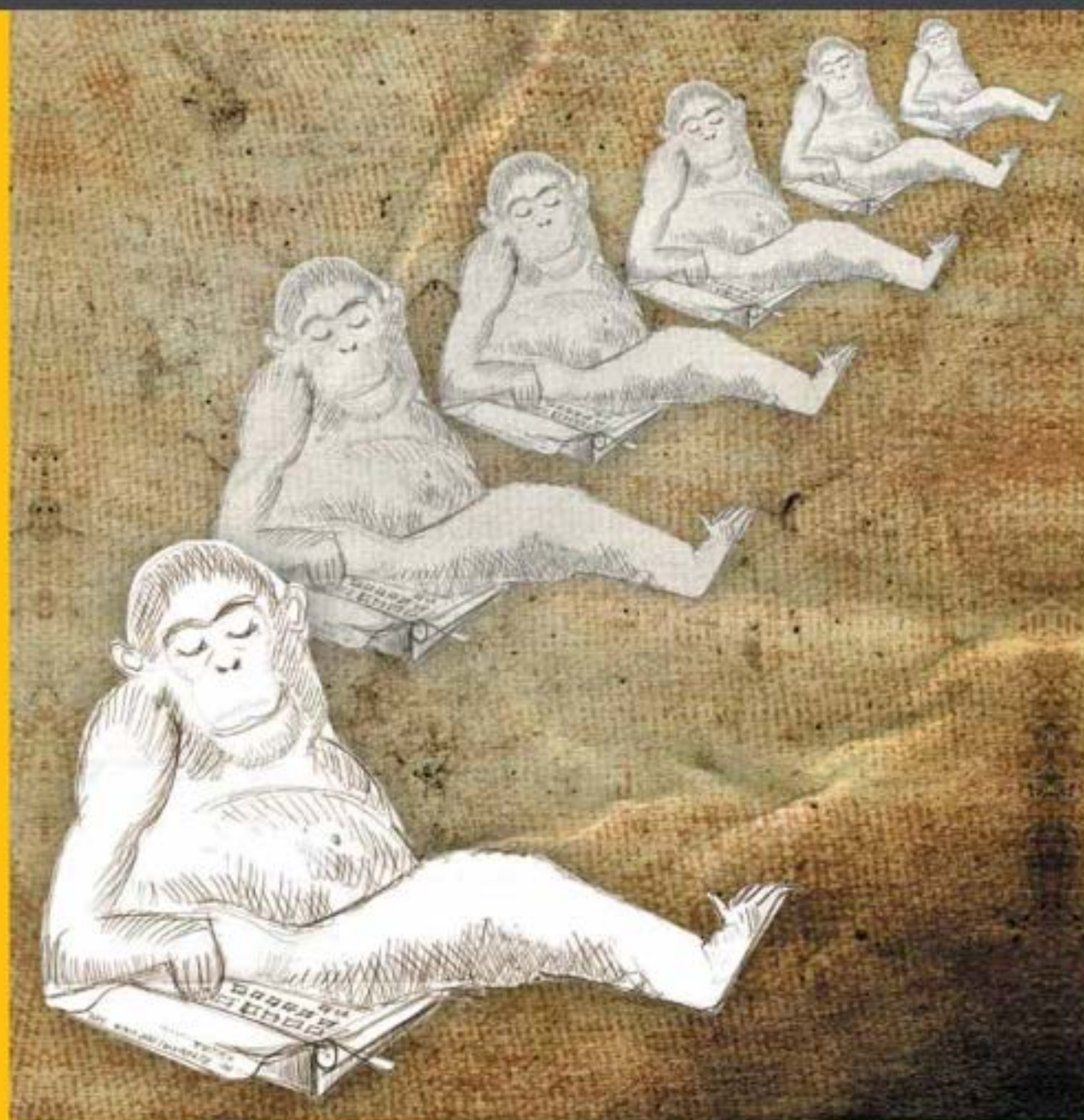


Rita Giuliano

# Argomenti di probabilità e statistica



Springer

## **Argomenti di probabilità e statistica**

Rita Giuliano

# Argomenti di probabilità e statistica



Springer

**Rita Giuliano**  
Dipartimento di Matematica  
Università di Pisa

ISBN 978-88-470-1758-0  
DOI 10.1007/978-88-470-1759-7

e-ISBN 978-88-470-1759-7

Springer Milan Dordrecht Heidelberg London New York

© Springer-Verlag Italia 2011

Quest'opera è protetta dalla legge sul diritto d'autore e la sua riproduzione è ammessa solo ed esclusivamente nei limiti stabiliti dalla stessa. Le fotocopie per uso personale possono essere effettuate nei limiti del 15% di ciascun volume dietro pagamento alla SIAE del compenso previsto dall'art. 68. Le riproduzioni per uso non personale e/o oltre il limite del 15% potranno avvenire solo a seguito di specifica autorizzazione rilasciata da AIDRO, Corso di Porta Romana n. 108, Milano 20122, e-mail [segreteria@aidro.org](mailto:segreteria@aidro.org) e sito web [www.aidro.org](http://www.aidro.org). Tutti i diritti, in particolare quelli relativi alla traduzione, alla ristampa, all'utilizzo di illustrazioni e tabelle, alla citazione orale, alla trasmissione radiofonica o televisiva, alla registrazione su microfilm o in database, o alla riproduzione in qualsiasi altra forma (stampata o elettronica) rimangono riservati anche nel caso di utilizzo parziale. La violazione delle norme comporta le sanzioni previste dalla legge.

L'utilizzo in questa pubblicazione di denominazioni generiche, nomi commerciali, marchi registrati, ecc. anche se non specificatamente identificati, non implica che tali denominazioni o marchi non siano protetti dalle relative leggi e regolamenti.

*Layout copertina:* Beatrice B., Milano

Immagine di copertina: "Il paradosso della scimmia di Borel", disegno di Nina Giuliano, elaborazione grafica di Riccardo Antonini. Riproduzione su autorizzazione

Impaginazione: PTP-Berlin, Protago TeX-Production GmbH, Germany ([www.ptp-berlin.eu](http://www.ptp-berlin.eu))  
Stampa: Isabel Litografia, Gessate (MI)

*Stampato in Italia*

Springer-Verlag Italia S.r.l., Via Decembrio 28, I-20137 Milano  
Springer-Verlag fa parte di Springer Science+Business Media ([www.springer.com](http://www.springer.com))

---

## Prefazione

Queste note sono rivolte agli studenti dei corsi di laurea triennali in cui è presente un modulo di Calcolo delle Probabilità e Statistica. Per la scelta e la quantità dei contenuti, e per la forma in cui essi sono presentati, queste pagine non costituiscono, non vogliono costituire, un vero libro di testo: oramai se ne trovano di ottimi anche in italiano (in bibliografia ne sono elencati alcuni) e non esiste certamente la necessità di produrne un altro.

Vorrei qui elencare alcuni dei “difetti” che un lettore esperto della materia sicuramente rileverà: il linguaggio usato è discorsivo, sebbene si sia cercato di mantenerlo comunque preciso e puntuale; manca una sezione di esercizi, sia svolti che semplicemente proposti: la stesura degli argomenti sarebbe risultata a mio avviso troppo frammentaria; gli esempi proposti si riferiscono a discipline diverse: quest’ultima caratteristica, se da una parte rende il testo auspicabilmente abbastanza flessibile e utilizzabile da parte di studenti di vari corsi di studio, dall’altra certamente ne è un limite, poiché diverse discipline hanno bisogno di diversi approfondimenti. Tuttavia ho ritenuto che uno studente interessato possa rimediare senza difficoltà proseguendo con letture più mirate.

Un altro “difetto” molto evidente è il fatto che alcuni punti della trattazione sono svolti in modo assai più particolareggiato di altri, e questo sicuramente provoca una certa sproporzione tra i vari argomenti; d’altra parte, per la mia pluridecennale esperienza di insegnamento, credo di avere ormai abbastanza chiare le difficoltà di uno studente con media preparazione matematica, ed i punti su cui si insiste nella trattazione sono quelli che, a mio avviso, risultano in generale più critici. Un esempio per tutti: il teorema di cambiamento di variabile negli integrali (per la ricerca della densità di una variabile funzione di un’altra) spesso viene usato in modo errato; ho deciso pertanto di dedicare a questo argomento un intero, lungo paragrafo.

Queste note devono dunque essere intese soprattutto come un aiuto per gli studenti, che, dovendo assimilare una quantità di nozioni nuove in breve tempo, per necessità o anche per scarsa maturità personale, spesso restano alla superficie della materia. L’obiettivo (forse troppo ambizioso) di queste

pagine è quello di aiutare il lettore a comprendere i concetti fondamentali della disciplina, limitando per quanto possibile gli strumenti tecnici (che, pur indispensabili, talvolta risultano preponderanti e oscurano le idee di base) e motivando in ogni caso l'introduzione di ogni nuova nozione con esempi semplici ma al tempo stesso significativi.

Desidero ringraziare il prof. M. Pratelli per avermi aiutato a rendere più completo il testo e la dott.ssa Bonadei della Springer per la pazienza con cui mi ha seguito nella stesura di questi appunti.

Livorno, febbraio 2011

*Rita Giuliano*

---

# Indice

<b>1</b>	<b>Spazi di probabilità</b>	1
1.1	Introduzione	1
1.2	Esempi e definizioni	2
1.3	Alcune conseguenze degli assiomi	6
1.4	Spazi di probabilità uniformi	10
1.5	Probabilità condizionale e indipendenza	11
1.6	Lo schema delle prove indipendenti	15
1.7	La formula di Bayes	17
<b>2</b>	<b>Variabili aleatorie</b>	21
2.1	Definizioni e prime proprietà	21
2.2	Funzione di ripartizione di una variabile aleatoria: definizione e proprietà	24
2.3	Variabili aleatorie discrete	27
2.4	Principali densità discrete	30
2.5	Vettori aleatori discreti	36
2.6	Variabili aleatorie indipendenti	40
<b>3</b>	<b>Speranza matematica e varianza</b>	45
3.1	Definizioni e proprietà	45
<b>4</b>	<b>Variabili aleatorie assolutamente continue</b>	57
4.1	Definizioni e prime proprietà	57
4.2	Alcune densità di variabili aleatorie assolutamente continue	60
4.3	Le principali densità della statistica; il teorema Limite Centrale; il teorema di Cochran	65
4.4	La funzione di ripartizione della legge normale standard	72
4.5	La Legge dei Grandi Numeri	74
4.6	Ricerca di leggi (e densità) di v. a. assolutamente continue	78

<b>5</b>	<b>La statistica inferenziale</b> .....	85
5.1	Introduzione ai problemi statistici .....	85
5.2	Il concetto di stimatore .....	86
5.3	I quantili .....	99
5.4	Intervalli di fiducia (o di confidenza) .....	101
5.5	Test statistici (parametrici) .....	108
5.6	Test per campioni gaussiani .....	112
<b>6</b>	<b>Il test del chi-quadro per l'adattamento (goodness-for-fit test)</b> .....	123
6.1	Motivazione .....	123
6.2	Descrizione del test del $\chi^2$ .....	123
6.3	Altre applicazioni del test del $\chi^2$ .....	132
<b>Appendice A    Richiami di calcolo combinatorio</b> .....		139
<b>Appendice B    Grafici e tavole</b> .....		143
<b>Bibliografia</b> .....		147



# Spazi di probabilità

## 1.1 Introduzione

Il Calcolo delle Probabilità nasce storicamente nel secolo diciassettesimo ad opera soprattutto del matematico francese Blaise Pascal, il quale ne introduce i primi concetti per risolvere alcuni problemi collegati ai giochi d'azzardo (famoso e importante è il suo carteggio con il Cavaliere de Méré, noto giocatore, che gli poneva domande riguardanti il gioco dei dadi, domande a cui Pascal rispondeva per scritto). In seguito però (soprattutto a partire dal secolo scorso e grazie all'opera di assiomatizzazione, grande merito del matematico russo Andrej Nikolaevič Kolmogorov) ci si rende conto che le stesse tecniche con cui si può rispondere a domande sui giochi d'azzardo possono essere utilizzate per trattare, più in generale, fenomeni dei quali non sia possibile avere sotto controllo tutti i fattori che ne determinano lo svolgimento, e dei quali, dunque, sia impossibile prevedere il risultato. Fenomeni di questo tipo vengono detti *casuali* o *aleatori* (dalla parola latina *alea* = *dado*: è famosa la frase di Cesare “alea iacta est ...”), e, in lingua inglese, *random*.

All'opposto dei fenomeni aleatori stanno i fenomeni *deterministici*, quelli cioè dei quali sono note tutte le condizioni iniziali e dei quali dunque è possibile stabilire lo svolgimento e l'esito, senza dover necessariamente ricorrere all'osservazione diretta. Un esempio di fenomeno deterministico è il moto di un corpo che si muova di moto rettilineo uniforme: supponendo che il corpo parta dall'origine e abbia velocità costante  $v$ , dopo un tempo (noto)  $t$  esso avrà percorso uno spazio  $s = vt$  (calcolabile dunque a partire dai dati  $v$  e  $t$ ).

In realtà ormai si ritiene superata la concezione secondo la quale, almeno in teoria, dovrebbe essere possibile controllare tutte le cause di un fenomeno; si tratta di una posizione concettuale di tipo “determinista” (originariamente proposta da un altro grande matematico francese, Pierre Simon Laplace, vissuto nel diciottesimo secolo). Piuttosto, si tende oggi a pensare ai fenomeni naturali come dotati di una loro “casualità intrinseca” e alle regole che la governano come decifrabili solo da un punto di vista probabilistico-statistico. È evidente infatti che in natura praticamente nessun fenomeno può essere con-

siderato deterministico (lo stesso esempio, fatto qui sopra, del mobile che si sposta di moto rettilineo uniforme deve essere considerato come un'astrazione ideale); è quindi altrettanto evidente che considerazioni di tipo “aleatorio” sono essenziali per lo studio di qualsiasi situazione reale, della natura più varia; le tecniche del Calcolo delle Probabilità si applicheranno allora in Biologia, Informatica, Sociologia, ecc. Per cominciare, i nostri esempi saranno tratti ancora dai giochi (lanci di monete, giochi di carte, estrazioni del lotto ecc), ma questo solo perché si tratta di esempi semplici. È bene ricordare che, come abbiamo detto sopra, il campo di applicazione del Calcolo delle Probabilità è assai più vasto.

## 1.2 Esempi e definizioni

**Esempio 1.2.1.** Si lancia una moneta non truccata. Senza effettuare materialmente il lancio, non siamo in grado di prevedere se essa darà “testa” oppure “croce”; ma, nonostante questo, c'è la possibilità di fare qualche affermazione su questo fenomeno? Intanto elenchiamo tutti i possibili risultati: sono evidentemente “testa” e “croce”, che indicheremo convenzionalmente con i simboli 1 e 0 (per motivi che chiariremo in seguito). In simboli scriviamo

$$\Omega = \{0, 1\},$$

e chiamiamo  $\Omega$  *spazio campione*. Il generico elemento di  $\Omega$  verrà indicato con il simbolo  $\omega$ . Ogni sottoinsieme di  $\Omega$  costituito da un solo  $\omega \in \Omega$  (da indicare con  $\{\omega\}$ ) si chiama *evento elementare*. Questa terminologia verrà discussa in seguito.

Tornando all'esperimento del lancio della moneta, un'altra cosa che tutti siamo capaci di fare è rispondere (anche se, per il momento, in modo solo intuitivo) alle domande: “qual è la probabilità che esca testa?” e “qual è la probabilità che esca croce?” Ad esse chiunque risponderebbe  $1/2$ : in simboli scriviamo

$$P(0) = P(1) = \frac{1}{2}.$$

Questo secondo passo, cioè, consiste nell'associare ad ogni evento elementare  $\{\omega\}$  la sua *probabilità*, che servirà a misurare, in qualche modo, il nostro *grado di fiducia* del verificarsi dell'evento considerato, o, se si preferisce un termine più semplice, l'*importanza* che vogliamo attribuirgli. Notiamo che una scrittura più precisa di quella usata sopra sarebbe

$$P(\{0\}) = P(\{1\}) = \frac{1}{2},$$

ma le parentesi interne non si scrivono per praticità.

**Esempio 1.2.2.** Si lanciano due monete equilibrate (non truccate). Cerchiamo di ripetere in questo secondo caso i passi dell'esempio precedente:

(i) costruzione dello spazio campione  $\Omega$ : si può scrivere (con evidente interpretazione dei simboli usati):

$$\Omega = \{(1, 1), (1, 0), (0, 1), (0, 0)\} = \{0, 1\}^2;$$

(ii) probabilità di ogni evento elementare: poiché le monete sono entrambe equilibrate, non c'è motivo di ritenere che uno dei possibili risultati sia *più probabile* oppure *meno probabile* di un altro, dunque scriveremo

$$P((1, 1)) = P((1, 0)) = P((0, 1)) = P((0, 0)) = \frac{1}{4}.$$

Può essere importante saper rispondere anche a domande un po' più complicate; ad esempio: qual è la probabilità che esca almeno una volta testa? Qual è la probabilità che esca esattamente una volta croce? Le risposte, molto semplici, sono rispettivamente  $3/4$  e  $1/2$ ; vediamo però un po' più in dettaglio qual è il ragionamento alla base di esse. Per il primo caso: gli eventi elementari che corrispondono al verificarsi dell'evento richiesto (cioè “esce almeno una volta testa”) sono evidentemente  $(1, 1), (1, 0), (0, 1)$ , e scriviamo

$$\{\text{esce almeno una volta testa}\} = \{(1, 1), (1, 0), (0, 1)\} = A,$$

identificando così l'evento che ci interessa con un preciso sottoinsieme  $A$  di  $\Omega$ . Per calcolare poi la probabilità, abbiamo sommato le probabilità degli eventi elementari  $\{\omega\}$  (dove  $\omega \in A$ ), che sono note. In simboli

$$P(A) = \sum_{\omega \in A} P(\omega) = \frac{3}{4}.$$

Per la seconda domanda abbiamo

$$\{\text{esce esattamente una volta croce}\} = \{(1, 0), (0, 1)\} = B;$$

$$P(B) = \sum_{\omega \in B} P(\omega) = \frac{2}{4}.$$

Come risulta evidente da questo esempio, ogni *evento* può (e sarà) identificato con un ben preciso sottoinsieme di  $\Omega$ . Inoltre, ad ogni evento abbiamo assegnato una probabilità. In altri termini, detta  $\mathcal{A}$  la famiglia degli eventi, abbiamo costruito una “funzione”  $P : \mathcal{A} \rightarrow \mathbb{R}^+$  (della quale fra poco studieremo le proprietà).

**Esempio 1.2.3.** Qual è la probabilità che alla roulette (non truccata) esca un numero pari oppure un numero dispari superiore a 13?

Si ha

$$\Omega = \{0, 1, 2, \dots, 36\}, \quad P(\omega) = \frac{1}{37};$$

poniamo poi

$$A = \{\text{esce un numero pari}\} = \{0, 2, 4, \dots, 34, 36\};$$

$$B = \{\text{esce un numero dispari superiore a 13}\} = \{15, 17, 19, \dots, 33, 35\};$$

si ha

$$P(A) = \frac{19}{37}, \quad P(B) = \frac{11}{37}.$$

Si chiede allora di calcolare  $P(A \cup B)$  (cioè la probabilità che si verifichi  $A$  oppure  $B$ ) e si ha

$$P(A \cup B) = \frac{19}{37} + \frac{11}{37} = \frac{30}{37}.$$

È chiaro che la somma è lecita perché i due eventi  $A$  e  $B$  sono disgiunti (se vogliamo usare un termine meno tecnico, due eventi cosiffatti si direbbero *incompatibili*).

La proprietà che abbiamo appena utilizzato, in modo per il momento solo intuitivo, è d'altra parte molto naturale per una funzione  $P : \mathcal{A} \rightarrow \mathbb{R}^+$  “candidata” per essere una probabilità.

Ricapitolando, per modellizzare dal punto di vista matematico un fenomeno aleatorio dovremo:

- (i) costruire lo spazio campione  $\Omega$ ;
- (ii) stabilire una famiglia  $\mathcal{A}$  di sottoinsiemi di  $\Omega$  che rappresentino la famiglia degli eventi “interessanti” per il nostro esperimento;
- (iii) definire una funzione  $P$  su  $\mathcal{A}$  e a valori in  $\mathbb{R}^+$  che fornisca le nostre valutazioni di probabilità per i singoli eventi.

Siamo ora in grado di dare le definizioni precise. Il lettore ne troverà facilmente le motivazioni riflettendo sugli esempi descritti sopra.

**Definizione 1.2.4.** Sia  $\Omega$  un insieme. Una famiglia  $\mathcal{A}$  di sottoinsiemi di  $\Omega$  è una  $\sigma$ -algebra se verifica le seguenti proprietà:

- (i)  $\Omega \in \mathcal{A}$ ;
- (ii)  $A \in \mathcal{A} \Rightarrow A^c \in \mathcal{A}$ ;
- (iii) se  $(A_n)_{n \in \mathbb{N}}$  è una successione di elementi di  $\mathcal{A}$ , allora  $\bigcup_{n \in \mathbb{N}} A_n \in \mathcal{A}$ .

**Osservazione 1.2.5.** (a) La famiglia  $\mathcal{A}$  viene detta  $\sigma$ -algebra degli eventi. La proprietà  $A \in \mathcal{A}$  si può esprimere a parole dicendo semplicemente: “ $A$  è un evento”.

- (b)  $\Omega$  viene chiamato anche *evento certo*;
- (c) se  $A$  è un evento,  $A^c$  (che è un evento per l'assioma (ii)) si chiama *evento contrario* di  $A$ ;
- (d) dalle proprietà (i) e (ii) segue facilmente che  $\emptyset = \Omega^c \in \mathcal{A}$ . Il sottoinsieme  $\emptyset$  viene detto anche *evento impossibile*;
- (e) l'assioma (iii) richiede *in particolare* che, se una famiglia finita di sottoinsiemi di  $\Omega$  è *interessante* per l'esperimento che stiamo considerando, anche la loro unione lo sia (ciò si esprime dicendo che  $\mathcal{A}$  è *stabile rispetto all'unione finita*). Questo è del tutto ragionevole: non si vede perché, se ad esempio siamo interessati al verificarsi di  $A$  e di  $B$ , non si debba esserlo anche al verificarsi di uno dei due ( $A$  oppure  $B$ ).

Ma (attenzione!) in realtà nell'assioma (iii) si pretende di più: precisamente si vuole che sia un evento anche l'unione di ogni successione *numerabile* di eventi (in altri termini,  $\mathcal{A}$  è *stabile rispetto all'unione numerabile*). Il motivo è strettamente tecnico: può succedere che alcuni sottoinsiemi *interessanti* di  $\Omega$  si esprimano in modo molto naturale come unioni *numerabili* di eventi, ma non come unioni finite, e dunque con la sola proprietà di stabilità finita non sarebbe possibile considerarli eventi. Noi stessi vedremo in seguito qualche esempio (Osservazione 2.4.9).

A noi capiterà comunque raramente di usare la proprietà di stabilità numerabile; in quasi ogni situazione ci sarà sufficiente la sola proprietà di stabilità finita.

(f) dagli assiomi (ii) e (iii) e dalle ben note leggi di De Morgan discende facilmente la seguente proprietà (*stabilità di  $\mathcal{A}$  rispetto all'intersezione numerabile*) (verifica per esercizio; qual è il significato intuitivo di questa proprietà?):

- (iv) se  $(A_n)_{n \in \mathbb{N}}$  è una successione di elementi di  $\mathcal{A}$ , allora  $\bigcap_{n \in \mathbb{N}} A_n \in \mathcal{A}$ .

**Definizione 1.2.6.** Siano  $\Omega$  un insieme,  $\mathcal{A}$  una  $\sigma$ -algebra di sottoinsiemi di  $\Omega$ . Un'applicazione  $P: \mathcal{A} \rightarrow \mathbb{R}^+$  si chiama *probabilità* (su  $\mathcal{A}$ ) se verifica le due condizioni seguenti:

- (i)  $P(\Omega) = 1$ ;
- (ii) se  $(A_n)_{n \in \mathbb{N}}$  è una successione di elementi di  $\mathcal{A}$  due a due disgiunti (cioè tali che  $\forall i \neq j$  si abbia  $A_i \cap A_j = \emptyset$ ) risulta  $P(\bigcup_{n \in \mathbb{N}} A_n) = \sum_{n \in \mathbb{N}} P(A_n)$ .

**Osservazione 1.2.7.** L'Esempio 1.2.3 mostra che l'assioma (ii) è assolutamente ragionevole, almeno nel caso di una famiglia finita di eventi (e in tal caso si dice anche che  $P$  gode della proprietà di *additività finita*). L'assioma (ii) nella sua versione estesa (detto di  $\sigma$ -*additività*) è dovuto alla necessità di assegnare una probabilità anche ad unioni *numerabili* di eventi.

**Osservazione 1.2.8.** Immaginiamo per un momento che  $\Omega$  sia un sottoinsieme del piano di area 1. Una funzione a valori reali definita sulla famiglia dei sottoinsiemi di  $\Omega$  e che abbia la proprietà di additività (o più in generale la proprietà (1.3.4) che vedremo tra poco) è la funzione *area*. Questa è in realtà più che una semplice analogia: se si riflette un attimo, si vede che l'area è proprio un modo per misurare l'importanza di un insieme. Questa è l'intuizione che ha condotto Kolmogorov (Introduzione di questo capitolo) a capire che la probabilità è una *misura*; in tal modo egli ha potuto utilizzare la teoria matematica della misura, già molto sviluppata all'epoca, e fondare così il Calcolo delle Probabilità come teoria moderna.

**Definizione 1.2.9.** Siano  $\Omega$  un insieme,  $\mathcal{A}$  una  $\sigma$ -algebra di sottoinsiemi di  $\Omega$ ,  $P$  una probabilità (su  $\mathcal{A}$ ). La terna  $(\Omega, \mathcal{A}, P)$  si chiama *spazio di probabilità*.

Ricapitolando, dunque, modellizzare un esperimento aleatorio significa costruire uno spazio di probabilità  $(\Omega, \mathcal{A}, P)$  tale che:

- (i)  $\Omega$  sia l'insieme di tutti i possibili risultati dell'esperimento;
- (ii)  $\mathcal{A}$  sia la famiglia degli eventi "interessanti" (dal punto di vista dell'esperimento in questione: se l'esperimento è per esempio il lancio di una moneta, sarà interessante chiedersi se uscirà testa, non lo sarà domandarsi se domani a Milano piove);
- (iii)  $P$  sia la valutazione di probabilità per ciascun evento, assegnata in modo da rispettare il più possibile la situazione pratica che abbiamo davanti. È bene però tenere presente che questa valutazione è comunque soggettiva e dipende dalla quantità di informazioni che lo sperimentatore ha a disposizione.

*Nota importante.* Nel caso di insiemi  $\Omega$  a cardinalità finita o numerabile (cioè con un numero di elementi finito o numerabile), tutti i sottoinsiemi di  $\Omega$  sono considerati eventi, cioè si pone *sempre*  $\mathcal{A} = \mathcal{P}(\Omega)$ .

Di seguito vedremo come gli assiomi imposti nelle definizioni permettano di effettuare il calcolo di probabilità di eventi di tipo particolare.

### 1.3 Alcune conseguenze degli assiomi

Sia  $(\Omega, \mathcal{A}, P)$  uno spazio di probabilità, fissato nel seguito una volta per tutte.

Siano  $A$  e  $B$  due eventi. Si può allora scrivere (fare un disegno)

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

Gli eventi  $(A \cap B)$  e  $(A^c \cap B)$  sono evidentemente disgiunti (gli elementi di  $A$  non possono appartenere anche ad  $A^c$  e viceversa); dunque, per la proprietà

di additività di  $P$  abbiamo

$$P(B) = P(A \cap B) + P(A^c \cap B). \quad (1.3.1)$$

Questa formula è già di per sé piuttosto importante nei calcoli: capita frequentemente di non saper calcolare direttamente la probabilità di un evento  $B$ , ma di saper “spezzare”  $B$  in due parti disgiunte (tramite un evento ausiliario  $A$ ) la cui probabilità è facilmente calcolabile. Nel seguito (formula (1.7.7)) vedremo come la (1.3.1) si generalizzi quando  $B$  deve essere scomposto in più di due parti (la chiameremo *formula della partizione dell'evento certo*).

