) - \frac{d^2}{d\theta^2} \log p(x_i; \theta) \right| \leq \epsilon.$$

Denotiamo ora con $\hat{\theta}_n$ la variabile casuale $\hat{\theta}_n(X_1, X_2, \dots, X_n)$, e

$$A_n := \{|\hat{\theta}_n - \theta| \leq \delta\}.$$

Per il Teorema 7.16 sappiamo che

$$\lim_{n \rightarrow +\infty} P_\theta(A_n) = 1. \quad (7.9)$$

Inoltre, per quanto appena visto, se $\omega \in A_n$

$$|B_n(\underline{X}(\omega), \hat{\theta}_n(\omega), \theta) - B_n(\underline{X}(\omega), \theta, \theta)| \leq \epsilon.$$

In altre parole

$$\{|B_n(\underline{X}, \hat{\theta}_n, \theta) - B_n(\underline{X}, \theta, \theta)| \leq \epsilon\} \supseteq A_n,$$

e quindi, per (7.9), per ogni $\epsilon > 0$

$$\lim_{n \rightarrow +\infty} P_\theta(|B_n(\underline{X}, \hat{\theta}_n, \theta) - B_n(\underline{X}, \theta, \theta)| \leq \epsilon) = 1. \quad (7.10)$$

Si noti ora che

$$B_n(\underline{X}, \theta, \theta) = \frac{1}{n} \sum_{i=1}^n \frac{d^2}{d\theta^2} \log p(X_i; \theta).$$

Le variabili casuali $\frac{d^2}{d\theta^2} \log p(X_i; \theta)$ sono i.i.d., ammettono momento secondo (Ipotesi G), e hanno media $-\frac{1}{\sigma^2(\theta)}$ per il Lemma 7.21. Perciò, per la legge debole di grandi numeri, per ogni $\epsilon > 0$

$$\lim_{n \rightarrow +\infty} P_\theta \left(\left| B_n(\underline{X}, \theta, \theta) + \frac{1}{\sigma^2(\theta)} \right| \leq \epsilon \right) = 1. \quad (7.11)$$

Mettendo assieme (7.10) e (7.11), abbiamo che, per ogni $\epsilon > 0$

$$\lim_{n \rightarrow +\infty} P_\theta \left(\left| B_n(\underline{X}, \hat{\theta}_n, \theta) + \frac{1}{\sigma^2(\theta)} \right| \leq \epsilon \right) = 1,$$

che è equivalente a (esercizio)

$$\lim_{n \rightarrow +\infty} P_\theta \left(\left| \left(-\frac{1}{B_n(\underline{X}, \hat{\theta}_n, \theta) \sigma^2(\theta)} \right) - 1 \right| > \epsilon \right) = 0. \quad (7.12)$$

Notiamo ora che, per $t, \theta \in \Theta$,

$$H_n(\underline{X}, t) = H_n(\underline{X}, \theta) + B_n(\underline{X}, t, \theta)(t - \theta).$$

Poichè $\hat{\theta}_n$ è un massimo locale per $L_n(\underline{X}, \theta)$, si ha che $H_n(\underline{X}, \hat{\theta}_n) = 0$. Perciò

$$H_n(\underline{X}, \theta) = -B_n(\underline{X}, \hat{\theta}_n, \theta)(\hat{\theta}_n - \theta). \quad (7.13)$$

Ricordiamo che

$$H_n(\underline{X}, \theta) = \frac{1}{n} \sum_{i=1}^n \frac{d}{d\theta} \log p(X_i; \theta).$$

Le variabili casuali $\frac{d}{d\theta} \log p(X_i; \theta)$ sono i.i.d., hanno media

$$E_\theta \left[\frac{d}{d\theta} \log p(X_i; \theta) \right] = \sum_x p(x; \theta) \frac{d}{d\theta} \log p(x; \theta) = \sum_x \frac{d}{d\theta} p(x; \theta) = 0$$

(Ipotesi F iii)), e varianza

$$E_\theta \left[\left(\frac{d}{d\theta} \log p(X_i; \theta) \right)^2 \right] = \sum_x p(x; \theta) \left(\frac{d}{d\theta} \log p(x; \theta) \right)^2 = \frac{1}{\sigma^2(\theta)},$$

per il teorema limite centrale

$$\sigma(\theta) \sqrt{n} H_n(\underline{X}, \theta) \xrightarrow{D} N(0, 1).$$

Usando il Lemma 7.20 con $Z_n = \sigma(\theta) \sqrt{n} H_n(\underline{X}, \theta)$ e $X_n = -\frac{1}{B_n(\underline{X}, \hat{\theta}_n, \theta) \sigma^2(\theta)}$, per (7.12) abbiamo

$$-\frac{1}{B_n(\underline{X}, \hat{\theta}_n, \theta) \sigma^2(\theta)} \sigma(\theta) \sqrt{n} H_n(\underline{X}, \theta) = \frac{1}{\sigma(\theta)} \sqrt{n} (\hat{\theta}_n - \theta) \xrightarrow{D} N(0, 1),$$

che è quanto si voleva dimostrare. \square

Osservazione 7.22. Illustriamo brevemente la modifica necessaria ad adattare questa dimostrazione al caso in cui non valga l'ipotesi F ii), ma la funzione $B(x, \theta)$ sia della forma speciale illustrata nell'Osservazione 7.18. L'ipotesi F ii) è stata usata solo per dimostrare che, per ogni $\epsilon > 0$

$$\lim_{n \rightarrow +\infty} P_\theta(|B_n(\underline{X}, \hat{\theta}_n, \theta) - B_n(\underline{X}, \theta, \theta)| \leq \epsilon) = 1. \quad (7.14)$$

Nelle ipotesi modificate abbiamo che

$$|B_n(\underline{X}, \hat{\theta}_n, \theta) - B_n(\underline{X}, \theta, \theta)| = \left| \frac{1}{n} \sum_{i=1}^n \psi(X_i) \right| \left| g(\hat{\theta}_n) - g(\theta) \right|.$$

Poiché g è continua, dato $\epsilon > 0$, argomentando in modo analogo alla dimostrazione precedente, possiamo mostrare che se $\omega \in A_n$, allora

$$\left| g(\hat{\theta}_n(\omega)) - g(\theta) \right| \leq \epsilon.$$

Sia

$$B_n := \left\{ \left| \frac{1}{n} \sum_{i=1}^n \psi(X_i) - m \right| \leq 1 \right\},$$

dove $m = E_\theta(\psi(X_1))$. Per la legge dei grandi numeri, $P_\theta(B_n) \rightarrow 1$ e, se $\omega \in B_n$,

$$\left| \frac{1}{n} \sum_{i=1}^n \psi(X_i(\omega)) \right| \leq m + 1.$$

Ma allora, se $\omega \in A_n \cap B_n$

$$|B_n(\underline{X}, \hat{\theta}_n, \theta) - B_n(\underline{X}, \theta, \theta)| \leq (m+1)\epsilon.$$

Poichè

$$P_\theta(A_n) \rightarrow 1, P_\theta(B_n) \rightarrow 1 \Rightarrow P_\theta(A_n \cap B_n) \rightarrow 1,$$

per l'arbitrarietà di ϵ (7.14) si deduce facilmente.

Vediamo ora alcuni esempi di stimatori MV.

Esempio 7.23. (i) Cominciamo col considerare il modello statistico dell'Esempio 7.5:

$$p(x; \theta) = e^{-\theta} \frac{\theta^x}{x!},$$

con $\Theta = (0, +\infty)$. L'Ipotesi B è verificata con $I = \mathbb{N}$.

$$L_n(\underline{x}, \theta) = \frac{1}{n} \sum_{i=1}^n \log p(x_i; \theta) = \frac{1}{n} \sum_{i=1}^n [-\theta + x_i \log \theta - \log x_i!] = -\theta + \bar{x}_n \log \theta + \frac{1}{n} \sum_{i=1}^n \log x_i!,$$

dove $\bar{x}_n = \frac{1}{n} \sum_{i=1}^n x_i$. Derivando rispetto a θ , si vede facilmente che

$$\hat{\theta}_n(\underline{x}) = \bar{x}_n.$$

Non è difficile mostrare che le Ipotesi C-G sono verificate, eccetto la F ii). Infatti si trova

$$B(x, \theta) = -\frac{x}{\theta^2}.$$

Valgono però le ipotesi modificate illustrate nell'Osservazione 7.18. In questo caso, tuttavia, la consistenza e la normalità asintotica di $\hat{\theta}_n(\bar{X})$ seguono direttamente dalla legge dei grandi numeri e dal teorema limite centrale.