La (1.3.1) ha alcune conseguenze importanti:

(i) per  $B = \Omega$  e per ogni  $A \in \mathcal{A}$  la (1.3.1) diventa (ricordando l'assioma (1.2.6 (i)):

$$1 = P(A) + P(A^c),$$

ovvero

$$P(A^c) = 1 - P(A). \quad (1.3.2)$$

La (1.3.2) permette dunque di calcolare la probabilità dell'evento contrario di  $A$  se è nota la probabilità di  $A$  (ed è intuitivamente accettata da tutti);

(ii) per  $A \subseteq B$  la (1.3.1) dà

$$P(B) = P(A) + P(A^c \cap B), \quad (1.3.3)$$

e quindi, poiché  $P(A^c \cap B) \geq 0$ , si ricava

$$P(A) \leq P(B), \quad \text{per } A \subseteq B.$$

Anche questa proprietà è del tutto naturale, e si chiama proprietà di *isotonia* della probabilità.

Una sua semplice conseguenza è il fatto che ogni evento ha probabilità inferiore o uguale a 1: infatti, se  $A$  è un evento, dalla ovvia relazione  $A \subseteq \Omega$  segue

$$P(A) \leq P(\Omega) = 1.$$

(In altre parole, ogni funzione *probabilità* assume valori nell'intervallo  $[0, 1]$ .)

La proprietà di isotonia si usa talvolta per mostrare che un dato evento  $A \subseteq B$  ha probabilità nulla, se si sa che  $P(B) = 0$ , oppure che un dato evento  $B \supseteq A$  ha probabilità 1, se si sa che  $P(A) = 1$  (dettagli per esercizio);

(iii) se  $A \subseteq B$ , si usa scrivere

$$A^c \cap B = B \setminus A,$$

e la (1.3.3) diventa

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

(iv) siano ora  $A$  e  $B$  due eventi qualsiasi (non necessariamente disgiunti). Possiamo scrivere (fare un disegno)

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

essendo i due eventi  $A$  e  $A^c \cap B$  disgiunti, dalla proprietà di additività della probabilità e dalla (1.3.1) si deduce

$$P(A \cup B) = P(A) + P(A^c \cap B) = P(A) + P(B) - P(A \cap B). \quad (1.3.4)$$

Questa formula permette di calcolare la probabilità dell'unione di due eventi  $A$  e  $B$  quando siano note  $P(A)$ ,  $P(B)$  e  $P(A \cap B)$ .

La (1.3.4) è un caso particolare di una formula, nota con il nome di *formula di inclusione-esclusione*, che si utilizza per il calcolo della probabilità dell'unione di una famiglia finita di eventi. Noi utilizzeremo quasi esclusivamente il caso di due eventi (cioè la (1.3.4)). Per completezza diamo (senza dimostrazione) la formula anche per il caso di tre eventi  $A$ ,  $B$  e  $C$ :

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

Il lettore provi per esercizio a immaginare la formula per il caso di quattro eventi;

(v) sia  $(A_n)_{n \in \mathbb{N}}$  una successione di eventi. Poiché

$$(\cup_{n \in \mathbb{N}} A_n)^c = (\cap_{n \in \mathbb{N}} A_n^c),$$

dalla (1.3.2) si ricava

$$P(\cup_{n \in \mathbb{N}} A_n) = 1 - P(\cap_{n \in \mathbb{N}} A_n^c). \quad (1.3.5)$$

Questa formula è utile perché spesso (ma non sempre ...) la probabilità di un'intersezione di eventi risulta più facile da calcolare della probabilità di un'unione.

**Esempio 1.3.6.** Si lancia  $n$  volte una moneta equilibrata. Quanto vale la probabilità che la faccia “testa” (contrassegnata convenzionalmente con il simbolo 1) esca almeno una volta nel corso degli  $n$  lanci?

Abbiamo già risposto a questa domanda per il caso  $n = 2$  nell'Esempio 1.2.2, dove abbiamo in pratica elencato tutti gli eventi elementari che rispondevano alla questione, sommando poi le loro probabilità. Nel caso generale ciò non è evidentemente possibile. Cominciamo allora con la costruzione dello spazio di probabilità che modella l'esperimento in questione. Analogamente al caso  $n = 2$  già visto, lo spazio campione  $\Omega$  sarà costituito da tutte



le sequenze di lunghezza  $n$  composte dai simboli 0 e 1, e cioè, per usare una scrittura comoda

$$\Omega = \{0, 1\}^n.$$

Anche in questo caso non c'è motivo di ritenere che una particolare sequenza sia più o meno probabile di un'altra, dato che la moneta è equilibrata, e dunque si pone

$$P(\omega) = \frac{1}{2^n}, \quad \forall \omega \in \Omega.$$

Poniamo ora, per  $k = 1, 2, \dots, n$

$$A_k = \{\text{esce 1 al } k\text{-esimo lancio}\}.$$

A noi interessa evidentemente calcolare  $P(\cup_{k=1}^n A_k)$ . Per la formula (1.3.5) si ha subito

$$P(\cup_{k=1}^n A_k) = 1 - P(\cap_{k=1}^n A_k^c).$$

D'altra parte è chiaro che

$$P(\cap_{k=1}^n A_k^c) = P(\{\text{esce 0 ad ogni lancio}\}) = \frac{1}{2^n},$$

(l'evento qui sopra è infatti costituito dall'unica sequenza di lunghezza  $n$  composta da soli simboli 0) e quindi la probabilità cercata vale

$$1 - \frac{1}{2^n}.$$

Tutte le proprietà viste fino a questo momento sono conseguenza della sola proprietà di additività finita della probabilità. Le ultime due, che enunciamo senza dimostrazione, utilizzano invece la  $\sigma$ -additività:

(vi) (*proprietà di passaggio al limite sulle successioni crescenti*). Sia  $(A_n)_{n \in \mathbb{N}}$  una successione *crescente* di eventi (ciò significa che,  $\forall n \in \mathbb{N}$ , risulta  $A_n \subseteq A_{n+1}$ ). Si ha allora

$$P(\cup_{n \in \mathbb{N}} A_n) = \lim_{n \rightarrow \infty} P(A_n);$$

(vii) (*proprietà di passaggio al limite sulle successioni decrescenti*). Sia  $(A_n)_{n \in \mathbb{N}}$  una successione *decrescente* di eventi (ciò significa che,  $\forall n \in \mathbb{N}$ , risulta  $A_n \supseteq A_{n+1}$ ). Si ha allora

$$P(\cap_{n \in \mathbb{N}} A_n) = \lim_{n \rightarrow \infty} P(A_n).$$

**Esercizio 1.3.7.** Dedurre la proprietà (vii) dalla (vi) e dagli assiomi di  $(\Omega, \mathcal{A}, P)$ .

## 1.4 Spazi di probabilità uniformi

In questo paragrafo vedremo un tipo di spazio di probabilità che si presenta frequentemente nelle applicazioni. In realtà, pur senza dirlo, negli esempi precedenti lo abbiamo già costruito.

**Esempio 1.4.1.** Riprendiamo l'esperimento degli  $n$  lanci di una moneta equilibrata (Esempio 1.3.6). Abbiamo già osservato che, per il modo in cui esso si svolge (cioè poiché la moneta è equilibrata), non c'è motivo di ritenere che una particolare sequenza di risultati si presenti con maggiore o minore frequenza (probabilità) di un'altra, e ciò ci ha consentito di assegnare probabilità pari a  $1/2^n$  a ciascun evento elementare. Ma perché proprio il numero  $1/2^n$ ? Notiamo che il numero  $2^n$  non è che la *cardinalità* (cioè il numero degli elementi) di  $\Omega = \{0, 1\}^n$ .

Supponiamo dunque di essere in presenza di un esperimento aleatorio *con un numero finito di possibili risultati* e che si svolge *in condizioni di equiprobabilità* (cioè, come nel caso dei lanci della moneta equilibrata, in modo che non ci sia motivo di pensare che un particolare risultato sia più o meno frequente degli altri), e sia  $\Omega$  lo spazio campione ad esso associato (che in questo paragrafo supponiamo dunque di cardinalità finita). La condizione di equiprobabilità impone evidentemente che sia

$$P(\omega) = \text{costante} = p, \quad \forall \omega \in \Omega.$$

Si tratta allora di determinare la costante  $p$ . Si ha l'eguaglianza ovvia

$$\Omega = \cup_{\omega \in \Omega} \{\omega\};$$

per gli assiomi (i) e (ii) della definizione (1.2.6) si deduce

$$1 = P(\Omega) = \sum_{\omega \in \Omega} P(\omega) = \sum_{\omega \in \Omega} p = p(\text{card } \Omega),$$

e da questa relazione segue subito

$$p = \frac{1}{\text{card } \Omega}.$$

Questo determina la probabilità di ciascun evento elementare. Per trovare ora la probabilità di un qualunque evento  $A$  si ragiona in modo sostanzialmente identico: dall'eguaglianza

$$A = \cup_{\omega \in A} \{\omega\}$$

si deduce

$$P(A) = \sum_{\omega \in A} P(\omega) = \sum_{\omega \in A} \frac{1}{\text{card } \Omega} = \frac{\text{card } A}{\text{card } \Omega}. \quad (1.4.2)$$

Questa formula definisce allora la funzione probabilità su tutta la  $\sigma$ -algebra degli eventi.

**Definizione 1.4.3.** Sia  $(\Omega, \mathcal{A}, P)$  uno spazio di probabilità, tale che  $\text{card } \Omega < +\infty$ .  $(\Omega, \mathcal{A}, P)$  si dice *spazio di probabilità uniforme* se

$$P(\omega) = \frac{1}{\text{card } \Omega}, \quad \forall \omega \in \Omega,$$

o, in modo equivalente, se

$$P(A) = \frac{\text{card } A}{\text{card } \Omega}, \quad \forall A \in \mathcal{P}(\Omega).$$

**Osservazione 1.4.4.** La formula (1.4.2) è nota a tutti: essa dice semplicemente che, *nel caso di uno spazio uniforme*, la probabilità di un evento si calcola come il rapporto tra il “numero dei casi favorevoli” e il “numero dei casi possibili”.

**Osservazione 1.4.5.** È bene ricordare, comunque, che *solo in alcuni casi* lo spazio di probabilità da costruire sarà uniforme. Si pensi per esempio all'esperimento che consiste nel lanciare  $n$  volte una moneta truccata: è chiaro che in questo caso le varie sequenze di risultati non potranno essere considerate equiprobabili.

## 1.5 Probabilità condizionale e indipendenza

Cominciamo con un esempio semplice.

**Esempio 1.5.1.** Da un'urna contenente 10 palline numerate da 1 a 10 si esegue un'estrazione. Quanto vale la probabilità che la pallina estratta porti un numero  $\leq 5$ ?

Sia  $\Omega = \{1, 2, 3, \dots, 10\}$ ; si tratta evidentemente di uno spazio uniforme e dunque per ogni evento elementare  $\{\omega\}$ , con  $\omega \in \Omega$ , poniamo

$$P(\omega) = \frac{1}{\text{card } \Omega} = \frac{1}{10}.$$

Poiché

$$A = \{\text{esce un numero } \leq 5\} = \{1, 2, 3, 4, 5\},$$

avremo

$$P(A) = \frac{\text{card } A}{\text{card } \Omega} = \frac{5}{10}.$$

In altre parole, le informazioni a nostra disposizione ci hanno indotto a costruire lo spazio di probabilità sopra detto, e di conseguenza abbiamo ottenuto, come “grado di fiducia” del verificarsi di  $A$ , il numero  $5/10$ .

Supponiamo adesso che un tizio che ha assistito all'estrazione ci abbia detto che è uscito un numero pari. Questa informazione supplementare ci induce a cambiare il nostro “grado di fiducia” del verificarsi di  $A$ ? In caso affermativo, quale nuova valutazione di probabilità dovremo dare?

La risposta è molto semplice: adesso il nostro spazio campione sarà ridotto al sottoinsieme di  $\Omega$

$$B = \{\text{esce un numero pari}\} = \{2, 4, 6, 8, 10\}.$$

Si tratterà ancora di uno spazio uniforme e quindi per ogni evento elementare  $\{\omega\}$  avremo

$$P(\omega) = \frac{1}{\text{card } B} = \frac{1}{5},$$

e, poiché due elementi di  $B$  sono  $\leq 5$ , (cioè  $\text{card}(A \cap B) = 2$ ), concludiamo che un numero inferiore o uguale a 5 esce con probabilità pari a

$$\frac{\text{card}(A \cap B)}{\text{card } B} = \frac{2}{5}.$$

Analizziamo un po' più in dettaglio il ragionamento fatto, per cercare di ricavarne il senso generale.

Possiamo scrivere

$$\frac{\text{card}(A \cap B)}{\text{card } B} = \frac{\text{card}(A \cap B)/\text{card } \Omega}{\text{card } B/\text{card } \Omega} = \frac{P(A \cap B)}{P(B)}.$$

In altre parole la nostra valutazione del verificarsi di  $A$  *sapendo che si è verificato*  $B$  è

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

Questo esempio giustifica la definizione seguente:

**Definizione 1.5.2.** Siano  $A, B$  due eventi, con  $P(B) > 0$ . Si chiama *probabilità condizionale di  $A$ , dato  $B$*  (oppure “sapendo che si è verificato  $B$ ”) la quantità

$$P(A|B) := \frac{P(A \cap B)}{P(B)} \quad (1.5.3)$$

(per chi non l'avesse mai vista, la scrittura  $:=$  significa semplicemente che con la parte destra della formula si definisce il simbolo nuovo scritto alla sua sinistra; in questo caso si tratta del simbolo  $P(A|B)$ , che viene definito dalla frazione al secondo membro).

**Osservazione 1.5.4.** Per quanto ovvio, si sottolinea il fatto che la scrittura  $P(A|B)$  non rappresenta la probabilità dell'evento “ $A|B$ ” (che non esiste!), ma solo un modo per indicare il rapporto che compare al secondo membro della (1.5.3).

Il nome di “probabilità” dato alla quantità  $P(A|B)$  è giustificato dalla seguente proposizione, di cui si lascia per esercizio al lettore la semplice dimostrazione:

**Proposizione 1.5.5.** *Sia  $B$  un evento, con  $P(B) > 0$ . L'applicazione  $Q : \mathcal{A} \rightarrow \mathbb{R}^+$  definita da  $Q(A) = P(A|B)$  è una probabilità (nel senso della Definizione 1.2.6).*

Può capitare che l'informazione supplementare “si è verificato  $B$ ” non cambi la nostra valutazione della probabilità del verificarsi di  $A$ .

**Esempio 1.5.6.** Si lancia due volte una moneta equilibrata. Poniamo

$$\begin{aligned} A &= \{\text{esce testa al secondo lancio}\}, \\ B &= \{\text{esce testa al primo lancio}\}. \end{aligned}$$

È immediato verificare che

$$P(A|B) = P(A) = \frac{1}{2}$$

(costruire lo spazio  $(\Omega, \mathcal{A}, P)$  che modella l'esperimento e svolgere per esercizio i calcoli).

Diamo dunque la seguente definizione (che sarà però provvisoria):

**Definizione 1.5.7.** Siano  $A, B$  due eventi, con  $P(B) > 0$ .  $A$  e  $B$  si dicono *indipendenti (tra loro)* se

$$P(A|B) = P(A). \quad (1.5.8)$$

Usando la (1.5.3), la (1.5.8) si può scrivere nella forma

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

o anche, moltiplicando per  $P(B)$ ,

$$P(A \cap B) = P(A)P(B). \quad (1.5.9)$$

La (1.5.9) è chiaramente simmetrica in  $A$  e  $B$  (del resto lo è anche l'espressione  *$A$  e  $B$  sono indipendenti (tra loro)* usata nella Definizione 1.5.7, mentre non è affatto evidente che lo sia anche la (1.5.8)). Inoltre, mentre la (1.5.8) perde di significato se  $B$  è un evento di probabilità nulla, la (1.5.9) ha ancora senso e, ciò che più importa, è ancora vera (verifica per esercizio). Essa risulta quindi più comoda come definizione di indipendenza; dunque sostituiremo la definizione provvisoria con la seguente:

**Definizione 1.5.10.** Due eventi  $A$  e  $B$  si dicono *indipendenti (tra loro)* se vale la (1.5.9).

Sarà necessario estendere la definizione di indipendenza ad una famiglia qualsiasi di eventi. Consideriamo per iniziare il caso di tre eventi  $A$ ,  $B$  e  $C$ . Per analogia con il caso di due eventi, potremmo pensare che sia corretto dire che essi sono indipendenti se

$$P(A \cap B \cap C) = P(A)P(B)P(C). \quad (1.5.11)$$

Purtroppo questa condizione non garantisce che gli eventi siano anche “due a due” indipendenti, come ci dice invece l’intuizione (se ad esempio  $A$ ,  $B$  e  $C$  sono tre eventi, con  $C = \emptyset$ , la relazione (1.5.11) è soddisfatta anche nel caso in cui  $A$  e  $B$  non siano indipendenti). La definizione ragionevole è allora la seguente:

**Definizione 1.5.12.** Tre eventi  $A$ ,  $B$  e  $C$  si dicono *indipendenti* se valgono *tutte* le condizioni seguenti

- (i)  $P(A \cap B \cap C) = P(A)P(B)P(C);$
- (ii)  $P(A \cap B) = P(A)P(B);$
- (iii)  $P(B \cap C) = P(B)P(C);$
- (iv)  $P(A \cap C) = P(A)P(C).$

La definizione precedente si generalizza in modo ovvio ad una famiglia qualsiasi di eventi: essi si diranno indipendenti se per ogni sottofamiglia finita la probabilità dell’intersezione è uguale al prodotto delle singole probabilità; in termini precisi:

**Definizione 1.5.13.** Gli eventi  $(A_i)_{i \in I}$  si dicono *indipendenti (fra loro)* se per ogni  $k \in \mathbb{N}$  e per ogni successione finita  $i_1, i_2, \dots, i_k$  di indici di  $I$ , si ha

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

**Esercizio 1.5.14.** Siano  $A$  e  $B$  due eventi. Mostrare che le affermazioni seguenti sono equivalenti:

- (a)  $A$  e  $B$  sono indipendenti;
- (b)  $A$  e  $B^c$  sono indipendenti;
- (c)  $A^c$  e  $B$  sono indipendenti;
- (d)  $A^c$  e  $B^c$  sono indipendenti.

Come si generalizza l’esercizio precedente al caso di una famiglia di  $n > 2$  eventi?

**Osservazione 1.5.15.** Per quanto riguarda i calcoli, l’indipendenza consiste sostanzialmente nella possibilità di fattorizzare la probabilità dell’intersezione di due eventi  $A$  e  $B$  nel prodotto delle probabilità di  $A$  e di  $B$ ; questo, dal

punto di vista computazionale, si traduce nella possibilità di effettuare calcoli che altrimenti sarebbero complicati o addirittura non fattibili; tuttavia questo vantaggio sarebbe poco apprezzabile se nella pratica situazioni di indipendenza non esistessero o si presentassero solo raramente; per fortuna le cose non vanno così (esempi saranno dati già nel paragrafo che segue), e comunque, negli esperimenti controllati, lo sperimentatore ha spesso la possibilità di mettersi volontariamente in condizioni di lavorare con eventi indipendenti. Si capisce bene quindi come l'indipendenza sia uno dei concetti più importanti del calcolo delle probabilità e della statistica.

L'ipotesi di indipendenza consente di costruire un particolare tipo di spazio di probabilità, utilizzabile nella pratica in una varietà di situazioni solo apparentemente differenti fra loro (cioè identificabili tra loro dal punto di vista della modellizzazione matematica). Questa costruzione sarà l'argomento del prossimo paragrafo.

## 1.6 Lo schema delle prove indipendenti

Vogliamo costruire uno spazio di probabilità per modellizzare l'esperimento che segue:

**Esempio 1.6.1.** Si lancia  $n$  volte una moneta, per la quale è  $p$  ( $0 < p < 1$ ) la probabilità di ottenere “testa” in un generico lancio (la moneta non è necessariamente equilibrata, o, ciò che è lo stesso,  $p$  non è necessariamente uguale a  $1/2$ ). L'esperimento si svolge in modo che il risultato di un singolo lancio non influenza e non è influenzato dagli altri, o, come si dice brevemente, in condizioni di indipendenza. Conveniamo, come al solito, di indicare il risultato “testa” (risp. “croce”) con il simbolo 1 (risp. 0).

Come abbiamo già fatto nell'Esempio 1.3.6, porremo

$$\Omega = \{0, 1\}^n.$$

In questo caso, tuttavia, lo spazio da costruire non sarà evidentemente uniforme (tranne che nel caso, già discusso in (1.3.6), in cui  $p = 1/2$ ). Quale sarà allora la probabilità ragionevole da assegnare ad un generico sottoinsieme di  $\Omega$ ? Grazie all'uguaglianza

$$P(A) = \sum_{\omega \in A} P(\omega)$$

già vista in (1.4.2), basterà farlo per gli eventi elementari

$$\omega = (\omega_1, \omega_2, \dots, \omega_n)$$

(con  $\omega_i \in \{0, 1\}$ ). Per chiarezza tratteremo dapprima un caso numerico.

**Esempio 1.6.2.** Poniamo  $n = 5$ , e consideriamo la successione di risultati  $\omega = (1, 1, 0, 1, 0)$ .

Per  $i = 1, \dots, 5$  consideriamo gli eventi

$$A_i = \{\text{esce 1 al lancio } i\text{-esimo}\}.$$

Evidentemente

$$\omega = A_1 \cap A_2 \cap A_3^c \cap A_4 \cap A_5^c.$$

Poiché l'esperimento si svolge in condizioni di indipendenza, è naturale ipotizzare l'indipendenza degli eventi  $A_i$ , e dunque per l'Esercizio 1.5.14 (generalizzato), avremo

$$P(\omega) = P(A_1 \cap A_2 \cap A_3^c \cap A_4 \cap A_5^c) = P(A_1)P(A_2)P(A_3^c)P(A_4)P(A_5^c).$$

D'altra parte, poiché 1 (risp. 0) esce con probabilità  $p$  (risp.  $1-p$ ), supporremo che, per ogni  $i = 1, \dots, 5$  risulti  $P(A_i) = p$ ,  $P(A_i^c) = 1-p$ .

Si conclude dunque che

$$P(\omega) = p^3(1-p)^2.$$

È facile capire che i numeri  $p$  e  $1-p$  compaiono con esponenti 3 e 2 rispettivamente perché, nella sequenza  $\omega$  considerata, 3 (risp. 2) è il numero di simboli uguali a 1 (risp. 0).

Il ragionamento dell'esempio è utilizzabile, evidentemente, anche nel caso generale; sia dunque  $\omega = (\omega_1, \omega_2, \dots, \omega_n)$  (con  $\omega_i \in \{0, 1\}$ ) un elemento di  $\Omega = \{0, 1\}^n$ ; osserviamo che le quantità

$$\sum_{i=1}^n \omega_i, \quad n - \sum_{i=1}^n \omega_i$$

non sono altro che il numero di simboli uguali a 1 e il numero di simboli uguali a 0 presenti nella sequenza  $\omega$ . Pertanto avremo

$$P(\omega) = p^{(\sum_{i=1}^n \omega_i)}(1-p)^{(n-\sum_{i=1}^n \omega_i)}.$$

L'importanza e l'uso del modello qui costruito vanno ben al di là della situazione dell'Esempio 1.6.1.

Supponiamo di dover eseguire  $n$  volte in condizioni di indipendenza un certo esperimento che produce due soli possibili risultati, che chiameremo convenzionalmente *successo* e *insuccesso* e che indicheremo rispettivamente con i simboli 1 e 0. Supponiamo inoltre di sapere che il *successo* (risp. l'*insuccesso*) si presenta con probabilità  $p$  (risp.  $1-p$ ) nella generica ripetizione dell'esperimento. Diamo alcuni esempi tratti da situazioni pratiche.

**Esempio 1.6.3.** In un laboratorio si sperimenta un nuovo farmaco somministrandolo a  $n$  cavie. È noto (per esperienza precedente) che una generica cavia presenta reazione al farmaco con probabilità  $p$  (e, al contrario, non presenta



alcuna reazione con probabilità  $1 - p$ ); supporremo inoltre che tutto avvenga in modo che il fatto che una delle cavie presenti o meno reazione non abbia influenza su quello che accade alle altre (ad esempio, si può immaginare che le cavie siano state scelte in modo che non ci sia alcun legame di parentela tra di loro ...). Ci interessa osservare quante e quali cavie reagiscano al farmaco, e con quali probabilità. Con il simbolo 1 (risp. 0) indichiamo il fatto che la generica cavia reagisca (risp. non reagisca) al farmaco. Allora, ad esempio, la sequenza  $\omega$  composta di soli simboli 0 si legge dicendo “nessuna cavia ha reagito al farmaco”, e ciò accade con probabilità  $(1 - p)^n$ . Naturalmente sarebbe stato possibile indicare con 1 il fatto che la generica cavia *non* presenta reazione (e dunque con 0 il fatto che la reazione c’è stata). In questo caso la sequenza di soli simboli 0 avrebbe significato che *tutte* le cavie hanno reagito, e la probabilità sarebbe stata uguale a  $p^n$ . In altre parole, i termini *successo* e *insuccesso* sono interscambiabili e non presentano nessuna connotazione positiva o negativa.

Negli esempi che seguono si lascia al lettore la cura di discutere il modello nei dettagli e di rispondere alle domande.

**Esempio 1.6.4.** In un’operazione di polizia, una pattuglia deve catturare  $n$  malviventi. La probabilità che un malvivente sia (risp. non sia) catturato è  $p$  (risp.  $1 - p$ ). Supponendo valida l’ipotesi di indipendenza, quanto vale la probabilità che al termine dell’operazione al più un malvivente sia riuscito a sfuggire alla cattura? Immaginare una situazione reale di indipendenza.

**Esempio 1.6.5.** In un appartamento ciascuna delle  $n$  stanze è illuminata da una lampadina. In una sera di temporale, un fulmine si abbatte nelle vicinanze. È noto che esso può fulminare una generica lampadina con probabilità  $p$ . Quanto vale la probabilità che l’appartamento resti al buio? (ipotizzare sempre l’indipendenza; cosa significa in questo esempio? è un’ipotesi realistica?).

**Osservazione 1.6.6.** Dagli Esempi 1.6.4 e 1.6.5 dovrebbe risultare chiaro che il termine “esperimento” è del tutto convenzionale.

Allo spazio di probabilità che abbiamo costruito in questo paragrafo si dà generalmente il nome di *schema delle prove indipendenti* (o anche *schema delle prove ripetute* o *schema di Bernoulli*).

## 1.7 La formula di Bayes

**Esempio 1.7.1.** Un segnale proviene il 40% delle volte da un’apparecchiatura  $A_1$  e per il restante 60% delle volte da una seconda apparecchiatura  $A_2$ . Esso può essere di due tipi: “lungo” ( $L$ ) oppure “breve” ( $B$ ).

È noto che  $A_1$  (risp.  $A_2$ ) trasmette un segnale di tipo  $L$  il 52% (risp. 37%) delle volte. In un certo istante viene ricevuto un segnale di tipo  $B$ ; qual è la probabilità che esso provenga da  $A_1$ ?

Consideriamo gli eventi

$$A_1 = \{\text{il segnale proviene da } A_1\};$$

$$A_2 = \{\text{il segnale proviene da } A_2\};$$

$$L = \{\text{il segnale risulta } L\};$$

$$B = \{\text{il segnale risulta } B\}.$$

Le informazioni in nostro possesso possono essere formalizzate come segue:

$$P(A_1) = 0.4; \quad P(A_2) = 0.6; \quad P(L|A_1) = 0.52; \quad P(L|A_2) = 0.37.$$

Di conseguenza si ha anche

$$P(B|A_1) = 1 - P(L|A_1) = 0.48; \quad P(B|A_2) = 1 - P(L|A_2) = 0.63.$$

L'esercizio chiede la probabilità condizionale  $P(A_1|B)$ . Il metodo che useremo per il calcolo sarà poi generalizzato per ottenere una formula molto importante (soprattutto in statistica), detta *formula di Bayes*.

Per definizione si ha

$$P(A_1|B) = \frac{P(A_1 \cap B)}{P(B)}.$$

Calcoleremo allora il numeratore e il denominatore della frazione precedente. Cominciamo dal calcolo di  $P(A_1 \cap B)$ . Si può scrivere, ancora utilizzando (in senso inverso!) la definizione di probabilità condizionale

$$P(A_1 \cap B) = P(B|A_1)P(A_1) = 0.48 \times 0.4 = 0.192.$$

Passiamo ora al calcolo di  $P(B)$ . Si può scrivere

$$B = B \cap \Omega = B \cap (A_1 \cup A_2) = (B \cap A_1) \cup (B \cap A_2);$$

dato che  $A_1$  e  $A_2$  sono due eventi tra loro disgiunti, anche gli eventi  $(B \cap A_1)$  e  $(B \cap A_2)$  lo saranno; dunque, per la proprietà di additività della probabilità, avremo

$$P(B) = P(B \cap A_1) + P(B \cap A_2).$$

$P(B \cap A_1)$  è già stata calcolata sopra e vale 0.192. Nello stesso modo si ha

$$P(A_2 \cap B) = P(B|A_2)P(A_2) = 0.63 \times 0.6 = 0.378.$$

Si conclude allora che

$$P(B) = 0.192 + 0.378 = 0.57,$$

ed infine

$$P(A_1|B) = \frac{P(A_1 \cap B)}{P(B)} = \frac{0.192}{0.57} = 0.3368.$$

Avendo in mente l'esercizio appena svolto, passiamo ad enunciare e dimostrare la *formula di Bayes*, di cui esso non è che un'applicazione. Premettiamo una definizione.

**Definizione 1.7.2.** Siano  $(\Omega, \mathcal{A}, P)$  uno spazio di probabilità fissato, e  $A_1, \dots, A_n$   $n$  eventi in  $\mathcal{A}$ . Si dice che  $A_1, \dots, A_n$  è una *partizione* di  $(\Omega, \mathcal{A}, P)$  se valgono entrambe le proprietà seguenti:

- (i)  $\bigcup_{i=1}^n A_i = \Omega$ ;
- (ii)  $A_i \cap A_j = \emptyset \quad \forall (i, j) \text{ con } i \neq j$ .

**Osservazione 1.7.3.** Gli eventi di una partizione vanno pensati come una famiglia di  $n$  eventualità che possono prodursi durante lo svolgersi dell'esperimento; esse esauriscono tutte le possibilità (sicuramente una di esse si verifica) e si escludono a vicenda (due di esse non possono verificarsi contemporaneamente). Nell'esercizio iniziale, ad esempio, si trattava di  $A_1$  e  $A_2$ : il segnale proviene sicuramente da una delle due apparecchiature, ma non da tutte e due contemporaneamente.

Ciò premesso, si ha il seguente:

**Teorema 1.7.4 (formula di Bayes).** Sia  $(\Omega, \mathcal{A}, P)$  uno spazio di probabilità. Sia  $A_1, \dots, A_n$  una partizione di  $\Omega$ , tale che  $P(A_i) > 0$  per ogni  $i = 1, \dots, n$ . Sia infine  $B$  un evento con  $P(B) > 0$ . Allora per ogni  $k = 1, \dots, n$  risulta

$$P(A_k|B) = \frac{P(B|A_k)P(A_k)}{\sum_{i=1}^n P(B|A_i)P(A_i)}.$$

**Osservazione 1.7.5.** La formula di Bayes è nota anche con il nome di formula delle *probabilità delle cause*: interpretando gli eventi della partizione come  $n$  possibili "cause" per il verificarsi di  $B$ , la formula permette il calcolo della probabilità che, sapendo che  $B$  si è verificato, la causa sia stata  $A_k$ . Per questo motivo viene spesso usato per calcolare le probabilità cosiddette *a posteriori* (*posterior probabilities* in inglese): per esempio, avendo osservato determinati sintomi in un paziente, si può utilizzare la formula per verificare la correttezza della diagnosi proposta.

*Dimostrazione (del teorema).* Come nell'esercizio, risulta

$$P(A_k \cap B) = P(B|A_k)P(A_k). \quad (1.7.6)$$

Inoltre si può scrivere

$$B = B \cap \Omega = B \cap (\cup_{i=1}^n A_i) = \cup_{i=1}^n (B \cap A_i).$$

Poiché gli eventi  $(B \cap A_i)$  sono tra loro due a due disgiunti (perché?), per l'assioma di additività della probabilità si ottiene

$$P(B) = \sum_{i=1}^n P(B \cap A_i). \quad (1.7.7)$$

La formula (1.7.7) va sotto il nome di *formula della partizione dell'evento certo*; ad essa abbiamo accennato all'inizio del paragrafo 1.3. Essa può essere utilizzata per calcolare la probabilità di  $B$  quando si sappia calcolare ciascuna delle probabilità che  $B$  si verifichi *contemporaneamente* ad una delle eventualità  $A_i$  (cioè che si verifichi l'evento  $B \cap A_i$ ).

Come abbiamo fatto in precedenza (formula (1.7.6)) si può scrivere

$$P(A_i \cap B) = P(B|A_i)P(A_i),$$

e quindi la (1.7.7) diventa

$$P(B) = \sum_{i=1}^n P(B|A_i)P(A_i). \quad (1.7.8)$$

Dalle formule (1.7.6) e (1.7.8) si ricava allora

$$P(A_k|B) = \frac{P(A_k \cap B)}{P(B)} = \frac{P(B|A_k)P(A_k)}{\sum_{i=1}^n P(B|A_i)P(A_i)},$$

ovvero la formula di Bayes che cercavamo. □

## Variabili aleatorie

---

### 2.1 Definizioni e prime proprietà

*Motivazione.* Molto spesso, l'esperimento che viene effettuato produce una quantità numerica  $X$  che deve essere studiata; anzi, si può dire che in moltissime situazioni è proprio la quantità  $X$  l'oggetto interessante per chi esegue l'esperimento (più che l'esperimento in sé e per sé).

**Esempio 2.1.1.** Un programma informatico produce sequenze aleatorie di lunghezza 5, composte dai simboli 0 e 1. Come sappiamo dal paragrafo 1.6, questo esperimento si modella con uno schema di  $n = 5$  prove indipendenti ( $p$  = probabilità di ottenere un simbolo “1”). È noto che il tempo impiegato per produrre un simbolo “1” (risp “0”) è di 2 (risp. 3)  $\mu\text{sec}$ . Il tecnico che utilizza il programma è interessato a sapere quanto tempo è necessario per produrre una sequenza; in altre parole, le domande a cui vorrebbe poter rispondere sono del tipo “qual è la probabilità che il programma impieghi più di  $x$   $\mu\text{sec}$ ”, oppure “qual è la probabilità che il programma impieghi un tempo compreso fra  $x$   $\mu\text{sec}$  e  $y$   $\mu\text{sec}$ ”, e così via, (piuttosto che domande come “qual è la probabilità che il programma produca (per esempio) la sequenza (1, 0, 0, 1, 1)”). Cioè in questa situazione siamo interessati a ottenere informazioni “di tipo probabilistico” (sarà chiaro fra poco il significato di questa espressione) sulla quantità  $X$  = tempo (in  $\mu\text{sec}$ ) impiegato dal programma per produrre una sequenza.

Prima di passare alla definizione matematica di variabile aleatoria, sviluppiamo ancora un po' l'esempio precedente. È chiaro che  $X$  dipende dalla particolare sequenza prodotta dal programma: ad esempio, se la sequenza è  $\omega_1 = (1, 0, 0, 1, 1)$  avremo  $X(\omega_1) = 12$ ; se è  $\omega_2 = (1, 1, 0, 1, 1)$  avremo  $X(\omega_2) = 11$ ; in altri termini,  $X$  è una funzione del risultato dell'esperimento: se, al solito, indichiamo con  $\Omega$  l'insieme di tutti i possibili risultati ( $\Omega = \{0, 1\}^5$  in questo caso),  $X$  sarà una funzione definita su  $\Omega$  e a valori in  $\mathbb{R}$ .

Cerchiamo ora di formalizzare domande del tipo “qual è la probabilità che il programma impieghi un tempo compreso fra  $x$   $\mu\text{sec}$  e  $y$   $\mu\text{sec}$ ”. Questa

domanda equivale a cercare la probabilità del sottoinsieme costituito dai possibili risultati  $\omega$  dell'esperimento per i quali risulti  $x < X(\omega) \leq y$ , ovvero la probabilità di

$$\{\omega \in \Omega : x < X(\omega) \leq y\}.$$

Pertanto una richiesta ragionevole sarà che i sottoinsiemi di  $\Omega$  del tipo precedente siano eventi (altrimenti su di essi non sarebbe definita la funzione *probabilità*).

Nell'esempio che stiamo trattando questo è certamente vero, dato che negli spazi a cardinalità finita ogni sottoinsieme è un evento. La richiesta diventa però indispensabile in situazioni più complicate, poiché in generale non è possibile considerare evento *ogni* sottoinsieme di  $\Omega$  (il motivo è di carattere matematico, e non ne tratteremo in questa sede perché va oltre i nostri scopi).

È facile far vedere che, se si suppone che siano eventi i sottoinsiemi

$$\{\omega \in \Omega : x < X(\omega) \leq y\},$$

allora sono eventi tutti i sottoinsiemi del tipo

$$\{\omega \in \Omega : X(\omega) \in I\},$$

dove  $I$  è un qualsiasi intervallo (limitato o no, chiuso o no) di  $\mathbb{R}$  (ved. l'Esercizio 2.1.5).

Osserviamo che è del tipo considerato anche ogni sottoinsieme del tipo

$$\{\omega \in \Omega : X(\omega) = x\},$$

dove  $x \in \mathbb{R}$  (basta prendere  $I = \{x\}$ , intervallo chiuso ridotto al solo punto  $x$ ).

**Osservazione 2.1.2.** Nelle ipotesi precedenti, più in generale si può far vedere che  $\{X \in I\}$  è un evento per una famiglia più vasta di sottoinsiemi  $I$  di  $\mathbb{R}$  (cioè non solo per gli intervalli). Tali insiemi sono detti *misurabili*. È opportuno sapere che *non tutti* i sottoinsiemi di  $\mathbb{R}$  sono misurabili; anche questo, comunque, è un problema di natura strettamente matematica e non vi insisteremo oltre.

*Notazione.* D'ora in avanti l'evento  $\{\omega \in \Omega : X(\omega) \in I\}$  sarà indicato con la notazione abbreviata  $\{X \in I\}$ . Allo stesso modo scriveremo  $\{x < X \leq y\}$ ,  $\{X \leq y\}$  al posto di  $\{\omega \in \Omega : x < X(\omega) \leq y\}$ ,  $\{\omega \in \Omega : X(\omega) \leq y\}$  e così via. Per evitare possibili equivoci, insistiamo sul fatto che quando si scrive, per esempio,  $\{1 \leq X \leq 2\}$ , non si intende l'intervallo  $[1, 2]$  (sottoinsieme di  $\mathbb{R}$ !) ma un *evento* (sottoinsieme di  $\Omega$ !).

Siamo ora in grado di dare la definizione matematica di variabile aleatoria:

**Definizione 2.1.3.** Sia  $(\Omega, \mathcal{A}, P)$  uno spazio di probabilità. Una *variabile aleatoria* (*random variable* in inglese) è una funzione  $X : \Omega \rightarrow \mathbb{R}$  tale che, per ogni  $t \in \mathbb{R}$  risulti  $\{X \leq t\} \in \mathcal{A}$ . Abbreviazione: *v.a.* (*r.v.* in inglese).

**Osservazione 2.1.4.** La precedente definizione è giustificata dal fatto che, affinché *tutti* i sottoinsiemi del tipo  $\{X \in I\}$ , con  $I$  insieme misurabile, siano eventi, è necessario e sufficiente supporre che lo siano i sottoinsiemi  $\{X \leq y\}$  (cioè quelli per i quali  $I$  è una semiretta sinistra chiusa del tipo  $(-\infty, y]$ ). La cosa è di semplice verifica ad esempio se  $I$  è un intervallo  $(a, b]$  (Esercizio 2.1.5). Più delicata è la dimostrazione per gli intervalli ridotti ad un solo punto ( $I = \{x\}$ ). Essa fa uso della proprietà di stabilità rispetto alle intersezioni *numerabili* della  $\sigma$ -algebra  $\mathcal{A}$ , e non sarà svolta in questa sede.

**Esercizio 2.1.5.** Supponiamo noto il fatto che ogni insieme del tipo  $\{X \in I\}$  è un evento quando  $I = (-\infty, b]$  e quando  $I = \{x\}$  (per ogni  $b$  e per ogni  $x$ ). Mostrare che la stessa cosa è vera per  $I$  del tipo  $(a, b]$ ,  $[a, b]$ ,  $[a, b]$ ,  $(-\infty, b)$ ,  $(-\infty, b]$ ,  $[a, +\infty)$ ,  $(a, +\infty)$ .

*Notazione.* Per brevità scriveremo semplicemente  $P(X \in I)$  anziché (come sarebbe corretto)  $P(\{X \in I\})$ .

**Definizione 2.1.6.** Sia  $X$  una v. a. su  $(\Omega, \mathcal{A}, P)$ . Si chiama *legge di X* l'applicazione

$$I \mapsto P(X \in I),$$

(definita sulla classe degli insiemi misurabili di  $\mathbb{R}$  e a valori in  $[0, 1]$ ).

La legge di una v. a.  $X$  va pensata come una “fotografia” delle varie probabilità assegnate a *tutti* gli eventi del tipo interessante (cioè  $\{X \in I\}$ ), e se ne intuisce dunque l'importanza. D'altra parte è evidente che nella pratica sarà impossibile “elencare” *tutte* le probabilità  $P(X \in I)$ , per ciascun  $I$ . Si capisce dunque la necessità di procurarsi alcuni strumenti non troppo complicati, che ci consentano di “risparmiare” calcoli.

Questo è lo scopo di ciò che faremo nel seguito. Il primo concetto che introduciamo è quello di funzione di ripartizione di  $X$ , che, come vedremo, è valido per qualunque tipo di variabile aleatoria. Nel paragrafo 2.3 e nel capitolo 4 studieremo poi due tipi importanti di variabili aleatorie (quelle discrete e quelle assolutamente continue) per le quali esistono ulteriori strumenti di calcolo.

## 2.2 Funzione di ripartizione di una variabile aleatoria: definizione e proprietà

**Definizione 2.2.1.** Sullo spazio di probabilità  $(\Omega, \mathcal{A}, P)$  sia  $X$  una v. a. assegnata. Si chiama *funzione di ripartizione* (o di *distribuzione*; *distribution function* in inglese) (f.d.r.) di  $X$  la funzione  $F: \mathbb{R} \rightarrow \mathbb{R}$  definita da

$$F(t) = P(X \leq t).$$

Quando saranno in gioco più variabili aleatorie  $X, Y, \dots$ , le rispettive funzioni di ripartizione verranno indicate con  $F_X, F_Y, \dots$ .

**Teorema 2.2.2 (Proprietà caratteristiche di una funzione di ripartizione).** (a) *Siano  $X$  una v. a.,  $F$  la sua funzione di ripartizione. Valgono le seguenti proprietà:*

- (i)  $0 \leq F(t) \leq 1, \forall t \in \mathbb{R}$ ;
- (ii)  $F$  è non decrescente, cioè, per ogni coppia  $x, y$  di numeri reali, con  $x \leq y$  risulta

$$F(x) \leq F(y);$$

- (iii) *valgono le relazioni*

$$\lim_{t \rightarrow -\infty} F(t) = 0; \quad \lim_{t \rightarrow +\infty} F(t) = 1;$$

- (iv)  $F$  è continua a destra, cioè per ogni  $x \in \mathbb{R}$  si ha

$$F(x^+) := \lim_{t \rightarrow x^+} F(t) = F(x).$$

(Per quanto evidente, si sottolinea il fatto che con la scrittura  $F(x^+)$  non si intende il valore di  $F$  nel punto (inesistente!)  $x^+$ , ma il limite da destra verso  $x$ .)

- (b) *Viceversa, sia  $F: \mathbb{R} \rightarrow \mathbb{R}$  una funzione, che gode di tutte le proprietà (i), (ii), (iii) (iv) elencate sopra. Allora esiste una variabile aleatoria  $X$ , definita su un opportuno spazio di probabilità, che ammette  $F$  come sua funzione di ripartizione.*

**Osservazione 2.2.3.** Il teorema precedente è utile quando si debba decidere se un'assegnata funzione  $F$  è una funzione di ripartizione: lo sarà se gode di tutte le proprietà (i)-(iv), non lo sarà se anche una sola di esse non è verificata.

**Esempio 2.2.4.** La funzione definita da

$$F(t) = \begin{cases} 0 & x \leq 1 \\ 1 & x > 1 \end{cases}$$

non è una funzione di ripartizione perché non è continua a destra.



Invece la funzione

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

verifica tutte le proprietà (i)–(iv), e dunque è una funzione di ripartizione.

*Dimostrazione.* Del Teorema 2.2.2 dimostreremo solo la parte (a). La (i) è ovvia, dato che i valori assunti da  $F$  sono probabilità di particolari eventi. Per la (ii) basta osservare che, se  $x \leq y$  risulta

$$\{X \leq x\} \subseteq \{X \leq y\},$$

e applicare poi la proprietà di isotonia della probabilità (cioè  $A \subseteq B \Rightarrow P(A) \leq P(B)$ ). La dimostrazione della (iii) richiede la proprietà di passaggio al limite della probabilità sulle successioni monotone di eventi (proprietà (vi) del paragrafo (1.3)). Dimostreremo la prima relazione (la seconda è analoga). Poiché  $F$  è una funzione limitata e monotona (per i punti (i) e (ii)), basterà verificare che

$$\lim_{\substack{n \rightarrow \infty \\ n \in \mathbb{N}}} F(n) = 1. \quad (2.2.5)$$

(In generale il fatto che una funzione abbia limite sulla successione dei numeri interi *non* garantisce che essa abbia limite per  $x \rightarrow \infty, x \in \mathbb{R}$ : basta pensare alla funzione  $F(x) = \sin(\pi x)$ . Tuttavia, se  $F$  è monotona e limitata, allora l'esistenza del limite per  $x \rightarrow \infty, x \in \mathbb{R}$  è garantita da un noto teorema di analisi, e dunque per trovarne il valore è sufficiente calcolarlo su una qualsiasi successione di numeri tendente a  $+\infty$ , per esempio quella dei numeri interi.)

Per verificare la (2.2.5), osserviamo che  $F(n)$  è la probabilità dell'evento  $A_n = \{X \leq n\}$ . La successione di eventi  $(A_n)_{n \in \mathbb{N}}$  è crescente ed inoltre  $\bigcup_{n \in \mathbb{N}} A_n = \Omega$  (verifiche per esercizio). Dunque, per la proprietà di passaggio al limite sulle successioni crescenti di eventi sopra ricordata, si ha

$$\lim_{n \rightarrow \infty} F(n) = \lim_{n \rightarrow \infty} P(A_n) = P(\bigcup_{n \in \mathbb{N}} A_n) = P(\Omega) = 1.$$

La dimostrazione di (iv) è analoga; basta verificare che

$$\lim_{n \rightarrow \infty} F(x + 1/n) = F(x).$$

Si lasciano i dettagli al lettore, dopo aver osservato che  $F(x + 1/n) = P(X \leq x + 1/n)$ . □

Mostriamo ora come si può usare la funzione di ripartizione per il calcolo della legge della variabile aleatoria. Prima di tutto enunciamo (senza dimostrazione) il seguente risultato:

**Teorema 2.2.6.** *La funzione di ripartizione individua (univocamente) la legge della variabile aleatoria. In altre parole se due variabili aleatorie hanno la stessa funzione di ripartizione, allora hanno la stessa legge.*

Vediamo in concreto il calcolo della legge di  $X$  a partire dalla sua funzione di ripartizione  $F$  in alcuni casi particolari (cioè il calcolo della quantità  $P(X \in I)$  per qualche tipo particolare di insieme misurabile  $I$ ):

(i)  $I = (a, b]$ . Si ha (verificare per esercizio la prima delle uguaglianze sottostanti):

$$P(a < X \leq b) = P(X \leq b) - P(X \leq a) = F(b) - F(a);$$

(ii)  $I = (-\infty, b)$ . In questo caso si fa uso della proprietà di passaggio al limite della probabilità sulle successioni monotone di eventi. Dato che qualche calcolo analogo è già stato svolto precedentemente, si lasciano i dettagli al lettore. Risulta

$$P(X < b) = \lim_{n \rightarrow \infty} P(X \leq b - 1/n) = \lim_{t \rightarrow b^-} F(t) = F(b^-);$$

come al solito, con il simbolo  $F(b^-)$  si è indicato il limite verso  $b$  da sinistra, (e non il valore nel punto inesistente  $b^-$ ), e il calcolo precedente dice che tale limite esiste e vale  $P(X < b)$ .

Nel seguito si lascia per esercizio al lettore il compito di giustificare i vari passaggi in modo completo.

(iii)  $I = \{x\}$ . Risulta

$$P(X = x) = P(X \leq x) - P(X < x) = F(x) - F(x^-);$$

(iv)  $I = [a, b)$ . In questo caso

$$P(a \leq X < b) = P(X < b) - P(X < a) = F(b^-) - F(a^-).$$

**Esempio 2.2.7.** Trovare formule analoghe alle precedenti per i casi

- (v)  $I = (a, b)$ ;
- (vi)  $I = [a, b]$ ;
- (vii)  $I = (a, +\infty)$ ;
- (viii)  $I = [a, +\infty)$ .

La formula del punto (iii) è interessante. Ricordando che ogni funzione di ripartizione è continua a destra, essa può essere scritta anche nella forma equivalente

$$P(X = x) = F(x^+) - F(x^-) \quad (2.2.8)$$

ed interpretata dicendo che la probabilità che  $X$  assuma il valore  $x$  è esattamente uguale al “salto” che la funzione di ripartizione di  $X$  presenta nel punto di ascissa  $x$ . Questa osservazione, intanto, è comoda in pratica quando si abbia a disposizione il grafico di  $F$ : i valori numerici che  $X$  assume con probabilità *strettamente* positiva sono niente altro che le ascisse dei “salti” (cioè i punti di discontinuità) nel grafico di  $F$ .

Invece i punti in cui  $F$  è continua sono tutti e soli i valori assunti dalla variabile  $X$  con probabilità nulla: questo segue immediatamente dalla relazione (2.2.8), osservando che  $F$  è continua in  $x$  se e soltanto se risulta  $F(x^+) - F(x^-) = 0$ .

Ciò giustifica la seguente:

**Definizione 2.2.9.** La variabile aleatoria  $X$  è detta *continua* se la sua funzione di ripartizione  $F$  è continua, o, ciò che è lo stesso, se  $X$  assume con probabilità nulla *ogni* valore reale  $x$  (cioè  $P(X = x) = 0 \forall x \in \mathbb{R}$ ).

**Esempio 2.2.10.** (a) La funzione di ripartizione dell'esempio (2.2.4) è continua.

(b) La funzione di ripartizione definita da

$$F(t) = \begin{cases} 0 & x < 1 \\ 1 & x \geq 1 \end{cases}$$

non è continua.

In particolare, la variabile aleatoria  $X$  avente  $F$  come f.d.r. assume il solo valore 1 con probabilità uguale a 1 (il “salto” della funzione di ripartizione).

## 2.3 Variabili aleatorie discrete

**Definizione 2.3.1.** Sia  $X$  una variabile aleatoria definita sullo spazio  $(\Omega, \mathcal{A}, P)$ .  $X$  si dice *discreta* se la sua immagine  $\mathcal{E}$  è un sottoinsieme finito o numerabile di  $\mathbb{R}$ , senza punti di accumulazione.

Scriveremo anche

$$\mathcal{E} = \{x_1, x_2, x_3 \dots\}.$$

**Esercizio 2.3.2.** Mostrare che gli eventi  $(\{X = x_i\})_{i \geq 1}$  costituiscono una partizione di  $\Omega$ .

Sappiamo che se  $X$  è una *qualsiasi* v. a. (cioè non necessariamente discreta), i sottoinsiemi di  $\Omega$  del tipo  $\{X = t\}$  sono eventi. Per le variabili discrete vale anche il viceversa, cioè:

**Proposizione 2.3.3.** Sia  $X$  una funzione definita su  $\Omega$  e a valori in un sottoinsieme  $\mathcal{E}$  di  $\mathbb{R}$  finito o numerabile. Allora  $X$  è una variabile aleatoria (discreta) se e solo se tutti i sottoinsiemi del tipo  $\{X = t\}$  sono eventi.

*Dimostrazione.* Basta dimostrare che, se tutti i sottoinsiemi del tipo  $\{X = t\}$  sono eventi, allora  $X$  è una variabile aleatoria. Osserviamo innanzitutto che, se  $t \notin \mathcal{E}$ , risulta  $\{X = t\} = \emptyset$ .

Occorre far vedere che gli insiemi del tipo  $\{X \leq t\}$  sono eventi. Grazie all'osservazione precedente, si ha

$$\{X \leq t\} = \bigcup_{\substack{x \in \mathbb{R} \\ x \leq t}} \{X = x\} = \bigcup_{\substack{x \in \mathcal{E} \\ x \leq t}} \{X = x\},$$

e dunque  $\{X \leq t\}$  è un evento in quanto unione al più numerabile di eventi (si ricordi che  $\mathcal{E}$  è al più numerabile per ipotesi).  $\square$

**Esercizio 2.3.4.** Sia  $X$  una v. a. discreta,  $I$  un sottoinsieme di  $\mathbb{R}$  (non necessariamente misurabile). Mostrare che il sottoinsieme di  $\Omega$   $\{X \in I\}$  è un evento.

**Definizione 2.3.5.** Sia  $X$  una v. a. discreta. L'applicazione  $p : \mathbb{R} \rightarrow \mathbb{R}$  definita da

$$p(x) = P(X = x)$$

(ben definita poiché  $\{X = x\}$  è un evento) si chiama *densità (discreta)* della v.a.  $X$ .

Quando saranno in gioco più variabili aleatorie  $X, Y, \dots$  useremo i simboli  $p_X, p_Y, \dots$  per indicare le rispettive densità.

Vediamo adesso come si può ricavare la legge di  $X$ , variabile discreta, conoscendone la densità  $p$ . Si ha

$$\{X \in I\} = \bigcup_{x \in \mathcal{E} \cap I} \{X = x\}$$

(dimostrazione per esercizio). L'unione precedente è al più numerabile (perché  $\mathcal{E}$  è al più numerabile), ed è costituita da insiemi due a due disgiunti (poiché, per l'Esercizio 2.3.2, si tratta di alcuni elementi di una partizione di  $\Omega$ ).

Per la proprietà di  $\sigma$ -additività della probabilità, si ha allora

$$P(X \in I) = \sum_{x \in \mathcal{E} \cap I} P(X = x) = \sum_{x \in \mathcal{E} \cap I} p(x) = \sum_{x \in I} p(x) \quad (2.3.6)$$

(giustificare l'ultimo passaggio). Questa formula (da ricordare bene!) è di fondamentale importanza nei calcoli.

**Osservazione 2.3.7.** Vediamone subito una semplice conseguenza: nel caso particolare  $I = \mathbb{R}$  si ottiene

$$\sum_{x \in \mathbb{R}} p(x) = \sum_{x \in \mathcal{E} \cap \mathbb{R}} p(x) = P(X \in \mathbb{R}) = P(\Omega) = 1. \quad (2.3.8)$$

**Osservazione 2.3.9.** Osservazione. Se  $x \notin \mathcal{E}$ , si ha evidentemente

$$p(x) = P(X = x) = P(\emptyset) = 0.$$

Si ha inoltre il seguente risultato:

**Teorema 2.3.10.** (a) *Sia  $p$  la densità di un'assegnata variabile aleatoria discreta  $X$ , e indichiamo con  $\mathcal{E}$  l'immagine di  $X$ ; la funzione  $p$  gode delle seguenti proprietà:*

- (i)  $p(x) \geq 0 \forall x \in \mathbb{R}$ ;
- (ii)  $p(x) = 0$  *tranne che su un'infinità al più numerabile di numeri  $x$ ;*
- (iii)  $\sum_{x \in \mathbb{R}} p(x) = 1$  *(e di conseguenza deve esistere almeno un  $x \in \mathbb{R}$  tale che  $p(x) > 0$ , dunque  $X$  non è continua); si osservi anche che, a causa della proprietà (ii), la somma sopra scritta è effettivamente una serie.*

(b) *Inoltre le proprietà (i), (ii) e (iii) caratterizzano le densità, nel senso che, se  $p$  è una funzione definita su  $\mathbb{R}$  e a valori in  $\mathbb{R}$ , che abbia tutte le proprietà citate, allora, su un opportuno spazio di probabilità  $(\Omega, \mathcal{A}, P)$  esiste una v. a. discreta  $X$  che ammette la funzione  $p$  come propria densità.*

*Dimostrazione.* La parte (a) del risultato appena enunciato è di semplice dimostrazione (le proprietà (ii) e (iii) sono appena state viste (formula 2.3.8) e Osservazione 2.3.9). Per la (b), sia  $\mathcal{E}$  l'insieme non vuoto (per la (iii)) e al più numerabile su cui  $p$  non è nulla (la sua esistenza è garantita dalla proprietà (ii)). Poniamo poi  $\Omega = \mathcal{E}$ ,  $\mathcal{A} = \mathcal{P}(\Omega)$ , e,  $\forall \omega \in \Omega$ ,  $P(\omega) = p(\omega)$ . Infine poniamo,  $\forall \omega \in \Omega$ ,  $X(\omega) = \omega$ . Si lascia al lettore il compito di verificare che la funzione  $X$  così costruita è la variabile aleatoria cercata.  $\square$

In modo analogo a quanto accade per la caratterizzazione delle funzioni di ripartizione, il teorema precedente si utilizza per decidere se un'assegnata funzione  $p$  è o no la densità di una v. a.:  $p$  sarà una densità se ha *tutte* le proprietà (i), (ii) e (iii); non lo sarà se anche una sola di esse non è valida.