(ii) Sia $T = (0, 1)$, e $X_1 \sim Ge(\theta)$, cioè, per $x \in I := \mathbb{N}$,

$$p(x; \theta) = \theta(1 - \theta)^x$$

Otteniamo

$$L_n(\underline{x}, \theta) = \log \theta + \bar{x}_n \log(1 - \theta).$$

Perciò

$$\frac{d}{d\theta} L_n(\underline{x}, \theta) = \frac{1}{\theta} - \bar{x}_n \frac{1}{1 - \theta}.$$

Ponendo $\frac{d}{d\theta} L_n(\underline{x}, \theta) = 0$, si trova

$$\hat{\theta}_n(\underline{x}) = \frac{1}{1 + \bar{x}_n},$$

che è facile vedere sia effettivamente l'unico massimo locale. Si noti che questo stimatore non soddisfa pienamente l'Ipotesi C, in quanto, se $\bar{x}_n = 0$, si ha $\hat{\theta}_n(\underline{x}) = 1 \notin \Theta$. Tuttavia, per ogni $\theta \in \Theta$

$$\lim_{n \rightarrow +\infty} P_\theta(\bar{X}_n = 0) = \lim_{n \rightarrow +\infty} \theta^n = 0.$$

Questo permette di apportare opportune modifiche alle dimostrazioni dei Teoremi 7.16 e 7.19, in modo da adattarli a questo caso. Si vede infatti che le Ipotesi C-G, a parte questo aspetto e l'ipotesi F ii) che va rimpiazzata con quella nell'Osservazione 7.18, sono verificate. Ne segue che

$$\frac{1}{1 + \bar{X}_n}$$

è uno stimatore consistente e asintoticamente normale per θ . Calcoliamo ora l'informazione di Fisher.

$$\frac{d^2}{d\theta^2} \log p(x; \theta) = -\frac{1}{\theta^2} - \frac{x}{(1-\theta)^2}.$$

Perciò, usando il Lemma 7.21,

$$\frac{1}{\sigma^2(\theta)} = -E_\theta \left[\frac{d^2}{d\theta^2} \log p(X_1; \theta) \right] = \frac{1}{\theta^2} + \frac{1}{(1-\theta)^2} E_\theta(X_1) = \frac{1}{\theta^2} + \frac{1}{(1-\theta)^2} \frac{1-\theta}{\theta} = \frac{1}{(1-\theta)\theta^2}.$$

Perciò

$$\sigma(\theta) = \theta\sqrt{1-\theta}.$$

È facile verificare che $\sigma(\theta)$ assume il suo massimo in $(0, 1)$ per $\theta = \frac{2}{3}$, da cui si ricava

$$\bar{\sigma} = \sup_{\theta \in (0,1)} \sigma(\theta) = \frac{2}{3\sqrt{3}}.$$

Quest'ultima identità consente di determinare un intervallo di confidenza per θ , come indicato nel paragrafo precedente.

(iii) Supponiamo ora che X_1 sia una variabile casuale assolutamente continua, con densità

$$f(x; \theta) = (\theta + 1)x^\theta \mathbf{1}_{(0,1)}(x),$$

dove $\theta \in \Theta = (-1, +\infty)$. L'Ipotesi A è soddisfatta con $I = (0, 1)$. Calcolando

$$\frac{d}{d\theta} \log f(x; \theta) = \frac{1}{\theta + 1} + \log x, \quad \frac{d^2}{d\theta^2} \log f(x; \theta) = -\frac{1}{(\theta + 1)^2},$$