Inoltre il Teorema 2.3.10(b) giustifica la seguente:

**Definizione 2.3.11.** Una funzione  $p : \mathbb{R} \rightarrow \mathbb{R}$  si dice *densità (discreta)* se gode di tutte le proprietà (i), (ii) e (iii).

**Osservazione 2.3.12.** Si noti che in questa definizione non è in gioco alcuna variabile aleatoria.

Un problema importante è il seguente: sia  $X$  una variabile aleatoria (non necessariamente discreta) definita su  $(\Omega, \mathcal{A}, P)$  e supponiamo assegnata una funzione  $\phi : \mathbb{R} \rightarrow \mathbb{R}$ . Ha senso considerare allora la funzione composta  $Y = \phi(X) = \phi \circ X$ . Ci chiediamo in quali casi  $Y$  risulta essere anch'essa una variabile aleatoria (definita su  $(\Omega, \mathcal{A}, P)$  ovviamente).

Dalla definizione di variabile aleatoria, si deduce che, affinché ciò accada, devono essere eventi tutti i sottoinsiemi di  $\Omega$  del tipo

$$\{Y \leq t\} = \{\phi(X) \leq t\} = \{X \in \phi^{-1}((-\infty, t])\}.$$

Se ne deduce che, per ogni  $t \in \mathbb{R}$ , l'insieme  $\phi^{-1}((-\infty, t])$  (cioè la controimmagine, secondo  $\phi$ , della semiretta  $(-\infty, t])$  deve essere un insieme misurabile;

chiameremo ancora *misurabili* le funzioni  $\phi$  che hanno questa proprietà. Sono misurabili ad esempio tutte le funzioni continue (non lo dimostreremo). Dunque  $Y = \phi \circ X$  è una variabile aleatoria se  $\phi$  è una funzione misurabile.

Nel caso molto particolare in cui  $X$  sia una variabile aleatoria discreta,  $Y = \phi \circ X$  è una variabile aleatoria *qualunque* sia la funzione  $\phi$  (anche non misurabile); infatti si ha

$$\{Y \leq t\} = \{\phi(X) \leq t\} = \{X \in \phi^{-1}((-\infty, t])\} = \bigcup_{x \in \mathcal{E} \cap (-\infty, t]} \{X = x\},$$

e l'unione qui sopra è un evento in quanto composta da un'infinità al più numerabile di eventi (si ricordi che  $\mathcal{E}$ , immagine di  $X$ , è al più numerabile).

Dalla formula precedente si deduce anche un modo per calcolare la funzione di ripartizione (e dunque la legge) di  $Y$ ; infatti, per la (2.3.6), indicando con  $p_X$  la densità discreta di  $X$ , si può scrivere

$$P(Y \leq t) = P(X \in \phi^{-1}((-\infty, t])) = \sum_{x \in \phi^{-1}((-\infty, t])} p_X(x) = \sum_{\phi(x) \leq t} p_X(x).$$

**Osservazione 2.3.13.** La formula precedente è di notevole importanza perché permette il calcolo della legge di  $Y$  a partire dalla conoscenza di  $\phi$  e della densità della v.a.  $X$  (cioè non è necessario utilizzare la densità di  $Y$ , come invece richiederebbe la formula (2.3.6), se applicata a  $Y$ ).

Nel prossimo paragrafo vedremo le principali densità che si incontrano nella pratica.

## 2.4 Principali densità discrete

**2.4.1. La densità binomiale.** Siano  $n \in \mathbb{N}$  e  $p \in (0, 1)$  due numeri fissati. Si chiama *densità binomiale* di parametri  $n$  e  $p$  (in simboli,  $\mathcal{B}(n, p)$ ) la funzione così definita

$$p(k) = \begin{cases} \binom{n}{k} p^k (1-p)^{n-k} & k = 0, 1, 2, \dots, n \\ 0 & \text{altrimenti.} \end{cases}$$

Vediamo adesso una tipica situazione in cui compare la densità binomiale. Sia  $(\Omega, \mathcal{A}, P)$  lo schema delle prove indipendenti di parametri (noti)  $n$  e  $p$ .

Ricordiamo che ciò significa che

$$\Omega = \{0, 1\}^n$$

ed inoltre, se  $\omega = (\omega_1, \dots, \omega_n)$  è un evento elementare (con  $\omega_i \in \{0, 1\}$  per ogni  $i = 1, \dots, n$ ), si pone

$$P(\omega) = p^{\sum_{i=1}^n \omega_i} (1-p)^{n - \sum_{i=1}^n \omega_i}.$$

Poniamo ora, per ogni  $i = 1, \dots, n$

$$X_i(\omega) = \omega_i;$$

$$X(\omega) = \sum_{i=1}^n X_i(\omega) = \sum_{i=1}^n \omega_i.$$

Le funzioni  $X_i$  e  $X$  sono variabili aleatorie (perché?) definite sullo spazio  $(\Omega, \mathcal{A}, P)$ . Diamone prima di tutto un'interpretazione concreta. Per fissare le idee, supponiamo che sia  $n = 5$ , e  $\omega = (1, 1, 0, 1, 0)$ . Ciò significa che, nello svolgimento delle 5 prove, la prima, la seconda e la quarta hanno dato "successo", la terza e la quinta "insuccesso".

Si ha subito

$$X_1(\omega) = X_2(\omega) = X_4(\omega) = 1, \quad X_3(\omega) = X_5(\omega) = 0.$$

Inoltre

$$X(\omega) = \sum_{i=1}^n X_i(\omega) = 1 + 1 + 1 = 3.$$

La v. a.  $X_i$  si chiama anche variabile *indicatrice* della prova  $i$ -esima (dice cosa è accaduto nella prova  $i$ -esima); la variabile  $X$  "conta" il numero di successi ottenuti nel corso di tutte le prove.

Tornando al caso generale, ci interessa calcolare la densità di  $X$ ; in altre parole, per ogni  $k \in \mathbb{R}$ , vogliamo calcolare  $P(X = k)$ .

È evidente che la v. a.  $X$  assume i soli valori interi  $k = 0, 1, 2, \dots, n$  (perché?). Pertanto le uniche probabilità  $P(X = k)$  (eventualmente) non nulle sono quelle che corrispondono a tali valori. Sia dunque  $k$  un intero fissato, con  $k \in \{0, 1, 2, \dots, n\}$ . L'evento  $\{X = k\}$  è costituito dalle sequenze  $\omega$  che contengono il simbolo "1" esattamente  $k$  volte, cioè tali che

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

Pertanto, per calcolare la probabilità di  $\{X = k\}$ , basterà sommare le probabilità di tutte le sequenze del tipo sopra detto. Poiché esse hanno tutte la medesima probabilità, pari a

$$P(\omega) = p^{\sum_{i=1}^n \omega_i} (1-p)^{n-\sum_{i=1}^n \omega_i} = p^k (1-p)^{n-k},$$

per avere la somma basterà moltiplicare tale valore per il numero di sequenze  $\omega$  appartenenti all'evento  $\{X = k\}$ . È facile capire che esso è esattamente il numero di combinazioni di  $n$  oggetti a  $k$  a  $k$  (è come se avessimo  $n$  caselle,  $k$  delle quali devono essere riempite con "1", mentre le restanti  $n - k$  con "0"). Questo numero è uguale a  $\binom{n}{k}$ , e quindi otteniamo

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

Si riconosce quindi la densità binomiale sopra definita.

**Osservazione 2.4.2.** È importante il caso particolare di densità  $\mathcal{B}(n, p)$  che si ha per  $n = 1$ ; la densità  $\mathcal{B}(1, p)$  che si ottiene si chiama anche *densità bernoulliana*. È immediato riconoscere che essa è la densità di una variabile aleatoria che assume i soli valori 1 e 0 con probabilità rispettive pari a  $p$  e  $1 - p$ . Esempi di variabili bernoulliane sono le v. a.  $X_i$  introdotte qui sopra (le variabili indicatrici delle singole prove).

**Esercizio 2.4.3.** Sia  $A$  un sottoinsieme di  $\Omega$ , e consideriamo la funzione  $1_A : \Omega \rightarrow \mathbb{R}$  definita da

$$1_A(\omega) = \begin{cases} 1 & \text{se } \omega \in A \\ 0 & \text{se } \omega \in A^c. \end{cases}$$

Essa si chiama *funzione indicatrice* dell'insieme  $A$ :

- (a) mostrare che  $1_A$  è una variabile aleatoria se e solo se  $A$  è un evento;
- (b) se  $A$  è un evento di probabilità  $p$ , qual è la densità di  $1_A$ ?

**2.4.4. La densità ipergeometrica.** Assegnati tre numeri interi strettamente positivi  $a, b$  e  $n$ , con  $n \leq a + b$ , poniamo  $k_1 = \max(n - b, 0)$ ;  $k_2 = \min(a, n)$  e osserviamo che è  $k_1 \leq k_2$  (fare i conti). La funzione definita nel modo seguente

$$p(k) = \begin{cases} \frac{\binom{a}{k} \binom{b}{n-k}}{\binom{a+b}{n}} & k = k_1, k_1 + 1, \dots, k_2 \\ 0 & \text{altrove} \end{cases}$$

è una densità discreta (per verificarlo si suggerisce di fare un ragionamento di tipo “combinatorio” e di non eseguire calcoli, che sono complicati).

Essa prende il nome di *densità ipergeometrica* di parametri  $a, b$  e  $n$ .

Vediamo una situazione in cui questa densità si presenta in modo naturale.

Da un'urna contenente  $a$  palline rosse e  $b$  palline verdi si eseguono  $n$  estrazioni senza rimpiazzo ( $1 \leq n \leq a + b$ ) e poniamo  $X$  = numero di palline rosse (tra le  $n$  estratte).

Costruiamo uno spazio di probabilità  $(\Omega, \mathcal{A}, P)$  con cui modellizzare l'esperimento: ogni possibile risultato è un gruppo di  $n$  palline prese dalle  $a + b$  palline presenti nell'urna; quindi l'insieme  $\Omega$  degli eventi elementari può essere identificato con l'insieme delle combinazioni di  $a + b$  oggetti a  $n$  a  $n$ .

Inoltre, poiché non c'è motivo di pensare che un particolare risultato sia più o meno probabile di ogni altro, assegniamo a ciascun  $\omega \in \Omega$  la stessa probabilità, pari a

$$P(\omega) = \frac{1}{\text{card}\Omega} = \frac{1}{\binom{a+b}{n}};$$

(si tratta cioè di uno *spazio uniforme*).



Ciò premesso, è immediato verificare che sullo spazio così costruito  $X$  è una variabile aleatoria a valori interi non negativi, e ciò che resta da fare è calcolarne la densità, cioè le quantità  $P(X = k)$  per ogni  $k$  intero (per  $k$  non intero tale probabilità vale zero).

È chiaro che  $P(X = k)$  sarà diversa da zero solo per gli interi  $k$  compresi fra  $k_1$  e  $k_2$  (infatti: il numero di palline rosse estratte non può superare né il numero di palline rosse a disposizione ( $= a$ ), né il numero totale di estrazioni effettuate ( $= n$ ). Inoltre, se  $n \geq b$ ,  $n - b$  è il minimo numero di palline rosse che possono essere estratte).

L'evento  $\{X = k\}$  corrisponde al sottoinsieme di  $\Omega$  costituito dai gruppi di  $n$  palline delle quali  $k$  sono rosse e  $n - k$  verdi, e dunque ha cardinalità pari a

$$\binom{a}{k} \binom{b}{n-k}.$$

Dato che stiamo lavorando su uno spazio uniforme si conclude che

$$P(X = k) = \frac{\text{card}\{X = k\}}{\text{card}\Omega} = \frac{\binom{a}{k} \binom{b}{n-k}}{\binom{a+b}{n}}.$$

**Osservazione 2.4.5.** Il modello di estrazioni senza rimpiazzo qui costruito si utilizza in situazioni nelle quali si è in presenza di una popolazione finita i cui componenti si differenzino per una (e una sola) caratteristica, e si debba effettuare da essa un *campionamento*, cioè, appunto, una serie di estrazioni senza rimpiazzo.

**2.4.6. La densità geometrica.** Sia  $p$  un numero assegnato, e supponiamo che sia  $0 < p < 1$ . Si chiama *densità geometrica* la funzione

$$p(k) = \begin{cases} p(1-p)^{k-1} & k = 1, 2, 3, \dots \\ 0 & \text{altrove.} \end{cases}$$

(Verificare per esercizio che si tratta effettivamente di una densità.)

Una situazione in cui si presenta la densità geometrica è la seguente.

Supponiamo di eseguire una successione infinita di lanci di una stessa moneta, per la quale è  $p$  (assegnata) la probabilità di ottenere “1” in un singolo lancio. Poniamo  $T$  = numero di lanci necessari per veder comparire la faccia “1” per la prima volta.

Ci interessa calcolare la densità di  $T$ . Prima di tutto, al solito, costruiamo il modello  $(\Omega, \mathcal{A}, P)$ . Poniamo, per  $i = 1, 2, 3, \dots$

$$\omega_i = (\underbrace{0, 0, 0, \dots}_{i-1 \text{ volte}}, 1),$$

ed inoltre

$$\omega_\delta = (0, 0, 0, \dots).$$

Il significato dei simboli qui sopra è abbastanza evidente. Notiamo solo che l'evento elementare  $\omega_\delta$  può essere espresso a parole dicendo “testa non esce mai”. Porremo dunque

$$\Omega = \{\omega_\delta, \omega_1, \omega_2, \omega_3, \dots\}.$$

L'insieme  $\Omega$ , questa volta non ha cardinalità finita, ma è comunque numerabile, e dunque anche in questo caso prenderemo  $\mathcal{A} = \mathcal{P}(\Omega)$ .

Dobbiamo ora definire la funzione  $P$ . Come vedremo fra un attimo, ciò equivale a calcolare la densità di  $T$ . Prima di tutto, osserviamo che  $T$  assume solo valori interi  $\geq 1$ ; inoltre, per  $k = 1, 2, 3, \dots$ , l'evento  $\{T = k\}$  non è altro che  $\{\omega_k\}$  (e questo implica che  $T$  è una v. a.). Si ha poi, per  $k \geq 1$

$$\{T = k\} = \{T > k - 1\} \setminus \{T > k\}$$

ed inoltre

$$\{T > k\} \subset \{T > k - 1\}$$

(dimostrazioni per esercizio).

Per formule note (quali?) si ha

$$P(T = k) = P(T > k - 1) - P(T > k).$$

Basterà allora calcolare  $P(T > k)$  per  $k = 0, 1, 2, \dots$ . Si ha intanto, in modo ovvio,

$$P(T > 0) = 1;$$

invece, per  $k \geq 1$ , dire che  $T$  assume un valore maggiore di  $k$  equivale a dire che fino al lancio  $k$ -esimo la faccia “1” non è mai uscita. Dunque, se con  $X_k$  si indica il numero di volte in cui è comparsa la faccia “1” in  $k$  lanci, l'evento  $\{T > k\}$  non è che  $\{X_k = 0\}$  e, poiché  $X_k$  ha densità  $\mathcal{B}(k, p)$  (si scrive anche  $X_k \sim \mathcal{B}(k, p)$ ), avremo

$$P(T > k) = P(X_k = 0) = \binom{k}{0} p^0 (1-p)^k = (1-p)^k.$$

La formula qui sopra è valida anche per il caso  $k = 0$ , poiché  $(1-p)^0 = 1$ . Dunque, per  $k \geq 1$  si avrà

$$P(T = k) = P(T > k - 1) - P(T > k) = (1-p)^{k-1} - (1-p)^k = p(1-p)^{k-1}.$$

**Esercizio 2.4.7.** C'è un passaggio non giustificato nel ragionamento precedente. Quale? In realtà la teoria matematica che serve per costruire  $T$  è ben più profonda di quella qui svolta: noi dobbiamo limitarci ai pochi cenni fatti sopra.

**Osservazione 2.4.8.** Lo spazio di probabilità considerato in questa situazione è un primo esempio di spazio a cardinalità non finita, ma numerabile.

La variabile aleatoria sopra considerata è una v. a. discreta, con immagine numerabile ma non finita (si tratta dell'insieme di tutti i numeri interi  $\geq 1$ ).

**Osservazione 2.4.9.** Nell'esperimento degli infiniti lanci di una moneta, potrebbe verificarsi l'evento elementare  $\{\omega_\delta\}$  (cioè potrebbe capitare di non ottenere *mai* la faccia “1”). Con quale probabilità ciò avviene (cioè quanto vale  $P(\omega_\delta)$ )?

Si può scrivere

$$\{\omega_\delta\}^c = \cup_{k=1}^{\infty} \{T = k\}.$$

(Perché? Spiegare a parole la formula precedente.) Poiché gli eventi  $\{T = k\}$  sono due a due disgiunti, per la proprietà di  $\sigma$ -additività di  $P$  (questa volta la sola additività finita non basta!) si ha

$$\begin{aligned} P(\omega_\delta) &= 1 - P\left(\cup_{k=1}^{\infty} \{T = k\}\right) = 1 - \sum_{k=1}^{\infty} P(T = k) \\ &= 1 - \sum_{k=1}^{\infty} p(1-p)^{k-1} = 1 - p \cdot \frac{1}{1 - (1-p)} = 0. \end{aligned}$$

(Ricordare la somma della serie geometrica!)

Cioè: per quanto piccola (purché strettamente positiva) sia la probabilità  $p$  di ottenere “1” in un singolo lancio, l'eventualità che “1” non esca mai è “quasi impossibile” (cioè di probabilità nulla). Questo risultato è noto sotto il pittoresco nome di “paradosso della scimmia”: una scimmia, battendo a caso sui tasti di una macchina da scrivere, prima o poi ne farà uscire l'intera *Divina Commedia*. Un nome più serio è “paradosso di Borel”.

**2.4.10. La densità di Poisson.** Sia  $\lambda > 0$  un numero reale assegnato. La funzione

$$p(k) = \begin{cases} \frac{\lambda^k}{k!} e^{-\lambda} & \text{per } k = 0, 1, 2, 3, \dots \\ 0 & \text{altrimenti} \end{cases}$$

è una densità (verifica per esercizio: si ricordi che, per ogni  $\lambda \in \mathbb{R}$  si ha

$$\sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^\lambda;$$

questa serie si chiama *serie esponenziale*).

Essa va sotto il nome di *densità di Poisson* di parametro  $\lambda$  (il simbolo usato è  $\Pi_\lambda$ ).

Per ottenere in modo naturale questa densità, consideriamo la seguente situazione: per ogni intero  $n$ , sia  $X_n$  una v. a. avente densità  $B(n, \lambda/n)$ , cioè

$$P(X_n = k) = \binom{n}{k} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k}, \quad k = 0, 1, 2, \dots, n.$$

Ci interessa trovare il limite di questa probabilità al tendere di  $n$  all'infinito (potremmo dire: ci interessa trovare a quali numeri si avvicinano i valori della densità binomiale quando il numero  $n$  di prove ripetute aumenta e la probabilità  $\lambda/n$  di successo in una singola prova diminuisce).

Fissato allora  $k$ , sia  $n$  abbastanza grande in modo che sia  $n \geq k$ . In questo caso si ha

$$\begin{aligned} P(X_n = k) &= \binom{n}{k} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} \\ &= \frac{n!}{k!(n-k)!} \frac{\lambda^k}{n^k} \left(1 - \frac{\lambda}{n}\right)^n \left(1 - \frac{\lambda}{n}\right)^{-k} \\ &= \frac{\lambda^k}{k!} \left(1 - \frac{\lambda}{n}\right)^n \left(1 - \frac{\lambda}{n}\right)^{-k} \underbrace{\frac{n}{n} \frac{n-1}{n} \frac{n-2}{n} \dots \frac{n-k+1}{n}}_{k \text{ fattori}}. \end{aligned}$$

Osservato che

- (i)  $\lim_{n \rightarrow \infty} \left(1 - \frac{\lambda}{n}\right)^n = e^{-\lambda}$ ;
- (ii)  $\lim_{n \rightarrow \infty} \frac{n}{n} \frac{n-1}{n} \frac{n-2}{n} \dots \frac{n-k+1}{n} = \lim_{n \rightarrow \infty} 1 \cdot \left(1 - \frac{1}{n}\right) \cdot \dots \cdot \left(1 - \frac{k-1}{n}\right) = 1$ ;
- (iii)  $\lim_{n \rightarrow \infty} \left(1 - \frac{\lambda}{n}\right)^{-k} = 1$ ,

se ne deduce che

$$\lim_{n \rightarrow \infty} P(X_n = k) = \frac{\lambda^k}{k!} e^{-\lambda},$$

e troviamo quindi la densità di Poisson di parametro  $\lambda$ , calcolata in  $k$ .

Questo risultato spiega l'uso della densità di Poisson quando la variabile da studiare si può pensare come il numero di successi in un numero molto grande di prove con bassa probabilità di successo nella singola prova. Esempi tipici sono il numero di telefonate che arrivano in un determinato intervallo di tempo ad un centralino, il numero di auto che passano da un casello autostradale in un dato intervallo di tempo, ecc.

## 2.5 Vettori aleatori discreti

*Motivazione.* Può capitare che l'esperimento sotto osservazione produca più variabili interessanti, e che di esse si debba studiare il comportamento “congiunto”: ad esempio, in uno schema di 5 prove indipendenti, consideriamo le due quantità  $X = \text{numero di successi nelle 5 prove}$  e  $Y = \text{numero di successi nelle prime 2 prove}$ ; in tal caso una domanda interessante a cui rispondere potrebbe essere: quanto vale la probabilità che su un totale di 3 successi, 2 di essi si siano presentati nelle prime 2 prove? In simboli, questo significa calcolare la “probabilità congiunta”

$$P(\{X = 3\} \cap \{Y = 2\}).$$

Supporremo fissato una volta per tutte lo spazio di probabilità  $(\Omega, \mathcal{A}, P)$ .

**Definizione 2.5.1.** Siano  $X_1, X_2, \dots, X_m$   $m$  variabili aleatorie discrete (tutte definite su  $(\Omega, \mathcal{A}, P)$ ). Si chiama *vettore aleatorio discreto  $m$ -dimensionale* (su  $(\Omega, \mathcal{A}, P)$ ) la  $m$ -upla di variabili aleatorie

$$X = (X_1, X_2, \dots, X_m);$$

si tratta della funzione  $X : \Omega \rightarrow \mathbb{R}^m$  definita da

$$X(\omega) = (X_1(\omega), X_2(\omega), \dots, X_m(\omega)).$$

**Osservazione 2.5.2.** L'attributo *aleatorio* dato al vettore  $X$  è giustificato dal fatto seguente: per ogni  $i = 1, 2, \dots, m$  sia  $x_i \in \mathbb{R}$ . Come sappiamo, il sottoinsieme di  $\Omega$   $\{X_i = x_i\}$  è un evento, dunque, per la proprietà di stabilità della  $\sigma$ -algebra  $\mathcal{A}$  rispetto all'intersezione finita, è un evento anche il sottoinsieme

$$\{X_1 = x_1\} \cap \{X_2 = x_2\} \cap \dots \cap \{X_m = x_m\},$$

che viene indicato più brevemente con la scrittura

$$\{X_1 = x_1, X_2 = x_2, \dots, X_m = x_m\}.$$

Ha quindi senso considerare la probabilità degli eventi del tipo precedente, e si ha la

**Definizione 2.5.3.** Indichiamo con  $x = (x_1, x_2, \dots, x_m)$  il generico punto di  $\mathbb{R}^m$ . Si chiama *densità* del vettore  $X$  (o *densità congiunta* delle variabili  $X_1, X_2, \dots, X_m$ ) la funzione  $p : \mathbb{R}^m \rightarrow \mathbb{R}$  definita da

$$p(x) = p(x_1, x_2, \dots, x_m) = P(X_1 = x_1, X_2 = x_2, \dots, X_m = x_m).$$

Per  $i = 1, 2, \dots, m$  la densità di  $X_i$  si chiama anche *densità marginale  $i$ -esima*.

**Osservazione 2.5.4.** (i) Per ogni  $i = 1, 2, \dots, m$  sia  $\mathcal{E}_i$  l'immagine (al più numerabile) di  $X_i$ . Risulta allora  $p(x_1, x_2, \dots, x_m) = 0$  se  $x_i \notin \mathcal{E}_i$  per qualche  $i$ . Infatti, in tal caso, gli eventi  $\{X_i = x_i\}$  e, a maggior ragione,  $\{X_1 = x_1, X_2 = x_2, \dots, X_m = x_m\}$  sono vuoti (perché?).

Pertanto la funzione  $p$  è diversa da zero su una parte dell'insieme (al più numerabile)  $\mathcal{E} = \mathcal{E}_1 \times \mathcal{E}_2 \times \dots \times \mathcal{E}_m$ .

(ii) Gli eventi  $\{X_1 = x_1, X_2 = x_2, \dots, X_m = x_m\}$  per  $x = (x_1, \dots, x_m) \in \mathcal{E}$  costituiscono una partizione di  $\Omega$  (verifica per esercizio). Ciò implica che

$$\begin{aligned} 1 &= P(\Omega) = P(\cup_{x \in \mathcal{E}} \{X_1 = x_1, X_2 = x_2, \dots, X_m = x_m\}) \\ &= \sum_{x \in \mathcal{E}} P(X_1 = x_1, X_2 = x_2, \dots, X_m = x_m) \\ &= \sum_{x \in \mathbb{R}^m} P(X_1 = x_1, X_2 = x_2, \dots, X_m = x_m), \end{aligned} \tag{2.5.5}$$

dove l'ultima eguaglianza è giustificata dal punto (i) precedente.

In particolare, dalla (2.5.5) discende il fatto che  $p$  non può essere identicamente nulla.

*Notazione.* Per  $m = 2$  scriveremo  $X, Y$  anziché  $X_1, X_2$ . Per  $m = 3$  analogamente scriveremo  $X, Y, Z$  anziché  $X_1, X_2, X_3$ . In particolare, fare attenzione al fatto che in questi contesti il simbolo  $X$  non rappresenta un vettore aleatorio a due o tre dimensioni, ma una variabile aleatoria *uno-dimensionale*.

Vedremo adesso in che modo si possono ricavare le densità marginali dalla congiunta. Per semplicità di scrittura, discuteremo il caso “bivariato” (cioè  $m = 2$ ). I calcoli e le formule sono del tutto analoghi nel caso generale.

Sia dunque  $(X, Y)$  un vettore aleatorio bidimensionale, di densità congiunta  $p$ . Cioè, per  $(x, y) \in \mathbb{R}^2$  si ha

$$p(x, y) = P(X = x, Y = y).$$

Ricordando che gli eventi  $\{Y = y\}$  rappresentano una partizione di  $\Omega$  (al variare di  $y$  in  $\mathbb{R}$ ), si ha

$$\begin{aligned} p_X(x) &= P(X = x) = P(\{X = x\} \cap \Omega) \\ &= P(\{X = x\} \cap (\cup_{y \in \mathbb{R}} \{Y = y\})) \\ &= \sum_{y \in \mathbb{R}} P(X = x, Y = y) = \sum_{y \in \mathbb{R}} p(x, y). \end{aligned}$$

In modo del tutto analogo si dimostra che

$$p_Y(y) = \sum_{x \in \mathbb{R}} p(x, y). \quad (2.5.6)$$

Vediamo qualche formula dello stesso tipo per il caso  $m = 3$ . Si ha, ad esempio

$$p_X(x) = \sum_{(y,z) \in \mathbb{R}^2} p(x, y, z);$$

$$p_Y(y) = \sum_{(x,z) \in \mathbb{R}^2} p(x, y, z);$$

e ancora

$$p_{X,Y}(x, y) := P(X = x, Y = y) = \sum_{z \in \mathbb{R}} p(x, y, z).$$

Dunque, ricapitolando, abbiamo visto che è possibile ottenere le densità marginali quando si conosce la densità congiunta. Vediamo ora sull'esempio che segue che, al contrario, in generale la conoscenza delle sole densità marginali non permette di ricavare la congiunta.

**Esempio 2.5.7.** (a) Da un'urna contenente 5 palline numerate da 1 a 5 si eseguono due estrazioni *con rimpiazzo* (significa che, dopo ogni estrazione, la pallina scelta viene rimessa nell'urna prima di eseguire l'estrazione successiva). Indichiamo con  $X$  (risp.  $Y$ ) il numero riportato sulla prima (risp. sulla seconda) pallina estratta. Calcoliamo la densità congiunta  $p$  della coppia di v. a.  $(X, Y)$  e le due densità marginali  $p_X$  e  $p_Y$ .

Evidentemente (si lasciano le verifiche per esercizio)

$$p(x, y) = P(X = x, Y = y) = \begin{cases} \frac{1}{25} & \text{per } (x, y) \in \{1, 2, 3, 4, 5\}^2 \\ 0 & \text{altrove.} \end{cases}$$

Inoltre

$$p_X(i) = p_Y(i) = \begin{cases} \frac{1}{5} & \text{per } i \in \{1, 2, 3, 4, 5\} \\ 0 & \text{altrove.} \end{cases}$$

(b) Rispondiamo ora alle stesse domande del punto (a) nel caso che le due estrazioni siano eseguite *senza rimpiazzo* (cioè, dopo ogni estrazione, la pallina scelta viene gettata via). Anche in questo caso si lasciano alcune verifiche per esercizio.

Posto

$$\Delta = \{(x, y) \in \mathbb{R}^2 : x = y\},$$

si vede facilmente che è

$$p(x, y) = P(X = x, Y = y) = \begin{cases} \frac{1}{20} & \text{per } (x, y) \in \{1, 2, 3, 4, 5\}^2 \setminus \Delta \\ 0 & \text{altrove.} \end{cases}$$

Inoltre è ovvio che la densità marginale di  $X$  (cioè  $p_X$ ) è identica a quella del caso (a). Calcoliamo la densità marginale di  $Y$  usando la formula (2.5.6). Risulta

$$p_Y(y) = \sum_{x \in \mathbb{R}} p(x, y) = \sum_{\substack{x=1, \dots, 5 \\ x \neq y}} \frac{1}{20} = \frac{4}{20} = \frac{1}{5}.$$

Si ottiene dunque ancora la densità del punto (a).

Ricapitolando, le densità marginali sono identiche a quelle del caso di estrazioni con rimpiazzo, e tuttavia le due densità congiunte sono diverse (in altri termini, le densità marginali non individuano in modo unico la congiunta).

Vediamo ora in che modo la densità congiunta viene usata per calcolare le probabilità di eventi più complessi. Anche adesso tratteremo per praticità il caso  $m = 2$ , ma la formula che si ottiene sarà del tutto generale.

Gli eventi che ci interessano sono della forma  $\{\omega \in \Omega : (X(\omega), Y(\omega)) \in S\}$ , dove  $S$  è un sottoinsieme di  $\mathbb{R}^2$  (di nuovo, scriveremo più semplicemente  $\{(X, Y) \in S\}$ ).

**Esercizio 2.5.8.** Verificare che il sottoinsieme di  $\Omega$  sopra scritto è effettivamente un evento.

**Osservazione 2.5.9.** Sono del tipo considerato eventi quali, ad esempio,  $\{X < Y\}$ ,  $\{X \geq Y^2\}$  e così via. (Nel primo caso basta prendere  $S = \{(x, y) \in \mathbb{R}^2 : x < y\}$ , nel secondo  $S = \{(x, y) \in \mathbb{R}^2 : x \geq y^2\}$ ).

L'evento  $\{(X, Y) \in S\}$  può evidentemente scriversi nella forma

$$\bigcup_{(x,y) \in S} \{X = x, Y = y\},$$

e l'unione precedente è fatta con eventi due a due disgiunti; dunque, passando alle probabilità, si trova

$$P((X, Y) \in S) = \sum_{(x,y) \in S} P(X = x, Y = y) = \sum_{(x,y) \in S} p(x, y),$$

dove, al solito,  $p$  rappresenta la densità congiunta del vettore  $(X, Y)$ .

Abbiamo cioè ottenuto la formula

$$P((X, Y) \in S) = \sum_{(x,y) \in S} p(x, y),$$

del tutto analoga a quella ricavata nel caso di una sola variabile (ved. (2.3.6)). La generalizzazione al caso  $m \geq 2$  viene lasciata per esercizio al lettore.

## 2.6 Variabili aleatorie indipendenti

Le variabili aleatorie considerate nella prima parte di questo paragrafo sono da intendersi generiche (cioè non necessariamente discrete). Supporremo al solito fissato lo spazio di probabilità  $(\Omega, \mathcal{A}, P)$ . Può accadere che due variabili aleatorie  $X$  e  $Y$  siano tali che il sapere che  $X$  assume un valore in un certo insieme  $A$  non cambi la valutazione di probabilità che  $Y$  assuma un valore in un dato insieme  $B$ : in altre parole, tali che gli eventi  $\{X \in A\}$  e  $\{Y \in B\}$  siano indipendenti, e ciò per ogni coppia di sottoinsiemi (misurabili)  $A$  e  $B$  di  $\mathbb{R}$ . Questo giustifica la seguente:

**Definizione 2.6.1.** Due variabili aleatorie  $X$  e  $Y$  si dicono tra loro *indipendenti* se, per ogni coppia di sottoinsiemi (misurabili)  $A$  e  $B$  di  $\mathbb{R}$ , risulta

$$P(X \in A, Y \in B) = P(X \in A)P(Y \in B). \quad (2.6.2)$$



Vediamo subito un semplice esempio.

**Esempio 2.6.3.** Si lancia due volte una moneta, per la quale è  $p \in (0, 1)$  (nota) la probabilità di ottenere “1” in un generico lancio; siano  $X$  e  $Y$  le variabili aleatorie indicatrici del primo e del secondo lancio. È immediato verificare che  $X$  e  $Y$  sono indipendenti secondo la definizione precedente (e ciò in accordo con l'intuizione).

La Definizione 2.6.1 si estende al caso di più variabili aleatorie come segue:

**Definizione 2.6.4.** Le variabili aleatorie  $X_1, X_2, \dots, X_m$  si dicono tra loro *indipendenti* se, per ogni famiglia  $A_1, A_2, \dots, A_m$  di sottoinsiemi (misurabili) di  $\mathbb{R}$ , risulta

$$P(X_1 \in A_1, X_2 \in A_2, \dots, X_m \in A_m) = \prod_{i=1}^m P(X_i \in A_i). \quad (2.6.5)$$

**Osservazione 2.6.6.** La condizione (2.6.5) è equivalente all'indipendenza della famiglia di eventi  $\{X_1 \in A_1\}, \{X_2 \in A_2\}, \dots, \{X_m \in A_m\}$ . A prima vista, può sembrare che la (2.6.5) sia più debole, poiché riguarda soltanto l'intersezione dell'intera famiglia  $\{X_i \in A_i\}, i = 1, \dots, m$  e non le intersezioni di tutte le sottofamiglie (come richiede la condizione di indipendenza). Tuttavia non è così: per ottenere la sottofamiglia composta dagli eventi  $\{X_{i_1} \in A_{i_1}\}, \{X_{i_2} \in A_{i_2}\}, \dots, \{X_{i_k} \in A_{i_k}\}$  basta porre  $A_i = \mathbb{R}$  per ogni  $i \notin \{i_1, i_2, \dots, i_k\}$ . In tal modo gli eventi corrispondenti sono tutti uguali all'intero spazio  $\Omega$  e non danno contributo all'intersezione; inoltre le relative probabilità valgono 1 e non danno contributo al prodotto delle probabilità.

La condizione delle Definizione 2.6.4 in generale è (per lo meno) scomoda per verificare se due o più variabili sono indipendenti. Cerchiamo allora condizioni più semplici, limitandoci per il momento al caso delle variabili aleatorie discrete. Per questo tipo di variabili si ha il seguente risultato:

**Proposizione 2.6.7.** Siano  $X_1, X_2, \dots, X_m$   $m$  variabili aleatorie discrete, aventi densità congiunta  $p(x_1, x_2, \dots, x_m)$  e densità marginali  $p_{X_1}(x_1), p_{X_2}(x_2), \dots, p_{X_m}(x_m)$ . Allora  $X_1, X_2, \dots, X_m$  sono fra loro indipendenti se e solo se,  $\forall (x_1, x_2, \dots, x_m) \in \mathbb{R}^m$ , vale la relazione

$$p(x_1, x_2, \dots, x_m) = p_{X_1}(x_1) \times p_{X_2}(x_2) \times \dots \times p_{X_m}(x_m) \quad (2.6.8)$$

(la funzione a secondo membro si chiama di *prodotto tensoriale*  $p_{X_1}, \dots, p_{X_m}$ ).

*Dimostrazione.* Al solito per semplicità tratteremo solo il caso  $m = 2$ . Siano  $X$  e  $Y$  due variabili aleatorie indipendenti; allora la condizione (2.6.2) vale in particolare per gli insiemi del tipo  $A = \{x\}$  e  $B = \{y\}$ , con  $x$  e  $y \in \mathbb{R}$ , e si

ottiene così la relazione (2.6.8), ossia

$$p(x, y) = p_X(x) \times p_Y(y).$$

Viceversa, supponiamo valida la (2.6.8), e siano  $A$  e  $B$  due sottoinsiemi di  $\mathbb{R}$ . Si ha allora

$$\begin{aligned} P(X \in A, Y \in B) &= P((X, Y) \in A \times B) \\ &= \sum_{(x, y) \in A \times B} p(x, y) = \sum_{x \in A, y \in B} p_X(x) \times p_Y(y) \\ &= \sum_{x \in A} p_X(x) \sum_{y \in B} p_Y(y) = P(X \in A) \times P(Y \in B). \quad \square \end{aligned}$$

Un altro risultato importante riguarda le variabili aleatorie che siano funzioni di variabili aleatorie indipendenti; precisamente:

**Proposizione 2.6.9.** *Siano  $X$  e  $Y$  due variabili aleatorie discrete indipendenti,  $\phi$  e  $\psi$  due funzioni definite su  $\mathbb{R}$  e a valori in  $\mathbb{R}$ . Allora le variabili aleatorie  $U = \phi \circ X$  e  $V = \psi \circ Y$  sono fra loro indipendenti.*

*Dimostrazione.* Per ogni coppia di sottoinsiemi  $A, B$  di  $\mathbb{R}$  si ha

$$\begin{aligned} P(U \in A, V \in B) &= P(\phi \circ X \in A, \psi \circ Y \in B) \\ &= P(X \in \phi^{-1}(A), Y \in \psi^{-1}(B)) \\ &= P(X \in \phi^{-1}(A)) \times P(Y \in \psi^{-1}(B)) \\ &= P(\phi \circ X \in A) \times P(\psi \circ Y \in B) \\ &= P(U \in A) \times P(V \in B). \quad \square \end{aligned}$$

**Osservazione 2.6.10.** Il risultato che abbiamo appena dimostrato si può estendere al caso di più di due variabili nel modo seguente: due funzioni di gruppi “separati” di variabili aleatorie discrete indipendenti sono tra loro indipendenti; in termini precisi:

**Proposizione 2.6.11.** *Siano  $X_1, X_2, \dots, X_m$   $m$  variabili aleatorie discrete indipendenti,  $I = \{i_1, i_2, \dots, i_k\}$  un sottoinsieme di  $\{1, 2, \dots, m\}$ . Consideriamo i due vettori aleatori discreti  $X = (X_i)_{i \in I}$  e  $Y = (X_i)_{i \in I^c}$ . Siano infine  $\phi : \mathbb{R}^k \rightarrow \mathbb{R}$  e  $\psi : \mathbb{R}^{m-k} \rightarrow \mathbb{R}$  due funzioni. Allora le variabili aleatorie  $U = \phi \circ X$  e  $V = \psi \circ Y$  sono fra loro indipendenti.*

Un’ultima nozione importante è la seguente:

**Definizione 2.6.12.** Siano  $X$  e  $Y$  due variabili aleatorie discrete, aventi densità congiunta  $p(x, y)$ . Si chiama *densità condizionale* di  $Y$ , dato  $X = x$ , la

funzione (della sola variabile  $y$ !) definita come segue

$$y \mapsto p_{Y|X}(y|x) := \begin{cases} \frac{P(Y=y, X=x)}{P(X=x)} = \frac{p(x,y)}{p_X(x)} & \text{se } p_X(x) \neq 0 \\ 0 & \text{altrimenti.} \end{cases}$$

**Osservazione 2.6.13.** La quantità  $p_{Y|X}(y|x)$ , nel caso in cui  $x$  sia tale che  $p_X(x) \neq 0$ , non è altro che la probabilità condizionale dell'evento  $\{Y = y\}$  dato l'evento  $\{X = x\}$ .

**Esercizio 2.6.14.** Mostrare che vale la formula

$$p(x, y) = p_{Y|X}(y|x)p_X(x)$$

(anche nel caso in cui  $x$  sia tale che  $p_X(x) = 0$ ).

## Speranza matematica e varianza

### 3.1 Definizioni e proprietà

**Esempio 3.1.1.** Si lancia una volta un dado equilibrato. Qual è il “risultato medio” che si ottiene? La risposta intuitiva è

$$\frac{1 + 2 + 3 + 4 + 5 + 6}{6} = 3.5.$$

Vediamo di capire un po’ più a fondo quel che abbiamo fatto.

Il risultato del lancio può essere descritto mediante una v. a.  $X$  che prenda i valori interi 1, 2, 3, 4, 5, 6 ciascuno con probabilità uguale a  $1/6$ . In altre parole  $X$  ha densità

$$p_X(x) = \begin{cases} 1/6 & \text{per } x \in \{1, 2, 3, 4, 5, 6\} \\ 0 & \text{altrove.} \end{cases}$$

Si riconosce allora che il valore medio scritto sopra è la somma dei valori assunti dalla  $X$ , ciascuno di essi moltiplicato per la probabilità con cui viene assunto:

$$1 \times \frac{1}{6} + 2 \times \frac{1}{6} + 3 \times \frac{1}{6} + 4 \times \frac{1}{6} + 5 \times \frac{1}{6} + 6 \times \frac{1}{6}.$$

Il calcolo precedente giustifica la seguente:

**Definizione 3.1.2.** Sia  $X$  una assegnata variabile aleatoria discreta sullo spazio  $(\Omega, \mathcal{A}, P)$ , avente densità  $p_X$ . Si dice che  $X$  ha *speranza matematica finita* se

$$\sum_{x \in \mathbb{R}} |x| p_X(x) < +\infty. \quad (3.1.3)$$

In tal caso si chiama *speranza matematica* di  $X$  la quantità

$$\mathbf{E}[X] = \sum_{x \in \mathbb{R}} x p_X(x). \quad (3.1.4)$$

**Osservazione 3.1.5.** La speranza matematica si chiama anche, semplicemente, *speranza*. Altri nomi usati sono *attesa*, *valore atteso*, *media*, *valore medio*. Alcuni termini inglesi sono *expectation*, *mean*, *average*.

**Osservazione 3.1.6.** La (3.1.3) non è altro che la condizione di convergenza assoluta della serie in (3.1.4), e dunque (per un ben noto risultato di analisi) garantisce la convergenza semplice della stessa serie (3.1.4). È ovvio che, se  $X$  assume un numero finito di valori, la (3.1.3) è soddisfatta. Essa diventa importante se  $X$  assume invece un'infinità numerabile di valori.

Il teorema che segue (che enunciamo senza dimostrazione) sarà usato per dimostrare le proprietà della speranza; esso è comunque utile anche nelle applicazioni.

**Teorema 3.1.7.** Sia  $X = (X_1, X_2, \dots, X_m)$  un vettore aleatorio discreto  $m$ -dimensionale, avente densità congiunta  $p(x) = p(x_1, x_2, \dots, x_m)$ . Sia poi  $\phi : \mathbb{R}^m \rightarrow \mathbb{R}$  una funzione, e consideriamo la variabile aleatoria  $Z = \phi \circ X$ . Allora  $Z$  ha speranza matematica finita se e solo se

$$\sum_{x \in \mathbb{R}} |\phi(x)| p_X(x) < +\infty.$$

In tal caso si ha

$$\mathbf{E}[Z] = \sum_{x \in \mathbb{R}} \phi(x) p_X(x).$$

**Osservazione 3.1.8.** Nelle applicazioni, il teorema precedente si usa quando si vuole calcolare la speranza di una v.a.  $Z = \phi \circ X$  senza dover prima trovare la densità di  $Z$ , come invece richiederebbe la definizione di speranza.

Dopo questa premessa, passiamo ad enunciare e dimostrare le principali proprietà della speranza.

**Proposizione 3.1.9.** Siano  $X$  e  $Y$  due v. a. discrete aventi entrambe speranza matematica finita. Allora:

(i) per ogni  $c \in \mathbb{R}$  la variabile aleatoria  $cX$  ha speranza matematica finita e si ha

$$\mathbf{E}[cX] = c\mathbf{E}[X];$$

(ii) la v. a.  $X + Y$  ha speranza matematica finita e si ha

$$\mathbf{E}[X + Y] = \mathbf{E}[X] + \mathbf{E}[Y].$$

(Le proprietà (i) e (ii) precedenti si possono riassumere dicendo che l'operatore  $X \mapsto \mathbf{E}[X]$  è lineare.)

*Dimostrazione.* Dimostriamo la (ii). La (i) è analoga, anzi molto più semplice. Poniamo  $Z = X + Y$ . Allora  $Z = \phi(X, Y)$ , dove  $\phi : \mathbb{R}^2 \rightarrow \mathbb{R}$  è la funzione

$$\phi(x, y) = x + y.$$

Sia  $p_{X,Y}$  la densità congiunta di  $(X, Y)$ . Si ha

$$\begin{aligned} \sum_{x,y} |x+y| p_{X,Y}(x,y) &\leq \sum_{x,y} (|x| + |y|) p_{X,Y}(x,y) = \\ &= \sum_{x,y} |x| p_{X,Y}(x,y) + \sum_{x,y} |y| p_{X,Y}(x,y) = \\ &= \sum_x |x| \sum_y p_{X,Y}(x,y) + \sum_y |y| \sum_x p_{X,Y}(x,y) = \\ &= \sum_x |x| p_X(x) + \sum_y |y| p_Y(y) < +\infty. \end{aligned}$$

Se si ripercorre il calcolo precedente togliendo tutti i valori assoluti (e di conseguenza il segno di  $\leq$  all'inizio diventa un segno di  $=$ ), si trova la formula da dimostrare, cioè

$$\mathbf{E}[X + Y] = \mathbf{E}[X] + \mathbf{E}[Y].$$

Per dimostrare il punto (i) della proposizione si usa di nuovo il teorema precedente applicato alla funzione  $\phi : \mathbb{R} \rightarrow \mathbb{R}$  definita da  $\phi(x) = cx$ .  $\square$

Un altro risultato utile è il seguente (del quale si lascia la dimostrazione per esercizio):

**Proposizione 3.1.10.** *Siano  $X$  e  $Y$  due variabili aleatorie con speranza matematica finita. Si hanno le seguenti proprietà:*

- (i) Se  $P(X \leq Y) = 1$ , allora  $\mathbf{E}[X] \leq \mathbf{E}[Y]$ ;
- (ii)  $|\mathbf{E}[X]| \leq \mathbf{E}[|X|]$ .

**Osservazione 3.1.11.** Il punto (i) della proposizione precedente si utilizza tra l'altro quando si vuole stimare la speranza di una v. a.  $X$ , ad esempio maggiorandola con quella di  $Y$ , perché non siamo in grado di calcolarla esplicitamente (ed invece sappiamo calcolare la speranza di  $Y$ ).

Vediamo ora cosa possiamo dire riguardo al comportamento dell'operatore  $X \mapsto \mathbf{E}[X]$  rispetto alla moltiplicazione tra variabili aleatorie. In generale è falso che la speranza del prodotto di due v. a. sia uguale al prodotto delle singole speranze (trovare un esempio). La cosa però è vera nel caso di due variabili aleatorie indipendenti; si ha infatti:

**Proposizione 3.1.12.** *Siano  $X$  e  $Y$  due v. a. indipendenti, entrambe con speranza matematica finita. Allora la v. a.  $Z = XY$  ha speranza matematica finita e*

$$\mathbf{E}[XY] = \mathbf{E}[X]\mathbf{E}[Y].$$

La dimostrazione è analoga a quella della Proposizione 3.1.9: si pone  $\phi(x, y) = xy$  e si usano il Teorema 3.1.7 e il fatto che la densità congiunta di due v.a. indipendenti si può fattorizzare nel prodotto tensoriale delle marginali.

Vediamo ora il calcolo della speranza in alcuni casi importanti.

**3.1.13.** (i) *Densità bernoulliana.* Sia  $X$  una v. a. avente densità  $\mathcal{B}(1, p)$ . Cioè:

$$X = \begin{cases} 1 & \text{con probabilità } p \\ 0 & \text{con probabilità } 1 - p. \end{cases}$$

Applicando allora la definizione di speranza, si ha

$$\mathbf{E}[X] = 1 \times p + 0 \times (1 - p) = p.$$

(ii) *Densità binomiale.* Sia  $X$  una v. a. avente densità binomiale  $\mathcal{B}(n, p)$ ; ciò significa che  $X$  assume i valori interi  $0, 1, 2, \dots, n$ , e che, per  $k = 0, 1, 2, \dots, n$  si ha

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

Se si applica anche ora la definizione, si trova

$$\mathbf{E}[X] = \sum_{k=0}^n k \binom{n}{k} p^k (1 - p)^{n-k}.$$

Il valore della somma precedente si può calcolare, e si trova che è  $np$ ; tuttavia il calcolo è un po' lungo (si lascia per esercizio). Qui preferiamo trovare la speranza di  $X$  per altra via.

Abbiamo visto in (2.4.1) che la v. a.  $Y =$  numero di successi in  $n$  prove indipendenti, tutte di legge  $\mathcal{B}(1, p)$ , ha densità  $\mathcal{B}(n, p)$  e si può scrivere come

$$Y = \sum_{i=1}^n X_i,$$

dove le  $X_i$  hanno legge  $\mathcal{B}(1, p)$ : poiché la speranza di una variabile dipende esclusivamente dalla sua legge, possiamo utilizzare la v. a.  $Y$  per il calcolo che ci interessa. Per la proprietà di linearità dell'operatore "speranza" e per il punto precedente si ha allora

$$\mathbf{E}[Y] = \sum_{i=1}^n \mathbf{E}[X_i] = \sum_{i=1}^n p = np.$$

(iii) *Densità di Poisson.* Sia  $X$  una v. a. con densità di Poisson.  $X$  assume tutti i valori interi e

$$P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}, \quad k = 0, 1, 2, \dots$$

Si ha dunque

$$\begin{aligned} \mathbf{E}[X] &= \sum_{k=0}^{\infty} k \frac{\lambda^k}{k!} e^{-\lambda} = \sum_{k=1}^{\infty} k \frac{\lambda^k}{k!} e^{-\lambda} \\ &= \sum_{k=1}^{\infty} \frac{\lambda^k}{(k-1)!} e^{-\lambda} = \lambda \sum_{h=0}^{\infty} \frac{\lambda^h}{h!} e^{-\lambda} = \lambda, \end{aligned}$$

poiché l'ultima somma scritta sopra non è altro che la somma dei valori della densità di Poisson, e quindi vale 1.

(iv) *Densità geometrica.* Sia infine  $X$  una v. a. con densità geometrica di parametro  $p$ . Allora

$$\mathbf{E}[X] = \sum_{k=1}^{+\infty} kp(1-p)^{k-1}.$$

Si può mostrare (ma non faremo i conti) che la serie sopra scritta ha somma  $\frac{1}{p}$ . Dunque, se interpretiamo  $X$  come il numero di prove necessarie per ottenere il primo successo, si trova che la media di questo numero, essendo l'inverso del parametro  $p$ , sarà molto grande se  $p$  è molto piccolo: la scimmia che scrive la *Divina Commedia* battendo a caso sui tasti della macchina da scrivere cesserà di vivere prima di aver terminato.

La speranza di una v. a.  $X$  è una delle cosiddette *misure di centralità* della legge di  $X$ : ciò significa che, in un certo senso, la speranza ci dà una valutazione del valore centrale di  $X$ ; tuttavia fino a questo momento non abbiamo nessuna indicazione su quanto grande sia la “tendenza” dei valori di  $X$  a disperdersi attorno a tale indice di centralità. Vediamo di chiarire con qualche esempio.

**Esempio 3.1.14.** (i) Sia  $X \equiv 0$ . Allora  $\mathbf{E}[X] = 0$ . I valori di  $X$  (cioè il solo valore 0) coincidono esattamente con la media.

(ii) Sia

$$X = \begin{cases} n & \text{con probabilità } 1/2 \\ -n & \text{con probabilità } 1/2. \end{cases}$$

Di nuovo si ha  $\mathbf{E}[X] = 0$ ; tuttavia, all'aumentare di  $n$ , i valori della variabile tendono a disperdersi sempre più attorno alla media.

Si presenta allora la questione di come misurare la “quantità di dispersione” di  $X$  attorno a  $\mathbf{E}[X]$ ; si tratta cioè di trovare un opportuno “indice di dispersione”.



Una prima idea potrebbe essere la seguente: lo scarto tra  $X$  e  $\mathbf{E}[X]$  è dato da  $|X - \mathbf{E}[X]|$ ; il valore medio di tale scarto è dunque

$$\mathbf{E}[|X - \mathbf{E}[X]|],$$

e questo numero potrebbe essere un ragionevole indice di dispersione; il problema è che il suo uso non è agevole perché i calcoli con esso risultano in genere troppo complicati. Si può allora rimediare così : per misurare lo scarto tra  $X$  e  $\mathbf{E}[X]$  usiamo la quantità  $|X - \mathbf{E}[X]|^2 = (X - \mathbf{E}[X])^2$  (una differenza consiste nel fatto che le unità di misura, che prima erano quelle di  $X$ , adesso sono elevate al quadrato). La media di questa nuova variabile è

$$\mathbf{E}[(X - \mathbf{E}[X])^2].$$

Fortunatamente, come vedremo anche noi tra poco, con questa nuova quantità i calcoli sono abbastanza semplici; inoltre, ed è la cosa più importante, si tratta di una “buona” misura della dispersione (giustificheremo fra qualche pagina questa affermazione, Proposizione 3.1.23). Prima di introdurla in modo rigoroso, abbiamo bisogno di qualche preliminare.

**Definizione 3.1.15.** (i) Sia  $X$  una variabile aleatoria (discreta), e sia  $k$  un numero intero  $\geq 1$ . Si dice che  $X$  ammette momento, o momento teorico (finito) di ordine  $k$  se la v. a.  $X^k$  ammette speranza finita, cioè se

$$\mathbf{E}[|X|^k] < +\infty.$$

In tal caso, si chiama *momento (teorico) di ordine  $k$*  di  $X$  il numero  $\mathbf{E}[X^k]$ .

(ii) Sia  $X$  una v. a. avente speranza finita e sia  $k$  un numero intero  $\geq 2$ . Si dice che  $X$  ammette momento *centrato* (finito) di ordine  $k$  se la v.a.  $(X - \mathbf{E}[X])^k$  ha speranza finita, ovvero se

$$\mathbf{E}[|X - \mathbf{E}[X]|^k] < +\infty,$$

e, in tal caso, si chiama *momento centrato di ordine  $k$*  di  $X$  il numero  $\mathbf{E}[(X - \mathbf{E}[X])^k]$ .

**Osservazione 3.1.16.** Il momento del primo ordine di  $X$  non è altro che la sua speranza matematica.

I momenti e i momenti centrati di una variabile aleatoria hanno qualche interessante proprietà.