e

$$L_n(\underline{x}; \theta) = \frac{1}{n} \sum_{i=1}^n \log f(x_i, \theta) = \log(\theta + 1) + \frac{\theta}{n} \sum_{i=1}^n \log x_i,$$

si può verificare agevolmente che le ipotesi B-G sono verificate, e

$$\hat{\theta}_n(\bar{X}) = -1 - \frac{1}{\frac{1}{n} \sum_{i=1}^n \log X_i},$$

che risulta dunque consistente e asintoticamente normale. Calcoliamo l'informazione di Fisher:

$$\frac{1}{\sigma^2(\theta)} = -E_\theta \left[\frac{d^2}{d\theta^2} \log f(X_1; \theta) \right] = \frac{1}{(\theta + 1)^2},$$

da cui

$$\sigma(\theta) = \theta + 1.$$

(iv) Supponiamo $X_1 \sim \Gamma(\alpha, \theta)$, dove $\alpha > 0$ si assume noto, e $\theta \in \Theta = (0, +\infty)$. In altre parole

$$f(x; \theta) = \frac{\theta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\theta x},$$

per $x \in I = (0, +\infty)$ (che rende soddisfatta l'Ipotesi B). Si trova

$$\log f(x; \theta) = \alpha \log \theta - \log \Gamma(\alpha) + (\alpha - 1) \log x - \theta x,$$

$$\frac{d}{d\theta} \log f(x; \theta) = \frac{\alpha}{\theta} - x, \quad \frac{d^2}{d\theta^2} \log f(x; \theta) = -\frac{\alpha}{\theta^2},$$

$$L_n(\underline{x}; \theta) = \alpha \log \theta - \log \Gamma(\alpha) + (\alpha - 1) \frac{1}{n} \log x_i - \theta \frac{1}{n} \sum_{i=1}^n x_i,$$

da cui si può verificare agevolmente che le ipotesi B-G sono verificate, e

$$\hat{\theta}_n(\bar{X}) = \frac{a}{\frac{1}{n} \sum_{i=1}^n X_i}$$

che risulta dunque consistente e asintoticamente normale. Calcoliamo l'informazione di Fisher:

$$\frac{1}{\sigma^2(\theta)} = -E_\theta \left[\frac{d^2}{d\theta^2} \log f(X_1; \theta) \right] = \frac{a}{\theta^2},$$

da cui

$$\sigma(\theta) = \frac{\theta}{\sqrt{a}}.$$

Appendice A

Somme infinite

Da scrivere

Appendice B

Un esempio di misura finitamente additiva ma non σ -additiva

Lo scopo di quanto segue è di dimostrare l'esistenza di una funzione $P : \mathcal{P}(\mathbb{N}) \rightarrow [0, 1]$ tale che $P(\mathbb{N}) = 1$, $P(A \cup B) = P(A) + P(B)$ per ogni coppia di sottoinsiemi disgiunti A e B di \mathbb{N} , ma P non è σ -additiva. Sia $l^\infty(\mathbb{N})$ l'insieme delle successioni limitate a valori in \mathbb{R} . Tale insieme ha una naturale struttura di spazio vettoriale reale: se $x = (x_n)$, $y = (y_n)$ sono elementi di $l^\infty(\mathbb{N})$, $\alpha, \beta \in \mathbb{R}$, definiamo $\alpha x + \beta y$ tramite $(\alpha x + \beta y)_n = \alpha x_n + \beta y_n$. Sia

$$V_0 = \{x \in l^\infty(\mathbb{N}) : x_n = 0 \text{ tranne che per un insieme finito di } n\}.$$