**Proposizione 3.1.17.** *Sia  $X$  una v. a. avente momento finito di ordine  $k$ . Allora*

- (i)  *$X$  ammette momento finito di ordine  $h$ , per ogni  $h \leq k$ ;*
- (ii) *se  $k \geq 2$ , allora  $X$  ammette momento centrato finito di ordine  $h$ , per ogni  $h \leq k$ .*

**Definizione 3.1.18.** Sia  $X$  una variabile aleatoria (discreta), avente momento del secondo ordine. In tal caso, per la proposizione precedente,  $X$  ammette momento centrato di ordine 2. Esso si chiama *varianza* di  $X$  e si indica con  $\mathbf{Var}X$ . Si pone cioè

$$\mathbf{Var}X = \mathbf{E}[(X - \mathbf{E}[X])^2].$$

Osserviamo che la media di una variabile aleatoria non negativa è un numero non negativo (perché?), dunque ha senso considerare  $\sqrt{\mathbf{Var}X}$ ; a questa quantità si dà il nome di *deviazione standard* di  $X$ .

**Osservazione 3.1.19.** Se  $X$  ha densità  $p_X$ , dal Teorema 3.1.7 segue la formula

$$\mathbf{Var}X = \sum_x [(x - \mathbf{E}[X])^2] p_X(x). \quad (3.1.20)$$

Vediamo adesso le principali proprietà della varianza.

Dalla formula della definizione si ricava

$$\begin{aligned} \mathbf{Var}X &= \mathbf{E}[(X - \mathbf{E}[X])^2] = \mathbf{E}[X^2 - 2\mathbf{E}[X]X + \mathbf{E}^2[X]] \\ &= \mathbf{E}[X^2] - 2\mathbf{E}^2[X] + \mathbf{E}^2[X] = \mathbf{E}[X^2] - \mathbf{E}^2[X]; \end{aligned}$$

questa è una seconda forma con cui calcolare la varianza (speranza del quadrato meno quadrato della speranza).

Si ha ora la:

**Proposizione 3.1.21.** *Sia  $X$  una v. a. avente varianza finita e sia  $c$  una costante. Allora*

- (i)  $\mathbf{Var}(cX) = c^2 \mathbf{Var}X;$
- (ii)  $\mathbf{Var}(X + c) = \mathbf{Var}X;$
- (iii)  $\mathbf{Var}X = 0 \Leftrightarrow X = c$

( $c$  costante).

*Dimostrazione.* (i) Si ha

$$\begin{aligned} \mathbf{Var}(cX) &= \mathbf{E}[c^2 X^2] - (\mathbf{E}[cX])^2 \\ &= c^2 \mathbf{E}[X^2] - (c\mathbf{E}[X])^2 \\ &= c^2 \mathbf{E}[X^2] - c^2 \mathbf{E}^2[X] = c^2 \mathbf{Var}X. \end{aligned}$$

(ii) Si ha

$$\begin{aligned} \mathbf{Var}(X + c) &= \mathbf{E}[(X + c - \mathbf{E}[X + c])^2] = \mathbf{E}[(X + c - \mathbf{E}[X] - c)^2] \\ &= \mathbf{E}[(X - \mathbf{E}[X])^2] = \mathbf{Var}X. \end{aligned}$$

(iii) È immediato verificare che la varianza di una costante è nulla. Viceversa, supponiamo che sia  $\mathbf{Var}X = 0$ . Allora, per la relazione (3.1.20) si ha

$$0 = \mathbf{Var}X = \sum_x (x - \mathbf{E}[X])^2 p_X(x).$$

Nella somma precedente tutti gli addendi sono non negativi. Pertanto la loro somma può essere nulla se e solo se ciascun addendo è nullo. Ciò può accadere se  $p_X(x) = 0$  (e in tal caso  $x$  è un valore assunto da  $X$  con probabilità nulla) oppure se  $x - \mathbf{E}[X] = 0$ , cioè se il valore  $x$  coincide con il valore medio. Dunque l'unico valore assunto da  $X$  con probabilità non nulla è  $\mathbf{E}[X]$ , e  $X$  è dunque costante.

**Osservazione 3.1.22.** La proprietà (ii) della Proposizione 3.1.21 è di facile interpretazione: se trasliamo la variabile  $X$  della quantità  $c$ , viene traslata anche la media della stessa quantità, e dunque la dispersione attorno alla media rimane invariata.

Le proprietà stabilite nella Proposizione 3.1.21 costituiscono una prima (ma in realtà piuttosto vaga) giustificazione del fatto che la varianza di una variabile aleatoria  $X$  viene assunta come misura della dispersione di  $X$  attorno alla sua media.

Una giustificazione molto più convincente è fornita dal seguente risultato, che dà una stima in termini della varianza di  $X$  della probabilità che  $X$  si discosti dalla sua media per più di ogni valore prefissato.

**Proposizione 3.1.23 (Diseguaglianza di Tchebyceff).** *Sia  $X$  una v.a. avente media e varianza finite. Allora, per ogni  $\epsilon > 0$*

$$P(|Y - \mathbf{E}[X]| > \epsilon) \leq \frac{\mathbf{Var}X}{\epsilon^2}.$$

Per la dimostrazione di questo risultato, serve la

**Proposizione 3.1.24 (Diseguaglianza di Markov).** *Sia  $Y$  una v. a. avente media finita. Allora, per ogni  $\epsilon > 0$ ,*

$$P(|Y| > \epsilon) \leq \frac{\mathbf{E}[|Y|]}{\epsilon}.$$

*Dimostrazione (della diseg. di Markov).* Supporremo (ma non sarebbe necessario) che  $X$  sia discreta, con densità discreta  $p$ . Conviene dimostrare la tesi nella forma equivalente

$$\mathbf{E}[|Y|] \geq \epsilon P(|Y| > \epsilon).$$

Poniamo  $A = \{y \in \mathbb{R} : |y| > \epsilon\}$ . Ricordando la formula (2.3.6) (per il calcolo della legge di una v. a.) si ha allora

$$P(|Y| > \epsilon) = P(Y \in A) = \sum_{y \in A} p(y),$$

e quindi

$$\begin{aligned}\mathbf{E}[|Y|] &= \sum_{y \in \mathbb{R}} |y|p(y) = \sum_{y \in A} |y|p(y) + \sum_{y \in A^c} |y|p(y) \geq \sum_{y \in A} |y|p(y) \\ &\geq \epsilon \sum_{y \in A} p(y) = \epsilon P(|Y| > \epsilon).\end{aligned}\quad \square$$

Per dimostrare la disuguaglianza di Tchebyceff, basta osservare che

$$P(|X - \mathbf{E}[X]| > \epsilon) = P((X - \mathbf{E}[X])^2 > \epsilon^2)$$

e applicare la disuguaglianza di Markov alla v. a. (non negativa)  $Y = (X - \mathbf{E}[X])^2$ .

Siano  $X$  e  $Y$  due assegnate v.a. Ci interessa trovare una quantità che ci dia qualche informazione sul “grado di dipendenza” che sussiste tra  $X$  e  $Y$ . È naturale pretendere che, se  $X$  e  $Y$  sono indipendenti, tale quantità valga zero; ricordando allora la Proposizione 3.1.12, una possibilità è considerare la differenza

$$\mathbf{E}[XY] - \mathbf{E}[X]\mathbf{E}[Y].$$

In effetti questa sarà la nostra scelta, ma di nuovo abbiamo bisogno di qualche preliminare.

**Definizione 3.1.25.** Si chiama *covarianza* di  $X$  e  $Y$  la quantità (se finita)

$$\mathbf{Cov}(X, Y) := \mathbf{E}[(X - \mathbf{E}[X])(Y - \mathbf{E}[Y])].$$

Si pone evidentemente il problema di decidere in quali casi la covarianza di due variabili  $X$  e  $Y$  è finita; a questo proposito si ha il seguente risultato (che enunciamo senza dimostrazione):

**Proposizione 3.1.26.** Se  $X$  e  $Y$  hanno entrambe varianza finita, allora

$$|\mathbf{Cov}(X, Y)| \leq \sqrt{\mathbf{Var} X} \sqrt{\mathbf{Var} Y}.$$

Di conseguenza, la covarianza di  $X$  e  $Y$  è anch'essa finita.

Sia  $p_{X,Y}$  la densità congiunta della coppia  $(X, Y)$ . Dal Teorema 3.1.7 segue la formula

$$\mathbf{Cov}(X, Y) = \sum_{x,y} (x - \mathbf{E}[X])(y - \mathbf{E}[Y])p_{X,Y}(x, y).$$

Inoltre risulta

$$\begin{aligned}\mathbf{Cov}(X, Y) &= \mathbf{E}[(X - \mathbf{E}[X])(Y - \mathbf{E}[Y])] \\ &= \mathbf{E}[XY - X\mathbf{E}[Y] - Y\mathbf{E}[X] + \mathbf{E}[X]\mathbf{E}[Y]] \\ &= \mathbf{E}[XY] - \mathbf{E}[X]\mathbf{E}[Y] - \mathbf{E}[X]\mathbf{E}[Y] + \mathbf{E}[X]\mathbf{E}[Y] \\ &= \mathbf{E}[XY] - \mathbf{E}[X]\mathbf{E}[Y]\end{aligned}$$

(cioè la covarianza è uguale alla speranza del prodotto meno il prodotto delle speranze).

Se ne deduce che

$$\mathbf{Cov}(X, Y) = 0 \Leftrightarrow \mathbf{E}[XY] = \mathbf{E}[X]\mathbf{E}[Y].$$

Questa osservazione ci conduce alla:

**Definizione 3.1.27.** Due variabili aleatorie  $X$  e  $Y$  si dicono *non correlate* se  $\mathbf{Cov}(X, Y) = 0$ , o, equivalentemente, se  $\mathbf{E}[XY] = \mathbf{E}[X]\mathbf{E}[Y]$ .

Dalla Proposizione 3.1.12 si deduce allora:

**Proposizione 3.1.28.** *Due v. a. indipendenti con varianza finita sono non correlate.*

**Osservazione 3.1.29.** Fare attenzione al fatto che il viceversa della proposizione precedente è falso: esistono coppie di v.a. non correlate ma non indipendenti. Facciamo un esempio semplice. Sia  $X$  una v. a. tale che

$$P(X = -1) = P(X = 0) = P(X = 1) = \frac{1}{3}.$$

Poniamo poi  $Y = X^2$ . Allora la v. a.  $Y$  è tale che

$$P(Y = 0) = \frac{1}{3}, \quad P(Y = 1) = \frac{2}{3}.$$

Dato che  $X^3 = X$  e  $\mathbf{E}[X] = 0$  si ha

$$\mathbf{E}[XY] = \mathbf{E}[X^3] = \mathbf{E}[X] = 0,$$

ed anche  $\mathbf{E}[X]\mathbf{E}[Y] = 0$ .

Dunque  $X$  e  $Y$  sono non correlate. Verifichiamo adesso che esse non sono indipendenti. Si ha infatti

$$P(X = 1, Y = 0) = P(X = 1, X^2 = 0) = 0,$$

(poiché l'evento considerato è vuoto). Invece

$$P(X = 1)P(Y = 0) = \frac{1}{3} \times \frac{1}{3} = \frac{1}{9}.$$

Calcoliamo ora la varianza della somma di due variabili  $X$  e  $Y$ . Risulta

$$\begin{aligned} \mathbf{Var}(X + Y) &= \mathbf{E}\left[\left((X + Y) - \mathbf{E}[X + Y]\right)^2\right] \\ &= \mathbf{E}[(X + Y - \mathbf{E}[X] - \mathbf{E}[Y])^2] \\ &= \mathbf{E}[((X - \mathbf{E}[X]) + (Y - \mathbf{E}[Y]))^2] \\ &= \mathbf{E}[(X - \mathbf{E}[X])^2 + (Y - \mathbf{E}[Y])^2 + 2(X - \mathbf{E}[X])(Y - \mathbf{E}[Y])] \\ &= \mathbf{E}[(X - \mathbf{E}[X])^2] + \mathbf{E}[(Y - \mathbf{E}[Y])^2] + 2\mathbf{E}[(X - \mathbf{E}[X])(Y - \mathbf{E}[Y])] \\ &= \mathbf{Var}X + \mathbf{Var}Y + 2\mathbf{Cov}(X, Y). \end{aligned}$$

**Osservazione 3.1.30.** Se  $X$  e  $Y$  sono non correlate, allora

$$\mathbf{Var}(X + Y) = \mathbf{Var}X + \mathbf{Var}Y.$$

Segue dalla proposizione 3.1.28 che ciò è vero in particolare se  $X$  e  $Y$  sono tra loro indipendenti.

Calcoliamo ora la varianza delle densità fondamentali che conosciamo.

**3.1.31.** (i) *Densità bernoulliana.* Sia  $X$  una v. a. avente legge  $\mathcal{B}(1, p)$ . Allora, poiché  $X^2 = X$ , per il punto (3.1.13)(i) si ha

$$\mathbf{Var}X = \mathbf{E}[X^2] - \mathbf{E}^2[X] = \mathbf{E}[X] - \mathbf{E}^2[X] = p - p^2 = p(1 - p).$$

(ii) *Densità binomiale.* Sia  $X$  una v. a. avente legge  $\mathcal{B}(n, p)$ . Allora, come già sappiamo,  $X$  ha la stessa legge di

$$\sum_{i=1}^n X_i,$$

dove le  $X_i$  sono tra loro indipendenti ed hanno tutte legge  $\mathcal{B}(1, p)$ . Per l'Osservazione 3.1.30 e per il punto precedente si ha dunque

$$\mathbf{Var}X = \mathbf{Var}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \mathbf{Var}X_i = \sum_{i=1}^n p(1 - p) = np(1 - p).$$

(iii) *Densità di Poisson.* Sia  $X$  una v. a. avente densità di Poisson di parametro  $\lambda$ . Per il calcolo della varianza, adoperiamo anche in questo caso l'espressione

$$\mathbf{Var}X = \mathbf{E}[X^2] - \mathbf{E}^2[X],$$

cominciando dal calcolo di  $\mathbf{E}[X^2]$ . Risulta

$$\begin{aligned} \mathbf{E}[X^2] &= e^{-\lambda} \sum_{k=0}^{\infty} k^2 \frac{\lambda^k}{k!} = e^{-\lambda} \lambda \sum_{k=1}^{\infty} k \frac{\lambda^{k-1}}{(k-1)!} \\ &= e^{-\lambda} \lambda \sum_{k=0}^{\infty} (k+1) \frac{\lambda^k}{k!} = \lambda \mathbf{E}[X+1] \\ &= \lambda(\lambda+1) = \lambda^2 + \lambda. \end{aligned}$$

Di conseguenza

$$\mathbf{Var}X = \lambda^2 + \lambda - \lambda^2 = \lambda.$$

(iv) *Densità geometrica*. Sia infine  $X$  una v. a. geometrica di parametro  $p$ . Si ha

$$\mathbf{Var}X = \mathbf{E}[X^2] - \mathbf{E}^2[X] = \sum_{k=1}^{+\infty} k^2 p(1-p)^{k-1} - \frac{1}{p^2},$$

e si può far vedere (ma al solito non faremo i conti) che la serie qui sopra ha somma pari a  $\frac{2-p}{p^2}$ . Pertanto

$$\mathbf{Var}X = \frac{2-p}{p^2} - \frac{1}{p^2} = \frac{1-p}{p^2}.$$

## Variabili aleatorie assolutamente continue

---

### 4.1 Definizioni e prime proprietà

In questo paragrafo studieremo un'altra classe importante di variabili aleatorie.

Abbiamo visto che, se  $X$  è una v. a. *discreta* di densità  $p$ , allora, per ogni  $A \subseteq \mathbb{R}$ , si ha

$$P(X \in A) = \sum_{x \in A} p(x). \quad (4.1.1)$$

**Definizione 4.1.2.** Una v. a.  $X$  sullo spazio  $(\Omega, \mathcal{A}, P)$  si dice *assolutamente continua* se esiste una funzione  $f : \mathbb{R} \rightarrow \mathbb{R}^+$  integrabile su  $\mathbb{R}$  e tale che, per ogni sottoinsieme misurabile  $A$  di  $\mathbb{R}$ , risulti

$$P(X \in A) = \int_A f(x) dx. \quad (4.1.3)$$

In tal caso la funzione  $f$  si chiama *densità* di  $X$  (o si dice che  $X$  ammette  $f$  come densità).

*Notazione.* Come al solito, quando saranno in gioco più variabili  $X, Y$ , ecc., indicheremo le rispettive densità con  $f_X, f_Y$ , ecc.

**Osservazione 4.1.4.** Si noti l'analogia tra le formule (4.1.1) e (4.1.3). La seconda si può formalmente ottenere dalla prima sostituendo il simbolo di sommatoria con quello di integrale. Questa analogia si presenterà in altre formule, e anzi la useremo per ottenere direttamente dal caso discreto le formule per il caso assolutamente continuo.

**Osservazione 4.1.5.** Cambiando il valore di  $f$  su un insieme finito di punti, l'integrale in (4.1.3) non cambia. Ciò implica che la densità di una v. a.  $X$  assolutamente continua non è unica.



**Osservazione 4.1.6.** Dalla formula (4.1.3) si deduce in particolare che la funzione di ripartizione  $F$  di una v. a. assolutamente continua  $X$  assume la forma

$$F(t) = P(X \leq t) = P(X \in (-\infty, t]) = \int_{-\infty}^t f(x)dx.$$

Inoltre si ha

$$P(a < X \leq b) = P(X \in (a, b]) = \int_a^b f(x)dx.$$

E ancora

$$P(X = a) = P(X \in \{a\}) = \int_{\{a\}} f(x)dx = 0,$$

da cui segue che la v. a.  $X$  è continua.

Infine si ha (notare ancora l'analogia con il caso discreto)

$$\int_{-\infty}^{+\infty} f(x)dx = P(X \in \mathbb{R}) = P(\Omega) = 1;$$

(in particolare, una funzione positiva che non abbia integrale uguale a 1 su tutta la retta non può essere la densità di alcuna v. a.  $X$ ).

Tuttavia, a differenza che nel caso discreto,  $f$  può assumere anche valori maggiori di 1.

Estenderemo ora al caso assolutamente continuo alcune nozioni già studiate per il caso discreto. La definizione di indipendenza è identica a quella vista per il caso discreto (Definizione 2.6.1), ma in questo caso deve essere data parlando solo di sottoinsiemi misurabili di  $\mathbb{R}$ . Precisamente:

**Definizione 4.1.7.** Due v. a.  $X, Y$  si dicono tra loro *indipendenti* se, per ogni coppia  $A, B$  di sottoinsiemi *misurabili* di  $\mathbb{R}$ , risulta

$$P(X \in A, Y \in B) = P(X \in A)P(Y \in B).$$

**Osservazione 4.1.8.** È facile far vedere che due v. a.  $X$  e  $Y$  sono indipendenti se e solo se, per ogni coppia di numeri reali  $s, t$ , risulta

$$P(X \leq s, Y \leq t) = F_X(s)F_Y(t),$$

dove con  $F_X$  (risp.  $F_Y$ ) si indica la funzione di ripartizione di  $X$  (risp.  $Y$ ).

La Proposizione 2.6.9 vale anche per v.a. assolutamente continue, purché le funzioni  $\phi$  e  $\psi$  siano misurabili.

Sia  $X$  una v. a. assolutamente continua avente densità  $f$ .

**Definizione 4.1.9.** Si dice che  $X$  ha *speranza matematica finita* se

$$\int_{-\infty}^{+\infty} |x|f(x)dx < +\infty;$$

in tal caso si chiama *speranza matematica* di  $X$  il numero

$$\mathbf{E}[X] := \int_{-\infty}^{+\infty} xf(x)dx.$$

**Osservazione 4.1.10.** Sia  $X$  una v. a. assolutamente continua e sia  $\phi : \mathbb{R} \rightarrow \mathbb{R}$  una funzione misurabile. Allora, come sappiamo (si vedano pp. 29–30) anche la funzione composta  $U = \phi(X)$  è una v. a. Si può mostrare (confrontare con il teorema 3.1.7) che  $U$  ha speranza finita se e solo se

$$\int_{-\infty}^{+\infty} |\phi(x)|f(x)dx < +\infty$$

e, in tal caso, si ha

$$\mathbf{E}[U] = \int_{-\infty}^{+\infty} \phi(x)f(x)dx.$$

In particolare, se finito, il momento  $k$ -esimo di  $X$  è dato da

$$\mathbf{E}[X^k] = \int_{-\infty}^{+\infty} x^k f(x)dx.$$

**Definizione 4.1.11.** Se  $X$  ha momento secondo  $\mathbf{E}[X^2]$  finito allora si definisce, come nel caso discreto,

$$\mathbf{Var}X := \mathbf{E}[(X - \mathbf{E}[X])^2].$$

Per l'Osservazione 4.1.10 si ha

$$\mathbf{Var}X = \int_{-\infty}^{+\infty} (x - \mathbf{E}[X])^2 f(x)dx.$$

Per la speranza e la varianza delle variabili aleatorie assolutamente continue valgono tutte le proprietà che abbiamo visto nel caso discreto, e non le ridimosteremo. In particolare si ha anche adesso

$$\mathbf{Var}X = E[X^2] - E^2[X] = \int_{-\infty}^{+\infty} x^2 f(x)dx - \left( \int_{-\infty}^{+\infty} xf(x)dx \right)^2.$$

Si noti di nuovo l'analogia delle definizioni: il simbolo di sommatoria che usavamo nel caso discreto è stato qui sostituito dal simbolo di integrale. Molte delle proprietà della speranza e della varianza per il caso assolutamente continuo si

possono formalmente provare semplicemente operando la stessa sostituzione nelle dimostrazioni che abbiamo già visto.

Infine, come accadeva nel caso discreto, è possibile provare che funzioni di gruppi separati di v.a. indipendenti sono v. a. tra loro indipendenti (cioè vale anche adesso la Proposizione 2.6.11, a condizione che nel suo enunciato sia soppresso l'aggettivo “discreto” e  $\varphi$  e  $\psi$  siano misurabili).

Nel prossimo paragrafo studieremo le principali densità di v.a. assolutamente continue.

## 4.2 Alcune densità di variabili aleatorie assolutamente continue

**4.2.1. Le leggi uniformi.** Si sceglie un punto *a caso* nell'intervallo  $[a, b]$  della retta  $\mathbb{R}$ . Indichiamo con  $X$  l'ascissa del punto scelto.

L'espressione *a caso* va interpretata nel senso che la probabilità che  $X$  appartenga ad un intervallo  $(s, t] \subset [a, b]$  è direttamente proporzionale alla lunghezza di  $(s, t]$ , e quindi, poiché la probabilità che  $X$  appartenga ad  $[a, b]$  è 1, dovremo richiedere che

$$P(s < X \leq t) = P(X \in (s, t]) = \frac{t - s}{b - a}.$$

È facile allora vedere che la funzione di ripartizione di  $X$  è

$$F(t) = \begin{cases} 0 & \text{se } t < a \\ \frac{t-a}{b-a} & \text{se } a \leq t \leq b \\ 1 & \text{se } t > b \end{cases}$$

e che una densità è

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{se } a \leq x \leq b \\ 0 & \text{altrimenti.} \end{cases}$$

Disegnare i grafici di  $F$  e di  $f$ .

**Definizione 4.2.2.** La legge qui descritta si chiama *legge uniforme* (sull'intervallo  $[a, b]$ ) e si indica con il simbolo  $\mathcal{U}([a, b])$ ; ogni variabile aleatoria che ammetta questa legge si dice *uniforme* (su  $[a, b]$ ).

**Esercizio 4.2.3.** Calcolare la media e la varianza della legge  $\mathcal{U}([a, b])$ .

**4.2.4. La legge esponenziale.** Sia  $\lambda$  un numero reale positivo assegnato,  $X$  una variabile aleatoria. Si dice che  $X$  ha *legge esponenziale* di parametro  $\lambda$  (il simbolo è  $\mathcal{E}(\lambda)$ ) se ammette come densità la funzione

$$f(x) = \begin{cases} 0 & \text{per } x < 0 \\ \lambda e^{-\lambda x} & \text{per } x \geq 0. \end{cases}$$

**Esercizio 4.2.5.** Verificare che  $\int_{-\infty}^{+\infty} f(x)dx = 1$ .

È facile vedere che la funzione di ripartizione della legge  $\mathcal{E}(\lambda)$ :

$$F(t) = \begin{cases} 0 & \text{per } t < 0 \\ 1 - e^{-\lambda t} & \text{per } t \geq 0. \end{cases}$$

Ogni variabile  $X$  avente legge esponenziale verifica la seguente relazione (la cui prova è lasciata per esercizio al lettore):

$$P(X > t + s | X > t) = P(X > s), \quad (4.2.6)$$

dove  $s$  e  $t$  sono due numeri positivi assegnati.

La relazione (4.2.6) viene detta *proprietà di assenza di memoria* o anche *proprietà di assenza di usura*. Per capire il significato di queste espressioni, immaginiamo che  $X$  rappresenti la durata di una lampadina, messa in funzione all'istante 0. Allora la (4.2.6) si interpreta così: sapendo che la lampadina non si è guastata fino all'istante  $t$ , la probabilità che essa duri più di ulteriori  $s$  istanti è uguale alla probabilità che la lampadina, messa in funzione all'istante 0, non si guasti fino all'istante  $s$ . In altri termini, la lampadina non è soggetta all'usura causata dai primi  $t$  istanti di funzionamento.

**Esercizio 4.2.7.** Calcolare la media e la varianza della legge  $\mathcal{E}(\lambda)$ .

**4.2.8. Le leggi di Weibull.** Siano  $\alpha > 0$  e  $\lambda > 0$  due numeri fissati, e consideriamo la funzione

$$f(x) = \begin{cases} \lambda \alpha x^{\alpha-1} e^{-\lambda x^\alpha} & \text{per } x > 0 \\ 0 & \text{per } x \leq 0. \end{cases}$$

È facile vedere che  $f$  è una densità (cioè è non negativa, integrabile su  $\mathbb{R}$  e  $\int_{-\infty}^{+\infty} f(x)dx = 1$ ). Osserviamo anche che, per  $\alpha = 1$ , si ottiene la densità esponenziale di parametro  $\lambda$  di cui abbiamo già parlato. Nel caso generale,  $f$  viene chiamata *densità di Weibull* (di parametri  $\alpha$  e  $\lambda$ ).

Sia  $X$  una v. a. avente  $f$  come densità. Ci interessa studiare la quantità  $P(X > t + s | X > t)$  per  $s > 0$  e  $t > 0$  (la abbiamo già incontrata parlando della densità esponenziale). In questo caso si ha

$$\begin{aligned} P(X > t + s | X > t) &= \frac{\int_{t+s}^{+\infty} \lambda \alpha x^{\alpha-1} e^{-\lambda x^\alpha} dx}{\int_t^{+\infty} \lambda \alpha x^{\alpha-1} e^{-\lambda x^\alpha} dx} \\ &= \frac{-e^{-\lambda x^\alpha} \Big|_{t+s}^{+\infty}}{-e^{-\lambda x^\alpha} \Big|_t^{+\infty}} \\ &= \frac{e^{-\lambda(t+s)^\alpha}}{e^{-\lambda t^\alpha}} = e^{-\lambda[(t+s)^\alpha - t^\alpha]}. \end{aligned}$$

Dunque, per ogni  $s$  fissato, la funzione  $t \mapsto P(X > t + s | X > t)$  è crescente (risp. costante, decrescente) quando la funzione  $t \mapsto (t + s)^\alpha - t^\alpha$  è decrescente (risp. costante, crescente). Ciò accade quando

$$\frac{d}{dt}[(t + s)^\alpha - t^\alpha] = \alpha[(t + s)^{\alpha-1} - t^{\alpha-1}] < 0$$

(risp.  $= 0$ ,  $> 0$ ) ovvero quando  $\alpha < 1$  (risp.  $= 1$ ,  $> 1$ ). Ricapitolando

$$t \mapsto P(X > t + s | X > t) \text{ è } \begin{cases} \text{crescente per } \alpha < 1 \\ \text{costante per } \alpha = 1 \\ \text{decrescente per } \alpha > 1. \end{cases}$$

Supponiamo che  $X$  rappresenti la durata di una lampadina soggetta ad usura. È naturale allora scegliere per  $X$  una densità per la quale la funzione  $t \mapsto P(X > t + s | X > t)$  sia decrescente, e per tale motivo in questo caso viene usata una variabile  $X$  avente densità di Weibull di parametro  $\alpha > 1$ .

Una situazione in cui sarebbe invece indicata una densità di Weibull con parametro  $\alpha < 1$  è ad esempio quella in cui  $X$  rappresenta la durata di vita di un neonato (è noto che la probabilità di mortalità infantile è massima nei primi giorni vita del bambino, per poi andare a diminuire mano a mano che i giorni passano).