Inoltre sia $\mathbf{1}$ la successione con tutti gli elementi uguali ad 1. Poniamo

$$V_1 = \{\alpha \mathbf{1} + x : \alpha \in \mathbb{R}, x \in V_0\},$$

cioè V_1 è lo spazio vettoriale generato da V_0 e $\mathbf{1}$. Definiamo su V_1 il funzionale lineare

$$\lambda(\alpha \mathbf{1} + x) = \alpha.$$

Teorema B.1. *Esiste un funzionale lineare $\Lambda : l^\infty(\mathbb{N}) \rightarrow \mathbb{R}$ che coincide con λ su V_1 , e tale che, per ogni $x \in l^\infty(\mathbb{N})$ si ha*

$$\inf_n x_n \leq \Lambda(x) \leq \sup_n x_n. \quad (\text{B.1})$$

Prima di dimostrare questo risultato (non banale!), vediamo come usarlo per il nostro scopo. Per $A \subseteq \mathbb{N}$, sia $x^{(A)}$ l'elemento di $l^\infty(\mathbb{N})$ definito da

$$x_n^{(A)} = \begin{cases} 1 & \text{se } n \in A \\ 0 & \text{altrimenti.} \end{cases}$$

Definiamo $P(A) = \Lambda(x^{(A)})$. Per (B.1), $P(A) \in [0, 1]$. Inoltre, $P(\mathbb{N}) = \Lambda(x^{(\mathbb{N})}) = \Lambda(\mathbf{1}) = \lambda(\mathbf{1}) = 1$, e $P(A) = \Lambda(x^{(A)}) = 0$ se A è un sottoinsieme finito di \mathbb{N} . Infine, se $A \cap B = \emptyset$, allora $x^{(A)} + x^{(B)} = x^{(A \cup B)}$, da cui, per la linearità di Λ , segue che $P(A \cup B) = P(A) + P(B)$. Dunque tale P è additiva. Non è però σ -additiva, in quanto

$$1 = P(\mathbb{N}) = P\left(\bigcup_{n \in \mathbb{N}} \{n\}\right) \neq \sum_{n \in \mathbb{N}} P(\{n\}) = 0.$$

Si tratta dunque di dimostrare il teorema enunciato sopra. La dimostrazione, come il lettore più esperto riconoscerà, è un adattamento della classica dimostrazione del teorema di Hahn-Banach.

Consideriamo l'insieme Θ delle coppie (W, Λ_W) dove W è un sottospazio di $l^\infty(\mathbb{N})$ che contiene V_1 , e Λ_W è un operatore lineare che estende λ e che soddisfa (B.1) per ogni $x \in W$. Tale insieme Θ è non vuoto, in quanto $(V_1, \lambda) \in \Theta$, e può essere dotato della seguente relazione d'ordine parziale: $(W, \Lambda_W) \leq (W', \Lambda_{W'})$ se W è sottospazio di W' e $\Lambda_{W'}(x) = \Lambda_W(x)$ per ogni $x \in W$. Sia ora Σ un sottoinsieme di Θ totalmente ordinato, cioè se $(W, \Lambda_W), (W', \Lambda_{W'}) \in \Sigma$ allora o $(W, \Lambda_W) \leq (W', \Lambda_{W'})$ oppure $(W, \Lambda_W) \geq (W', \Lambda_{W'})$. Definiamo $\bar{W} = \bigcup_{W \in \Sigma} W$ e, per $x \in \bar{W}$,

$$\bar{\Lambda}(x) = \Lambda_W(x) \quad \text{se } x \in W.$$

Dal fatto che Σ è totalmente ordinato, si deduce che \bar{W} è un sottospazio di $l^\infty(\mathbb{N})$, e che la definizione di $\bar{\Lambda}$ è ben posta. Inoltre $\bar{\Lambda}$ è lineare, ed ovviamente estende λ . Dunque $(\bar{W}, \bar{\Lambda}) \in \Theta$, e $(\bar{W}, \bar{\Lambda}) \geq (W, \Lambda_W)$ per ogni $(W, \Lambda_W) \in \Sigma$. Abbiamo perciò dimostrato che ogni sottoinsieme totalmente ordinato di Θ ammette un maggiorante. Per il Lemma di Zorn, possiamo concludere che Θ ammette un elemento massimale, cioè esiste $(V, \Lambda) \in \Theta$ tale che $(V, \Lambda) \geq (W, \Lambda_W)$ per ogni $(W, \Lambda_W) \in \Theta$. Se $V = l^\infty(\mathbb{N})$, allora il teorema è dimostrato. Altrimenti, sia $y \in l^\infty(\mathbb{N}) \setminus V$, e V' lo spazio vettoriale generato da V e y . Mostriamo che il funzionale Λ si può estendere ad un funzionale Λ' su V' che soddisfa (B.1) su tutto V' . Osserviamo che $V' = \{x + \alpha y : x \in V, \alpha \in \mathbb{R}\}$. Ogni funzionale Λ' che estende Λ dev'essere della forma