**4.2.9. Le leggi Gamma.** Per introdurre le leggi Gamma bisogna prima definire la *funzione Gamma di Eulero*. Sia  $\alpha > 0$  un numero reale assegnato. Utilizzando noti criteri di confronto per gli integrali impropri, è abbastanza facile vedere che la quantità

$$\Gamma(\alpha) := \int_0^{+\infty} x^{\alpha-1} e^{-x} dx$$

è finita (calcoli per esercizio).

**Definizione 4.2.10.** La funzione  $\alpha \mapsto \Gamma(\alpha)$ , definita per  $\alpha > 0$ , è la *funzione Gamma di Eulero*.

**Esercizio 4.2.11.** Utilizzando la definizione, calcolare  $\Gamma(1)$ ,  $\Gamma(2)$ ,  $\Gamma(3)$ .

**4.2.12. Alcune proprietà importanti.** (i) Per ogni  $\alpha > 0$  si ha

$$\Gamma(\alpha + 1) = \alpha \Gamma(\alpha).$$

Infatti, integrando per parti, si ha

$$\Gamma(\alpha + 1) = \int_0^{+\infty} x^\alpha e^{-x} dx = -x^\alpha e^{-x} \Big|_0^{+\infty} + \alpha \int_0^{+\infty} x^{\alpha-1} e^{-x} dx = \alpha \Gamma(\alpha),$$

dato che  $\lim_{x \rightarrow +\infty} x^\alpha e^{-x} = 0$ .

(ii) Applicando il punto (i) al caso  $\alpha = n$  (intero  $\geq 1$ ), si trova la relazione  $\Gamma(n+1) = n\Gamma(n)$ , e, dato che  $\Gamma(1) = 1$  (per calcolo diretto), si ottiene  $\Gamma(2) = 1\Gamma(1) = 1!$ ,  $\Gamma(3) = 2\Gamma(2) = 2 \cdot 1 = 2!$ ,  $\Gamma(4) = 3\Gamma(3) = 3 \cdot 2! = 3!$ , ecc. Per induzione su  $n$ , si trova dunque che  $\Gamma(n) = (n-1)!$  (dimostrarlo rigorosamente).

(iii) Vedremo più tardi (Esempio 4.6.6) che  $\Gamma(\frac{1}{2}) = \sqrt{\pi}$ .

Siano ora  $\alpha > 0$  e  $\lambda > 0$  due numeri reali assegnati, e consideriamo la funzione  $f: \mathbb{R} \rightarrow \mathbb{R}^+$  definita

$$f(x) = \begin{cases} \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} & \text{per } x > 0 \\ 0 & \text{per } x \leq 0. \end{cases}$$

È facile vedere che  $f$  è una densità di probabilità. Infatti

$$\int_{-\infty}^{+\infty} f(x) dx = \int_0^{+\infty} \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} dx,$$

e, con il cambiamento di variabile  $y = \lambda x$ , l'integrale qui sopra si trasforma in

$$\int_0^{+\infty} \frac{1}{\Gamma(\alpha)} y^{\alpha-1} e^{-y} dy = 1,$$

per la definizione della funzione Gamma di Eulero.

**Definizione 4.2.13.** La densità sopra definita si chiama *densità Gamma di parametri  $\alpha$  e  $\lambda$* , e si indica con il simbolo  $\Gamma(\alpha, \lambda)$ . La legge individuata da tale densità si chiama legge  $\Gamma(\alpha, \lambda)$ .

**Osservazione 4.2.14.** La famiglia delle densità Gamma comprende la densità esponenziale (caso particolare, per  $\alpha = 1$ ).

**Esercizio 4.2.15.** Sia  $X$  una v. a. avente densità  $\Gamma(\alpha, \lambda)$ . Mostrare che  $X$  ha speranza finita e

$$\mathbf{E}[X] = \frac{\alpha}{\lambda}.$$

Dato che  $X$  è una v. a. positiva (perché?), possiamo verificare che  $X$  ha speranza finita contemporaneamente al calcolo del valore stesso di  $\mathbf{E}[X]$ . Per definizione di media si ha

$$\begin{aligned} \mathbf{E}[X] &= \int_{-\infty}^{+\infty} x f(x) dx = \int_0^{+\infty} \frac{\lambda^\alpha}{\Gamma(\alpha)} x^\alpha e^{-\lambda x} dx \\ &= \frac{\Gamma(\alpha+1)}{\lambda \Gamma(\alpha)} \int_0^{+\infty} \frac{\lambda^{\alpha+1}}{\Gamma(\alpha+1)} x^{(\alpha+1)-1} e^{-\lambda x} dx. \end{aligned}$$

Nella funzione integranda qui sopra si riconosce la densità  $\Gamma(\alpha + 1, \lambda)$ , e quindi l'integrale vale 1. Pertanto, continuando il calcolo e utilizzando la Proprietà 4.2.12 (i), si trova

$$\mathbf{E}[X] = \frac{\Gamma(\alpha + 1)}{\lambda \Gamma(\alpha)} = \frac{\alpha \Gamma(\alpha)}{\lambda \Gamma(\alpha)} = \frac{\alpha}{\lambda}.$$

**Osservazione 4.2.16.** Per calcolare l'integrale, abbiamo scritto le cose in modo da trasformarlo nell'integrale su  $\mathbb{R}$  di una densità di probabilità. A questo punto il conto diventa immediato, dato che ogni densità ha integrale 1 su  $\mathbb{R}$ . Questo trucco di calcolo è utilizzato frequentemente.

**Esercizio 4.2.17.** Sia  $X$  una v. a. avente densità  $\Gamma(\alpha, \lambda)$ :

- (i) utilizzando un accorgimento analogo a quello illustrato nell'osservazione precedente, calcolare  $\mathbf{E}[X^2]$ ;
- (ii) calcolare  $\mathbf{Var}X$ .

La famiglia delle leggi Gamma possiede un'importante proprietà di stabilità, enunciata nel risultato seguente, che non dimostreremo.

**Teorema 4.2.18.** *Siano  $X$  e  $Y$  due v.a. tra loro indipendenti, aventi legge  $\Gamma(\alpha_1, \lambda)$  e  $\Gamma(\alpha_2, \lambda)$  rispettivamente. Allora la v.a.  $U = X + Y$  ha legge  $\Gamma(\alpha_1 + \alpha_2, \lambda)$ .*

**Osservazione 4.2.19.** Il risultato precedente si estende subito al caso della somma di più di due variabili: se, per esempio,  $X$ ,  $Y$  e  $Z$  sono tre v. a. aventi legge  $\Gamma(\alpha_1, \lambda)$ ,  $\Gamma(\alpha_2, \lambda)$  e  $\Gamma(\alpha_3, \lambda)$  rispettivamente, allora  $U = X + Y$  ha legge  $\Gamma(\alpha_1 + \alpha_2, \lambda)$  per il Teorema (4.2.18) ed è indipendente da  $Z$  (perché funzione del vettore  $(X, Y)$ , che è indipendente da  $Z$ ).

Dunque alla coppia  $U, Z$  si può applicare di nuovo il Teorema 4.2.18 e si trova che  $(X + Y) + Z = U + Z$  ha legge  $\Gamma((\alpha_1 + \alpha_2) + \alpha_3, \lambda) = \Gamma(\alpha_1 + \alpha_2 + \alpha_3, \lambda)$ . Il risultato generale si ottiene per induzione (svolgere i passaggi per esercizio).

**Osservazione 4.2.20.** Il teorema precedente è falso in generale se le v.a.  $X$  e  $Y$  non sono indipendenti. Sia ad esempio  $X$  una v.a. di legge  $\Gamma(\alpha, \lambda)$ , e poniamo  $Y = X$ . Allora  $X$  e  $Y$  non sono indipendenti (perché?). Verificare per esercizio che  $X + Y = 2X$  ha legge  $\Gamma(\alpha, \lambda/2)$  (e non  $\Gamma(2\alpha, \lambda)$ , come richiederebbe il Teorema 4.2.18, se fosse vero anche in questo caso).

Nel prossimo paragrafo parleremo delle principali densità che compaiono in statistica.

### 4.3 Le principali densità della statistica; il teorema Limite Centrale; il teorema di Cochran

**4.3.1. Le densità normali.** Siano  $\mu \in \mathbb{R}$  e  $\sigma \in \mathbb{R}^+$  due numeri assegnati. Si chiama *densità normale* (o *gaussiana*) di parametri  $\mu$  e  $\sigma^2$  la funzione

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

Il fatto che  $f$  sia una densità (cioè che abbia integrale 1 sulla retta reale) è conseguenza di alcune nozioni che non sono alla nostra portata, e lo assumeremo dunque senza dimostrazione (si ricordi che le funzioni del tipo di  $x \mapsto e^{-kx^2}$  non hanno una primitiva esprimibile per via elementare).

La legge di cui  $f$  è una densità si indica con il simbolo  $\mathcal{N}(\mu, \sigma^2)$  (il significato dei parametri  $\mu$  e  $\sigma^2$  sarà chiaro fra poco); ogni variabile aleatoria avente tale legge si chiama variabile aleatoria *normale* o *gaussiana*.

Nella famiglia delle leggi normali, sopra definite, particolare rilevanza ha la  $\mathcal{N}(0, 1)$  (cioè la legge di parametri  $\mu = 0$  e  $\sigma^2 = 1$ ). Scriviamone esplicitamente la densità:

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

Appunto per la sua importanza, questa legge viene detta *normale standard* (o *gaussiana standard*).

Una prima proprietà della famiglia delle leggi normali è la sua stabilità rispetto alle trasformazioni affini; precisamente si ha:

**Proposizione 4.3.2.** *Sia  $X$  una variabile aleatoria avente legge gaussiana  $\mathcal{N}(m, s^2)$ ; siano inoltre  $a \neq 0$  e  $b$  due numeri reali. Allora la variabile aleatoria  $Y = aX + b$  ha legge  $\mathcal{N}(am + b, a^2s^2)$ .*

*Dimostrazione.* Per la definizione di densità (caso assolutamente continuo) e poiché la funzione di ripartizione individua la legge della variabile, basta verificare che

$$P(Y \leq y) = \int_{-\infty}^y \frac{1}{\sqrt{2\pi}|a|s} \exp\left(-\frac{(x - am - b)^2}{2a^2s^2}\right) dx.$$

Se  $a > 0$ , si ha

$$\begin{aligned} P(Y \leq y) &= P(aX + b \leq y) = P\left(X \leq \frac{y-b}{a}\right) \\ &= \int_{-\infty}^{\frac{y-b}{a}} \frac{1}{\sqrt{2\pi}s} \exp\left(-\frac{(t-m)^2}{2s^2}\right) dt \\ &= \int_{-\infty}^y \frac{1}{\sqrt{2\pi}|a|s} \exp\left(-\frac{(x-am-b)^2}{2a^2s^2}\right) dx, \end{aligned}$$



dove l'ultima eguaglianza si ottiene effettuando il cambiamento di variabile

$$t = \frac{x - b}{a},$$

(dato che che  $a = |a|$ ). Nel caso  $a < 0$  si ha invece (ricordando che  $|a| = -a$ )

$$\begin{aligned} P(Y \leq y) &= P(aX + b \leq y) = P\left(X \geq \frac{y - b}{a}\right) \\ &= \int_{\frac{y-b}{a}}^{+\infty} \frac{1}{\sqrt{2\pi}s} \exp\left(-\frac{(t-m)^2}{2s^2}\right) dt \\ &= \int_y^{-\infty} \frac{1}{\sqrt{2\pi}as} \exp\left(-\frac{(x-am-b)^2}{2a^2s^2}\right) dx \\ &= \int_{-\infty}^y \frac{1}{\sqrt{2\pi}|a|s} \exp\left(-\frac{(x-am-b)^2}{2a^2s^2}\right) dx. \quad \square \end{aligned}$$

Dalla precedente proposizione si deduce:

**Corollario 4.3.3.** *Siano  $\mu \in \mathbb{R}$  e  $\sigma \in \mathbb{R}^+$  due numeri assegnati.*

- (i) se  $X \sim \mathcal{N}(0, 1)$ , allora  $Y := \sigma X + \mu \sim \mathcal{N}(\mu, \sigma^2)$ ;
- (ii) se  $X \sim \mathcal{N}(\mu, \sigma^2)$ , allora  $Y := (X - \mu)/\sigma \sim \mathcal{N}(0, 1)$ .

*Dimostrazione.* Per dimostrare il punto (i) si applica a  $X$  la proposizione precedente con  $a = \sigma$ ,  $b = \mu$ ,  $m = 0$ ,  $s = 1$ .

Per il punto (ii) si applica a  $X$  la stessa proposizione, questa volta con  $a = 1/\sigma$ ,  $b = -\mu/\sigma$ ,  $m = \mu$ ,  $s = \sigma$ .  $\square$

Calcoliamo ora la speranza e la varianza di una variabile aleatoria  $X$  avente legge  $\mathcal{N}(\mu, \sigma^2)$ , cominciando dal caso particolare  $\mu = 0$ ,  $\sigma = 1$ .

Sia dunque  $X \sim \mathcal{N}(0, 1)$ . Si ha prima di tutto, per noti risultati di analisi

$$\begin{aligned} \int_{-\infty}^{+\infty} |x| \frac{1}{\sqrt{2\pi}} \exp(-x^2/2) dx &= 2 \int_0^{+\infty} x \frac{1}{\sqrt{2\pi}} \exp(-x^2/2) dx \\ &\leq C \int_0^{+\infty} \frac{1}{x^2} dx < +\infty, \end{aligned}$$

(dove la prima eguaglianza segue dal fatto che la funzione  $x \mapsto |x|e^{-x^2/2}$  è pari), e dunque  $X$  ha speranza matematica finita. Inoltre, poiché la funzione  $x \mapsto x \exp(-x^2/2)$  è dispari, si ha

$$\begin{aligned} \mathbf{E}[X] &= \int_{-\infty}^{+\infty} x \frac{1}{\sqrt{2\pi}} \exp(-x^2/2) dx \\ &= \int_{-\infty}^0 x \frac{1}{\sqrt{2\pi}} \exp(-x^2/2) dx + \int_0^{+\infty} x \frac{1}{\sqrt{2\pi}} \exp(-x^2/2) dx \end{aligned}$$

$$= - \int_0^{+\infty} x \frac{1}{\sqrt{2\pi}} \exp(-x^2/2) dx + \int_0^{+\infty} x \frac{1}{\sqrt{2\pi}} \exp(-x^2/2) dx = 0.$$

dove nel penultimo passaggio si è usato il cambiamento di variabili  $y = -x$  (dettagli per esercizio). Alternativamente il conto può essere svolto così:

$$\mathbf{E}[X] = \int_{-\infty}^{+\infty} x \frac{1}{\sqrt{2\pi}} \exp(-x^2/2) dx = \frac{1}{\sqrt{2\pi}} \exp(-x^2/2) \Big|_{-\infty}^{+\infty} = 0.$$

Passiamo ora al calcolo della varianza (che, poiché  $\mathbf{E}[X] = 0$ , in questo caso coincide con  $\mathbf{E}[X^2]$ ).

Per verificare che

$$\int_{-\infty}^{+\infty} x^2 \frac{1}{\sqrt{2\pi}} \exp(-x^2/2) dx < +\infty$$

si ragiona in modo analogo a quanto fatto sopra per la speranza di  $X$ . Passiamo dunque al calcolo vero e proprio:

$$\begin{aligned} \mathbf{Var}X &= \mathbf{E}[X^2] = \int_{-\infty}^{+\infty} x^2 \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx \\ &= \int_{-\infty}^{+\infty} x \left( x \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) \right) dx \\ &= \int_{-\infty}^{+\infty} x \frac{d}{dx} \left( -\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) \right) dx \\ &= -x \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) \Big|_{-\infty}^{+\infty} - \int_{-\infty}^{+\infty} -\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx \\ &= 0 + \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx = 1. \end{aligned}$$

Qui si è usato il fatto che la funzione  $x \mapsto \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$  è una densità, e dunque ha integrale uguale a 1 su  $\mathbb{R}$ . (Si noti inoltre che in questo caso il calcolo non può essere dello stesso tipo di quello eseguito nel caso della speranza di  $X$  perché la funzione  $x \mapsto x^2 e^{-x^2/2}$  non ha una primitiva esprimibile per via elementare: infatti, in generale, le funzioni del tipo  $x \mapsto x^n e^{-kx^2}$ ,  $n \in \mathbb{N}$  hanno una primitiva esprimibile per via elementare se e solo se  $n$  è dispari).

A questo punto siamo in grado di calcolare  $\mathbf{E}[X]$  e  $\mathbf{Var}X$  anche per una variabile aleatoria  $X$  avente legge  $\mathcal{N}(\mu, \sigma^2)$ . Per far questo usiamo il punto (ii) del Corollario 4.3.3, che ci assicura che la variabile aleatoria definita da

$$Y = \frac{X - \mu}{\sigma}$$

ha legge  $\mathcal{N}(0, 1)$  (e dunque, in particolare, si ha  $\mathbf{E}[Y] = 0$ ,  $\mathbf{Var}Y = 1$ ). D'altra parte si può scrivere

$$X = \sigma Y + \mu,$$

e quindi, per le note proprietà della speranza e della varianza, risulta

$$\begin{aligned}\mathbf{E}[X] &= \sigma \mathbf{E}[Y] + \mu = \mu; \\ \mathbf{Var} X &= \sigma^2 \mathbf{Var} Y = \sigma^2.\end{aligned}$$

Il teorema che segue, che enunciamo senza dimostrazione, descrive il comportamento delle variabili aleatorie normali rispetto alla somma:

**Teorema 4.3.4.** *Siano  $X_1, X_2, \dots, X_n$   $n$  variabili aleatorie indipendenti, tali che, per ogni  $i$ , con  $i = 1, 2, \dots, n$  si abbia*

$$X_i \sim \mathcal{N}(\mu_i, \sigma_i^2).$$

*Allora la variabile aleatoria*

$$Z = X_1 + X_2 + \dots + X_n$$

*ha legge  $\mathcal{N}(\sum_{i=1}^n \mu_i, \sum_{i=1}^n \sigma_i^2)$ .*

**Osservazione 4.3.5.** La tesi del teorema può non essere vera senza l'ipotesi di indipendenza: se ad esempio  $X_1$  è una variabile normale standard e si prende  $X_2 = X_1$ , allora  $X_1$  e  $X_2$  non sono indipendenti (perché?), e la loro somma  $Z$ , essendo uguale a  $2X_1$ , ha legge  $\mathcal{N}(0, 4)$  per la Proposizione 4.3.2, e non  $\mathcal{N}(0, 2)$  (come invece dovrebbe essere se valesse il teorema qui sopra).

**Esempio 4.3.6.** Siano  $X_1, X_2, \dots, X_n$   $n$  variabili indipendenti e tutte con la stessa legge (non necessariamente gaussiana!). Poniamo

$$\overline{X} = \frac{X_1 + X_2 + \dots + X_n}{n}.$$

La variabile aleatoria  $\overline{X}$  viene detta *media campionaria* ed è di uso comune in statistica.

Come applicazione (utile in statistica) del Teorema 4.3.4 e degli altri risultati visti sopra, calcoleremo la legge di  $\overline{X}$  nel caso particolare in cui le  $X_i$ ,  $i = 1, 2, \dots, n$  abbiano tutte legge  $\mathcal{N}(\mu, \sigma^2)$ .

Posto  $Z = X_1 + X_2 + \dots + X_n$ , per il teorema precedente si ha

$$Z \sim \mathcal{N}(n\mu, n\sigma^2);$$

dunque, per la Proposizione 4.3.2 applicata a  $Z$  (con  $a = 1/n$  e  $b = 0$ ) si ha

$$\overline{X} \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right).$$

Come conseguenza, per il Corollario 4.3.3 (ii) si deduce che la variabile aleatoria centrata (= con media nulla) e ridotta (= con varianza unitaria)

$$\frac{\overline{X} - \mu}{\sqrt{\sigma^2/n}} = \sqrt{n} \frac{\overline{X} - \mu}{\sigma} = \frac{X_1 + X_2 + \dots + X_n - n\mu}{\sigma\sqrt{n}}$$

ha legge normale standard.

*Notazione.* D'ora in avanti la funzione di ripartizione della legge  $\mathcal{N}(0, 1)$  verrà indicata con il simbolo  $\Phi$ ; abbiamo cioè

$$\Phi(t) = \int_{-\infty}^t \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx, \quad t \in \mathbb{R}.$$

Della funzione  $\Phi$  qui definita parleremo in modo più approfondito nel Paragrafo 4.4.

Abbiamo ora visto che, se si hanno  $n$  variabili indipendenti  $X_1, X_2, \dots, X_n$  tutte con legge  $\mathcal{N}(\mu, \sigma^2)$  e a partire da esse si costruisce la variabile aleatoria centrata e ridotta associata alla loro media campionaria, cioè

$$S_n^* = \frac{X_1 + X_2 + \dots + X_n - n\mu}{\sigma\sqrt{n}},$$

la variabile  $S_n^*$  risulta avere legge normale standard. Una delle più importanti proprietà delle leggi normali è contenuta nel teorema che segue, che può considerarsi uno dei teoremi fondamentali del Calcolo delle Probabilità:

**Teorema Limite Centrale 4.3.7 (CLT nella letteratura di lingua anglosassone).** *Sia  $(X_n)_{n \geq 1}$  una successione di variabili aleatorie indipendenti e tutte con la stessa legge (si abbrevia con la sigla: i.i.d. = indipendenti identicamente distribuite, cioè indipendenti e tutte con la stessa legge), aventi media  $\mu$  e varianza  $\sigma^2$  finite. Per ogni intero  $n \geq 1$  poniamo*

$$S_n^* = \frac{X_1 + X_2 + \dots + X_n - n\mu}{\sigma\sqrt{n}}.$$

*Allora, per ogni  $x \in \mathbb{R}$  si ha*

$$\lim_{n \rightarrow \infty} P(S_n^* \leq x) = \Phi(x).$$

**Osservazione 4.3.8.** Si ha cioè il fatto (abbastanza sorprendente!) che la legge di  $S_n^*$  tende ad avvicinarsi alla legge normale standard *qualunque sia la legge di partenza* delle variabili  $(X_n)$  (non necessariamente normali, dunque, come accadeva nell'Esempio 4.3.6, e neppure necessariamente con densità!).

Qui di seguito vedremo un'importante applicazione del TLC.

**4.3.9. L'approssimazione normale.** Sia  $(X_n)$  una successione di variabili aleatorie i.i.d. Un problema che si incontra frequentemente è il calcolo della legge della somma delle prime  $n$  variabili della successione:

$$S_n = X_1 + X_2 + \dots + X_n.$$

Questo problema è spesso di non facile soluzione. Tuttavia, se  $n$  è sufficientemente grande, e ci si accontenta di un valore approssimato (in genere basta  $n > 30$  per avere un'approssimazione soddisfacente) si può procedere così:

$$P(S_n \leq x) = P\left(S_n^* \leq \frac{x - n\mu}{\sigma\sqrt{n}}\right) \simeq \Phi\left(\frac{x - n\mu}{\sigma\sqrt{n}}\right),$$

per il TLC; a questo punto si usano le tavole della legge normale, delle quali parleremo fra poco.

**Osservazione 4.3.10.** C'è un errore teorico nel modo in cui abbiamo usato il TLC per derivare la formula dell'approssimazione normale. Qual è? Questo errore può peraltro essere eliminato ricorrendo a risultati teorici più profondi di quelli che possiamo trattare in questi appunti. L'approssimazione normale può quindi essere utilizzata con tranquillità.

Introdurremo adesso altre due densità di uso comune in statistica.

**Definizione 4.3.11.** Siano  $X_1, X_2, \dots, X_n$   $n$  variabili aleatorie indipendenti e tutte con densità normale standard. Poniamo

$$Y = X_1^2 + X_2^2 + \dots + X_n^2.$$

La densità di  $Y$  viene chiamata *densità del chi quadrato a  $n$  gradi di libertà*; essa si indicherà con il simbolo  $\chi^2(n)$ .

Vedremo in seguito (Esempio 4.6.6) che, se  $X$  è una v. a. avente legge  $\mathcal{N}(0, 1)$ , la legge di  $X^2$  è una  $\Gamma(\frac{1}{2}, \frac{1}{2})$  (e dunque, in base alla definizione precedente, la legge  $\Gamma(\frac{1}{2}, \frac{1}{2})$  non è altro che una  $\chi^2(1)$ ). D'altra parte, sommando  $n$  v.a. indipendenti di legge  $\Gamma(\frac{1}{2}, \frac{1}{2})$ , per il Teorema 4.2.18 si ottiene una  $\Gamma(\frac{n}{2}, \frac{1}{2})$  che dunque, sempre per la Definizione 4.3.11, coincide con la  $\chi^2(n)$ .

Pertanto, come per tutte le leggi Gamma, della densità della  $\chi^2(n)$  ( $= \Gamma(\frac{n}{2}, \frac{1}{2})$ ) esiste un'espressione esplicita (scriverla!), che però per i nostri scopi non sarà necessaria. Osserviamo soltanto che la variabile aleatoria  $Y$  della Definizione 4.3.11 è positiva (in quanto somma di quadrati), e dunque la densità del  $\chi^2(n)$  è identicamente nulla sulla semiretta negativa di  $\mathbb{R}$  (spiegare perché). Inoltre è facile calcolare  $\mathbf{E}[Y]$  (pur senza conoscere la forma della densità). Infatti si ha

$$\begin{aligned}\mathbf{E}[Y] &= \mathbf{E}[X_1^2 + X_2^2 + \dots + X_n^2] = \mathbf{E}[X_1^2] + \mathbf{E}[X_2^2] + \dots + \mathbf{E}[X_n^2] \\ &= \mathbf{Var}X_1 + \mathbf{Var}X_2 + \dots + \mathbf{Var}X_n = n.\end{aligned}$$

Il grafico della densità  $\chi^2(n)$  è riportato a p. 143.

**Definizione 4.3.12.** Siano  $X \sim \mathcal{N}(0, 1)$  e  $Y \sim \chi^2(n)$  due variabili aleatorie indipendenti. Poniamo

$$Z = \sqrt{n} \frac{X}{\sqrt{Y}}.$$

La densità di  $Z$  si chiama *densità  $t$  di Student a  $n$  gradi di libertà*; la indicheremo con il simbolo  $t(n)$ .

Anche della densità  $t(n)$  è possibile dare un'espressione esplicita, che a noi non interessa. Per i nostri scopi ci basterà sapere che si tratta di una densità pari (e dunque la relativa variabile aleatoria sarà simmetrica, ved. il paragrafo 4.4) e che il suo grafico è un grafico “a campana” che somiglia a quello della gaussiana standard (ved. p. 143); anzi, per  $n \rightarrow \infty$  il grafico della densità  $t(n)$  tende a diventare quello della gaussiana.

Le tre densità  $\mathcal{N}(0, 1)$ ,  $\chi^2(n)$  e  $t(n)$  sono collegate da un importante teorema, che è fondamentale in statistica:

**Teorema 4.3.13.** *Siano  $X_1, X_2, \dots, X_n$   $n$  variabili aleatorie indipendenti, tutte con legge  $\mathcal{N}(\mu, \sigma^2)$ . Indicheremo al solito con  $\bar{X}$  la loro media campionaria e porremo inoltre*

$$S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}.$$

*Consideriamo le due variabili aleatorie*

$$Z = \sqrt{n} \frac{\bar{X} - \mu}{\sigma}$$

e

$$W = \frac{S^2}{\sigma^2} (n-1) = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{\sigma^2}.$$

*Allora  $Z$  ha densità  $\mathcal{N}(0, 1)$ ,  $W$  ha densità  $\chi^2(n-1)$  ed inoltre  $Z$  e  $W$  sono tra loro indipendenti.*

Il risultato precedente è un caso particolare di un importante teorema, dovuto allo statistico W.G. Cochran; sebbene un po' impropriamente, chiameremo 4.3.13 ancora “Teorema di Cochran”. Da esso e dalla definizione di densità di Student segue:

**Corollario 4.3.14.** *Nelle ipotesi e con le notazioni del teorema precedente la variabile aleatoria*

$$T = \sqrt{n-1} \frac{Z}{\sqrt{W}}$$

*ha densità  $t(n-1)$ .*

**Osservazione 4.3.15.** La variabile  $T$  sopra definita si può scrivere anche in una forma che risulterà utile in statistica. Si ha

$$T = \sqrt{n-1} \frac{Z}{\sqrt{W}} = \sqrt{n-1} \sqrt{n} \frac{\bar{X} - \mu}{\sigma} \frac{\sigma}{\sqrt{n-1} S} = \sqrt{n} \frac{\bar{X} - \mu}{S},$$

dove con  $S$  si è indicata la radice quadrata aritmetica di  $S^2$ .