$$\Lambda'(x + \alpha y) = \Lambda(x) + \alpha c,$$

per un $c \in \mathbb{R}$. Scrivendo $\sup(x)$ e $\inf(x)$ in luogo di $\sup_n x_n$ e $\inf_n x_n$, dobbiamo verificare che è possibile scegliere c in modo tale che

$$\inf(x + \alpha y) \leq \Lambda(x) + \alpha c \leq \sup(x + \alpha y) \quad (\text{B.2})$$

per ogni $x \in V$ e $\alpha \in \mathbb{R}$. Se $\alpha > 0$, osservato che $\inf(\alpha z) = \alpha \inf(z)$ e $\sup(\alpha z) = \alpha \sup(z)$, dividendo per α si vede che (B.1) vale per ogni $\alpha > 0, x \in V$ se e solo se

$$\inf(x + y) \leq \Lambda(x) + c \leq \sup(x + y) \quad (\text{B.3})$$

per ogni $x \in V$. Analogamente, se $\alpha < 0$, dividendo per $-\alpha$ si vede che (B.1) vale per ogni $\alpha < 0, x \in V$ se e solo se

$$\inf(x - y) \leq \Lambda(x) - c \leq \sup(x - y) \quad (\text{B.4})$$

per ogni $x \in V$. Mettendo tutto assieme e ricavando c in (B.3) e (B.4), abbiamo che l'estensione Λ' desiderata esiste se esiste $c \in \mathbb{R}$ per cui per ogni $x, z \in V$

$$\begin{aligned} \inf(x + y) - \Lambda(x) &\leq c \leq \sup(x + y) - \Lambda(x) \\ \Lambda(z) - \sup(z - y) &\leq c \leq \Lambda(z) - \inf(z - y). \end{aligned}$$

Osservando che una famiglia di intervalli chiusi ha intersezione non vuota se e solo se ogni estremo superiore è maggiore o uguale di ogni estremo inferiore, abbiamo che l'esistenza di un $c \in \mathbb{R}$ soddisfacente alle disuguaglianze precedenti è equivalente alla validità delle seguenti disuguaglianze, per ogni $x, z \in V$:

- (1) $\inf(x + y) - \Lambda(x) \leq \sup(z + y) - \Lambda(z)$;
- (2) $\inf(x + y) - \Lambda(x) \leq \Lambda(z) - \inf(z - y)$;
- (3) $\Lambda(z) - \sup(z - y) \leq \Lambda(x) - \inf(x - y)$;
- (4) $\Lambda(z) - \sup(z - y) \leq \sup(x + y) - \Lambda(x)$.

Mostriamo (1), essendo (2), (3) e (4) del tutto analoghe. Essa equivale a

$$\Lambda(x - z) \geq \inf(x + y) - \sup(z + y). \quad (\text{B.5})$$

Ma Λ soddisfa (B.1), perciò $\Lambda(x - z) \geq \inf(x - z)$. Dunque, per dimostrare (B.5), basta mostrare che

$$\inf(x - z) \geq \inf(x + y) - \sup(z + y) \quad (\text{B.6})$$

per ogni $x, z \in \Lambda$. Questo non è difficile. Per $m \in \mathbb{N}$

$$x_m - z_m = x_m + y_m - (z_m + y_m) \geq \inf(x + y) - \sup(z + y),$$

da cui (B.6) segue facilmente. Si lascia per esercizio la dimostrazione delle altre disuguaglianze.

Abbiamo dunque costruito (V', Λ') tale che $(V', \Lambda') > (V, \Lambda)$. Questo contraddice la massimalità di (V, Λ) . Non resta quindi che concludere che $V = l^\infty(\mathbb{N})$, il che conclude la dimostrazione.