Osserviamo ora che:

- (i) la variabile aleatoria  $T$  non dipende dal parametro  $\sigma$ ;
- (ii) l'espressione adesso ricavata per  $T$ , e cioè

$$T = \sqrt{n} \frac{\bar{X} - \mu}{S},$$

a parte la sostituzione del parametro  $\sigma$  con  $S$ , è identica a quella di  $Z$ .

**Osservazione 4.3.16.** Il teorema di Cochran non è facile da memorizzare. Per quanto riguarda la v.a.  $W$ , la sua forma somiglia a quella della v.a.

$$U = \frac{\sum_{i=1}^n (X_i - \mu)^2}{\sigma^2} = \sum_{i=1}^n \left( \frac{X_i - \mu}{\sigma} \right)^2;$$

osservando che le v.a.  $\frac{X_i - \mu}{\sigma}$  sono indipendenti e tutte di legge  $\mathcal{N}(0, 1)$  (Corollario 4.3.3 (ii)), si vede subito che  $U$  ha legge  $\chi^2(n)$ , per la Definizione 4.3.11. La differenza tra  $U$  e  $W$  consiste nella sostituzione di  $\mu$  (media delle  $X_i$ ) con  $\bar{X}$  (media campionaria delle  $X_i$ ), che ne è uno “stimatore” (il significato di questo termine sarà chiarito nella sezione di statistica, Cap. 5). Questo produce la perdita di un grado di libertà nella legge  $\chi^2$  risultante (per  $W$  si hanno infatti  $n - 1$  gradi anziché  $n$ , come per  $U$ ), per la “regola” secondo la quale “ogni parametro stimato riduce di 1 il numero di gradi di libertà” (Osservazione 6.2.10). Per la v. a.  $T$ , abbiamo già osservato in (4.3.15) (ii) che la sua forma somiglia a quella di  $Z$ , con la sostituzione di  $\sigma$  con il suo “stimatore”  $S$ ; tale sostituzione fa cambiare la legge: da una  $\mathcal{N}(0, 1)$  si passa una  $t(n)$ , che comunque, almeno per  $n$  grande, è molto vicina alla  $\mathcal{N}(0, 1)$ .

**Osservazione 4.3.17.** Sappiamo (Proposizione 2.6.11) che due v.a. che siano funzioni di gruppi “separati” di v.a. indipendenti sono indipendenti. Se i due gruppi non sono separati, in generale non si può concludere niente a proposito dell'indipendenza delle due variabili risultanti: è facile costruire esempi in cui esse non sono indipendenti, ma il teorema di Cochran fornisce un caso in cui invece le cose vanno diversamente: le v.a.  $Z$  e  $W$  sono entrambe funzioni di *tutte* le  $X_i$ , e tuttavia sono tra loro indipendenti.

## 4.4 La funzione di ripartizione della legge normale standard

In questo paragrafo vedremo alcune proprietà della funzione  $\Phi$  e parleremo dell'uso delle relative tavole (riportate a p. 144).

**Definizione 4.4.1.** Sia  $X$  una variabile aleatoria. Si dice che  $X$  è *simmetrica* se  $X$  e  $-X$  hanno la stessa legge, o, in modo equivalente, la stessa funzione

di ripartizione:

$$F_X(t) = P(X \leq t) = P(-X \leq t) = F_{-X}(t) \quad \forall t \in \mathbb{R}.$$

**Osservazione 4.4.2.** (i) Sia  $X$  una v. a. continua. La precedente relazione si può scrivere anche nella forma

$$\begin{aligned} F_X(t) &= P(X \leq t) = P(-X \leq t) = P(X \geq -t) = 1 - P(X < -t) \\ &= 1 - P(X \leq -t) = 1 - F_X(-t), \quad \forall t \in \mathbb{R} \end{aligned},$$

ovvero anche

$$F_X(t) + F_X(-t) = 1. \quad (4.4.3)$$

Dalla relazione (4.4.3) si deduce in particolare che, se  $X$  è simmetrica, allora  $2F_X(0) = 1$ , e cioè

$$F_X(0) = \frac{1}{2}. \quad (4.4.4)$$

(ii) Se  $X$  ammette densità  $f$ , allora  $X$  è simmetrica se e solo se,  $\forall t \in \mathbb{R}$ , si ha

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

Vale la seguente:

**Proposizione 4.4.5.** *Sia  $X$  una variabile aleatoria assolutamente continua. Allora  $X$  è simmetrica se e solo se ammette una densità  $f$  pari, cioè tale che*

$$f(x) = f(-x) \quad \forall x \in \mathbb{R}$$

(dimostrazione per esercizio).

Segue dalla Proposizione 4.4.5 che ogni variabile aleatoria  $X$  avente legge normale standard è simmetrica (poiché  $x \mapsto \exp(-x^2/2)$  è una funzione pari). Le relazioni (4.4.3) e (4.4.4) applicate alla funzione di ripartizione  $\Phi$  danno allora rispettivamente

$$\Phi(t) + \Phi(-t) = 1; \quad (4.4.6)$$

$$\Phi(0) = \frac{1}{2}.$$

Vediamo ora in che modo si usano le tavole della legge normale standard. Esistono vari tipi di tavole. Le più comuni riportano il valore di  $\Phi(x)$  per  $x \in (0, 3)$  (le tavole di p. 144 addirittura per  $x \in (0, 4.4)$ ). Il valore  $x$ , fino al primo decimale, si legge nella colonna di sinistra; se  $x$  ha anche un secondo decimale, esso va cercato nella riga in alto. Allora  $\Phi(x)$  si trova, nella tavola, nella posizione in cui si incrociano la riga e la colonna prescelte. Ad esempio, se  $x = 1.64$ , il valore 1.6 si legge nella colonna di sinistra. Ad esso si aggiunge



il numero 0.04, che troviamo nella riga in alto. Dunque  $\Phi(1.64) = 0.9495$  (è il numero che si trova all'intersezione della riga marcata 1.6 con la colonna marcata 0.04).

Questo per quanto riguarda i numeri  $x \in (0.3)$ . Se invece  $x \in (-3, 0)$  si usa la relazione (4.4.6). Ad esempio, per  $x = -1.64$  si trova

$$\Phi(-1.64) = 1 - \Phi(1.64) = 1 - 0.9495 = 0.0505.$$

Infine, se  $x \geq 3$ , il valore di  $\Phi(x)$  si approssima con 1 (e, di conseguenza,  $\Phi(x) \approx 0$  per  $x \leq -3$ ).

Un altro tipo di tavole è quello che riporta i valori di

$$x \mapsto \Delta(x) := P(0 \leq X \leq x) = \int_0^x \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx,$$

sempre per  $x \in (0.3)$ . Il legame di questa funzione con la funzione di ripartizione è il seguente

$$\begin{aligned}\Phi(x) &= P(X \leq x) = P(X < 0) + P(0 \leq X \leq x) \\ &= \Phi(0) + P(0 \leq X \leq x) = 0.5 + \Delta(x).\end{aligned}$$

Infine altre tavole meno usate danno il valore di  $P(X > x) = 1 - \Phi(x)$ .

## 4.5 La Legge dei Grandi Numeri

Sullo spazio di probabilità  $(\Omega, \mathcal{A}, P)$  siano  $(X_n)_{n \in \mathbb{N}}$  una successione di variabili aleatorie,  $X$  una variabile aleatoria.

**Definizione 4.5.1.** Si dice che, per  $n \rightarrow \infty$ ,  $(X_n)$  converge *in probabilità* verso  $X$ , e si scrive

$$X_n \xrightarrow{P} X,$$

se, per ogni  $\epsilon > 0$ , risulta

$$\lim_{n \rightarrow \infty} P(|X_n - X| > \epsilon) = 0.$$

Vale la seguente:

**Proposizione 4.5.2 (Criterio per la convergenza in probabilità verso una costante).** *Su  $(\Omega, \mathcal{A}, P)$  sia  $(Z_n)_{n \geq 1}$  una successione di v. a. aventi ciascuna media e varianza finite. Se*

$$\lim_{n \rightarrow \infty} \mathbf{E}[Z_n] = c < +\infty, \quad \lim_{n \rightarrow \infty} \mathbf{Var} Z_n = 0,$$

*allora la successione  $(Z_n)_{n \geq 1}$  converge in probabilità verso  $Z \equiv c$ .*

*Dimostrazione.* Sia  $\epsilon > 0$  un numero fissato. Dobbiamo vedere che tende a 0 la successione delle probabilità degli eventi

$$A_n := \{|Z_n - c| > \epsilon\}.$$

Per la disuguaglianza di Markov 3.1.24 si ha

$$P(A_n) = P((Z_n - c)^2 > \epsilon^2) \leq \frac{\mathbf{E}[(Z_n - c)^2]}{\epsilon^2}$$

e d'altra parte

$$\begin{aligned} \mathbf{E}[(Z_n - c)^2] &= \mathbf{E}[\{(Z_n - \mathbf{E}[Z_n]) + (\mathbf{E}[Z_n] - c)\}^2] \\ &= \mathbf{E}[(Z_n - \mathbf{E}[Z_n])^2] + \underbrace{2\mathbf{E}[(Z_n - \mathbf{E}[Z_n])(\mathbf{E}[Z_n] - c)]}_{=0} + \mathbf{E}[(\mathbf{E}[Z_n] - c)^2] \\ &= \mathbf{Var}Z_n + (\mathbf{E}[Z_n] - c)^2 \rightarrow 0, \quad n \rightarrow \infty \end{aligned}$$

per le ipotesi fatte (spiegare perché l'addendo centrale è nullo!). □

La proposizione precedente ha una conseguenza importante. Il termine “debole” usato qui sotto verrà spiegato più avanti in (4.5.13).

**Corollario 4.5.3 (Legge Debole dei Grandi Numeri).** *Sia  $(X_n)_{n \geq 1}$  una successione di variabili aleatorie i.i.d., con media  $\mu$  e varianza  $\sigma^2$  finite. Allora la successione delle medie campionarie*

$$\frac{X_1 + X_2 + \cdots + X_n}{n}$$

*converge in probabilità verso  $\mu$  per  $n \rightarrow \infty$ :*

$$\frac{X_1 + X_2 + \cdots + X_n}{n} \xrightarrow{P} \mu, \quad n \rightarrow \infty.$$

*Dimostrazione.* Basta applicare il criterio della proposizione precedente alle v. a.

$$Z_n = \frac{X_1 + X_2 + \cdots + X_n}{n}.$$

Infatti, in questo caso, per la proprietà di linearità della speranza,

$$\mathbf{E}[Z_n] = \mathbf{E}\left[\frac{\sum_{k=1}^n X_k}{n}\right] = \frac{1}{n} \sum_{k=1}^n \mathbf{E}[X_k] = \frac{n\mu}{n} = \mu;$$

inoltre, per l'indipendenza delle  $(X_n)$  e per le proprietà della varianza

$$\mathbf{Var} Z_n = \mathbf{Var} \left( \frac{\sum_{k=1}^n X_k}{n} \right) = \frac{1}{n^2} \sum_{k=1}^n \mathbf{Var} X_k = \frac{n \sigma^2}{n^2} = \frac{\sigma^2}{n} \rightarrow 0,$$

per  $n \rightarrow \infty$ . □

**Esempio 4.5.4.** Sia  $(X_n)_{n \geq 1}$  una successione di variabili aleatorie indipendenti, tutte con legge  $B(1, p)$ . L'  $n$ -esima variabile può essere pensata come la variabile indicatrice dell'  $n$ -esima prova in una successione infinita di prove indipendenti del tipo successo-insuccesso. La media campionaria

$$\frac{X_1 + X_2 + \cdots + X_n}{n}$$

rappresenta allora la percentuale di successi nelle prime  $n$  prove. Il fatto che all'aumentare di  $n$  tale percentuale si avvicini a  $p$  (probabilità di successo nella generica prova) è accettato da tutti, ma in realtà è conseguenza della Legge dei Grandi Numeri.

**Osservazione 4.5.5.** Nella dimostrazione della Legge dei Grandi Numeri, l'ipotesi di equidistribuzione delle  $(X_n)$  serve solo ad assicurare che queste variabili abbiano tutte la stessa media e la stessa varianza. Il risultato vale in effetti anche in queste ipotesi più deboli (purché ci sia ancora l'ipotesi di indipendenza).

Applicando direttamente la disuguaglianza di Tchebyceff nel caso delle medie campionarie considerato nella Legge dei Grandi Numeri, si ottiene la disuguaglianza

$$P \left( \left| \frac{X_1 + X_2 + \cdots + X_n}{n} - \mu \right| > \epsilon \right) \leq \frac{\sigma^2}{n\epsilon^2}, \quad (4.5.6)$$

(dettagli per esercizio) che costituisce una stima dall'alto della probabilità che lo scarto tra la media campionaria e il vero valore della media superi il valore di soglia fissato  $\epsilon$ . Nell'esempio che segue diamo una valutazione della stessa probabilità ottenuta mediante la formula dell'approssimazione normale.

**Esempio 4.5.7.** Sia  $\epsilon > 0$  un numero fissato. Nelle ipotesi della Legge dei Grandi Numeri, calcoliamo un valore approssimato per

$$P \left( \left| \frac{X_1 + X_2 + \cdots + X_n}{n} - \mu \right| > \epsilon \right).$$

Utilizzando l'approssimazione normale vista nel Paragrafo 4.3, si ha

$$\begin{aligned}
 & P\left(\left|\frac{X_1 + X_2 + \cdots + X_n}{n} - \mu\right| > \epsilon\right) \\
 &= P\left(\frac{X_1 + X_2 + \cdots + X_n}{n} - \mu > \epsilon\right) \\
 &\quad + P\left(\frac{X_1 + X_2 + \cdots + X_n}{n} - \mu < -\epsilon\right) \\
 &= P(X_1 + X_2 + \cdots + X_n > n\mu + n\epsilon) \\
 &\quad + P(X_1 + X_2 + \cdots + X_n < n\mu - n\epsilon) \\
 &= 1 - P(X_1 + X_2 + \cdots + X_n \leq n\mu + n\epsilon) \\
 &\quad + P(X_1 + X_2 + \cdots + X_n < n\mu - n\epsilon) \\
 &\approx 1 - \Phi\left(\frac{n\mu + n\epsilon - n\mu}{\sigma\sqrt{n}}\right) + \Phi\left(\frac{n\mu - n\epsilon - n\mu}{\sigma\sqrt{n}}\right) \\
 &= 1 - \Phi\left(\frac{\sqrt{n}\epsilon}{\sigma}\right) + \Phi\left(-\frac{\sqrt{n}\epsilon}{\sigma}\right) = 2\left(1 - \Phi\left(\frac{\sqrt{n}\epsilon}{\sigma}\right)\right).
 \end{aligned}$$

**Osservazione 4.5.8.** Le approssimazioni che si ottengono con l'uso della diseguaglianza di Tchebyceff sono in generale poco accurate (si possono addirittura trovare esempi in cui il confine superiore della diseguaglianza è maggiore di 1!). Confrontiamo ad esempio il risultato (4.5.6) con quello dell'Esempio 4.5.7 nel caso di una successione  $(X_n)$  di v. a. indipendenti aventi legge  $\mathcal{B}\left(1, \frac{1}{2}\right)$  con  $n = 100$ ,  $\epsilon = \frac{1}{10}$ . Ricordando che  $\sigma^2 = p(1-p) (= 0.25$  in questo caso) la (4.5.6) dà il valore 0.25, mentre da (4.5.7) si ottiene  $2(1 - \Phi(2)) = 2(1 - 0.9772) = 0.0456$ , molto migliore del precedente, se si tiene conto che il valore effettivo della probabilità cercata è 0.038.

Per uso futuro accenniamo molto brevemente anche ad un altro tipo di convergenza di variabili aleatorie.

Sullo spazio di probabilità  $(\Omega, \mathcal{A}, P)$  sia  $(X_n)_{n \in \mathbb{N}}$  una successione di variabili aleatorie e sia  $X$  una variabile aleatoria.

**Definizione 4.5.9.** Si dice che, per  $n \rightarrow \infty$ ,  $(X_n)$  converge *quasi certamente* verso  $X$  e si scrive

$$X_n \xrightarrow{q.c.} X,$$

se esiste un evento  $E \in \mathcal{A}$  avente probabilità 1 tale che, per ogni  $\omega \in E$  si abbia

$$\lim_{n \rightarrow \infty} X_n(\omega) = X(\omega).$$

**Osservazione 4.5.10.** Sarebbe più semplice dire che  $(X_n)$  converge quasi certamente verso  $X$  se

$$P(\omega \in \Omega : \lim_{n \rightarrow \infty} X_n(\omega) = X(\omega)) = 1.$$

Cercare di spiegare il motivo per il quale la definizione viene data nel modo leggermente più complicato (4.5.9).

Sussiste una gerarchia tra i concetti di convergenza quasi certa e convergenza in probabilità visto sopra; precisamente:

**Teorema 4.5.11.** *Se  $X_n \xrightarrow{q.c.} X$ , allora  $X_n \xrightarrow{P} X$ .*

Vale inoltre il risultato seguente:

**Teorema 4.5.12 (Legge Forte dei Grandi Numeri).** *Sia  $(X_n)_{n \geq 1}$  una successione di variabili aleatorie i.i.d., con media finita  $\mu$ . Allora la successione delle medie campionarie*

$$\frac{X_1 + X_2 + \cdots + X_n}{n}$$

*converge quasi certamente verso  $\mu$  per  $n \rightarrow \infty$ :*

$$\frac{X_1 + X_2 + \cdots + X_n}{n} \xrightarrow{q.c.} \mu, \quad n \rightarrow \infty.$$

**4.5.13.** Dato che, come abbiamo visto in 4.5.11, la convergenza quasi certa implica la convergenza in probabilità, il teorema precedente implica il Corollario 4.5.3. Questo spiega gli attributi di “Forte” e “Debole” utilizzati rispettivamente per i due risultati. Ci siamo limitati alla dimostrazione di 4.5.3 (addirittura con l’ulteriore ipotesi di finitezza della varianza) perché è la sola alla nostra portata.

## 4.6 Ricerca di leggi (e densità) di v. a. assolutamente continue

Conoscendo la densità  $g$  di una v. a. assolutamente continua  $Y$  è possibile dedurre (almeno in teoria) la sua funzione di ripartizione, e dunque la sua legge, per mezzo della formula

$$F_Y(t) = P(Y \leq t) = \int_{-\infty}^t g(x) dx. \quad (4.6.1)$$

Spesso però capita di dover effettuare il procedimento inverso, cioè abbiamo a disposizione la f.d.r. di  $Y$ , e dobbiamo cercare una sua densità (ammesso

che esista). In questo paragrafo ci occuperemo appunto di alcuni metodi per trovare la densità di una v. a. partendo dalla conoscenza della sua funzione di ripartizione.

La situazione più semplice si ha quando la funzione di ripartizione di  $Y$  è già scritta, per qualche motivo fortunato, nella “forma giusta”, cioè come nella formula (4.6.1): la funzione  $g$  che compare nell'integrale è una densità per  $Y$  (rivedere la definizione di v. a. assolutamente continua). Questo è capitato per esempio nella dimostrazione della Proposizione 4.3.2 (in cui si aveva  $Y = aX + b$ , cioè  $Y$  era una particolare funzione di un'altra v. a.  $X$ ): nell'integrale dell'uguaglianza

$$P(Y \leq y) = \int_{-\infty}^y \frac{1}{\sqrt{2\pi}|a|s} \exp\left(-\frac{(x - am - b)^2}{2a^2s^2}\right) dx$$

abbiamo riconosciuto la densità  $g$  della legge  $\mathcal{N}(am + b, a^2s^2)$ , e abbiamo concluso che  $Y$  segue appunto tale legge.

Il procedimento utilizzato nella Proposizione 4.3.2 era un opportuno cambiamento di variabile, che trasformava l'ultimo integrale della formula ( $a > 0$  per brevità)

$$\begin{aligned} P(Y \leq y) &= P(aX + b \leq y) = P\left(X \leq \frac{y - b}{a}\right) \\ &= \int_{-\infty}^{\frac{y-b}{a}} \frac{1}{\sqrt{2\pi}s} \exp\left(-\frac{(t - m)^2}{2s^2}\right) dt \end{aligned}$$

nell'integrale desiderato.

Esattamente la stessa tecnica può essere usata in generale, nel caso in cui  $Y$  sia funzione ( $Y = \phi(X)$ ) di un'altra variabile assolutamente continua  $X$ , di cui sia nota la densità. Più precisamente si ha il seguente risultato (consiglio: leggerne la dimostrazione parallelamente alla dimostrazione della 4.3.2; si riconoscerà facilmente che il metodo di prova è esattamente il medesimo):

**Proposizione 4.6.2.** *Sia  $\phi : (a, b) \rightarrow (c, d)$  una funzione differenziabile, surgettiva e strettamente monotona (dove gli intervalli considerati possono essere non limitati da una o ambedue le parti). Sia  $X$  una v. a. assolutamente continua di densità  $f$ , tale che  $f$  sia nulla fuori di  $(a, b)$  (in altri termini,  $X$  prende quasi certamente valori in  $(a, b)$ ). Allora la v. a.  $Y = \phi(X)$  è assolutamente continua, e la sua densità è data dalla formula*

$$g(y) = \begin{cases} f(\phi^{-1}(y)) \left| \frac{d}{dy} \phi^{-1}(y) \right| & \text{per } y \in (c, d) \\ 0 & \text{altrove.} \end{cases}$$

*Dimostrazione.* Cominciamo con l'osservare che i valori di  $Y$  appartengono (quasi certamente) all'intervallo  $(c, d)$  (infatti, dato che  $X$  prende valori in

$(a, b)$ , si avrà  $P(Y \in (c, d)) = P(X \in (a, b)) = 1$  e quindi, per prima cosa,

$$P(Y \leq y) = \begin{cases} 0 & \text{per } y \leq c \\ 1 & \text{per } y \geq d \end{cases}$$

(dettagli per esercizio; si intende che, se  $c = -\infty$  (risp.  $d = +\infty$ ), la prima (risp. seconda) parte della formula precedente non deve essere scritta).

Sia ora  $y \in (c, d)$ ; supponiamo dapprima che  $\phi$  sia strettamente crescente e scriviamo la f.d.r. della  $Y$  in termini della  $X$  (che è la v. a. sulla quale abbiamo le informazioni). Si ha

$$P(Y \leq y) = P(c \leq Y \leq y) = P(\phi(a) \leq \phi(X) \leq y) = P(a \leq X \leq \phi^{-1}(y))$$

(applicando la funzione  $\phi^{-1}$ , anch'essa crescente), e, continuando,

$$P(a \leq X \leq \phi^{-1}(y)) = \int_a^{\phi^{-1}(y)} f(x) dx.$$

Se in quest'ultimo integrale effettuiamo il cambiamento di variabili  $z = \phi(x)$ , cioè  $x = \phi^{-1}(z)$ ,  $dx = \frac{d}{dz}\phi^{-1}(z)$ , si vede che esso è uguale a

$$\int_{\phi(a)=c}^y f(\phi^{-1}(z)) \frac{d}{dz}\phi^{-1}(z) dz = \int_{-\infty}^y f(\phi^{-1}(z)) \left| \frac{d}{dz}\phi^{-1}(z) \right| dz,$$

osservando che:

- (i) se  $z \leq c$ , allora  $\phi^{-1}(z) \leq \phi^{-1}(c) = a$ , perciò  $f(\phi^{-1}(z)) = 0$ ;
- (ii)  $\phi^{-1}$  è crescente e quindi la sua derivata è positiva, dunque uguale al suo valore assoluto (il motivo per il quale la formula viene scritta con il valore assoluto sarà chiaro tra poco).

Se  $\phi$  è decrescente, si scrive

$$\begin{aligned} P(Y \leq y) &= P(c \leq \phi(X) \leq y) = P(\phi(b) \leq \phi(X) \leq y) \\ &= P(\phi^{-1}(y) \leq X \leq b) = \int_{\phi^{-1}(y)}^b f(x) dx. \end{aligned}$$

A questo punto si procede con lo stesso cambiamento di variabili di prima e si ottiene l'uguaglianza

$$\begin{aligned} P(Y \leq y) &= \int_y^{\phi(b)=c} f(\phi^{-1}(z)) \frac{d}{dz}\phi^{-1}(z) dz \\ &= - \int_c^y f(\phi^{-1}(z)) \frac{d}{dz}\phi^{-1}(z) dz = - \int_{-\infty}^y f(\phi^{-1}(z)) \frac{d}{dz}\phi^{-1}(z) dz \\ &= \int_{-\infty}^y f(\phi^{-1}(z)) \left| \frac{d}{dz}\phi^{-1}(z) \right| dz \end{aligned}$$

perché in questo caso  $\phi^{-1}$  è decrescente e quindi ha derivata negativa.  $\square$

**Osservazione 4.6.3.** Fare attenzione al fatto che la proposizione precedente richiede che la trasformazione  $\phi$  sia *biunivoca* da  $(a, b)$  a  $(c, d)$ . Se questa ipotesi non è valida, la proposizione non può essere applicata. L'esempio seguente mostra come fare in casi di questo genere.

**Esempio 4.6.4.** Sia  $X$  una v. a. avente legge  $\mathcal{U}(-1, 1)$ . Si vuole calcolare la legge di  $Y = X^2$ . La trasformazione  $\phi : x \mapsto x^2$  applica l'intervallo  $(-1, 1)$  sull'intervallo  $(0, 1)$  ma non è biunivoca. Partendo come al solito dal calcolo della funzione di ripartizione di  $Y$ , per  $t \in (0, 1)$  si ha facilmente

$$P(Y \leq t) = P(X^2 \leq t) = P(-\sqrt{t} \leq X \leq \sqrt{t}) = \int_{-\sqrt{t}}^{\sqrt{t}} f(x) dx;$$

a questo punto, volendo ancora applicare (come nella Proposizione 4.6.2) il cambiamento di variabili  $z = \phi(x)$ , dobbiamo scrivere

$$\int_{-\sqrt{t}}^{\sqrt{t}} f(x) dx = \int_{-\sqrt{t}}^0 f(x) dx + \int_0^{\sqrt{t}} f(x) dx,$$

in modo che nei due intervalli  $(-\sqrt{t}, 0)$  e  $(0, \sqrt{t})$  la trasformazione sia biunivoca (ipotesi richiesta dal teorema di cambiamento di variabili per gli integrali). Il seguito dei calcoli viene lasciato per esercizio.

I calcoli precedenti (sia che si possa utilizzare la Proposizione 4.6.2, sia che si debba procedere come nell'Esempio 4.6.4) sono spesso fastidiosi; un errore comune è appunto quello di applicare la Proposizione anche quando non si potrebbe. Nel seguito indichiamo un metodo, meno rigoroso ma più semplice per il calcolo della legge di  $Y = \phi(X)$ . Calcolando come sempre la f.d.r. di  $Y$ , e ponendo  $A_t = \{x \in \mathbb{R} : \phi(x) \leq t\}$ , si ha

$$P(Y \leq t) = P(X \in A_t) = \int_{A_t} f(x) dx.$$

Questo integrale è una funzione di  $t$ , diciamola  $G$ , e quindi si arriva all'uguaglianza  $P(Y \leq t) = G(t)$ : se il calcolo effettivo dell'integrale è possibile, la funzione  $G$  risultante è intanto la f.d.r. di  $Y$ .

Volendo calcolare la densità  $g$  (ammesso che esista) ricordiamo prima di tutto che essa deve soddisfare la relazione

$$G(t) = \int_{-\infty}^t g(y) dy.$$

Se sapessimo che  $g$  è continua, il teorema fondamentale del Calcolo (riguardarlo!) ci direbbe che  $G$  è derivabile e

$$G'(t) = g(t),$$



e inoltre questa formula esprimerebbe  $g$  in termini di  $G$ , che è nota. Il problema è che non abbiamo informazioni sulla continuità di  $g$ , quindi il ragionamento precedente non è corretto; e del resto molte densità non sono continue su tutto  $\mathbb{R}$  (pensare per es. alle densità uniformi). Tuttavia procederemo così:  $G$  in generale è derivabile salvo al più in un insieme finito di punti  $\mathcal{F}$ . Deriviamo allora  $G$  dove possibile, ottenendo una funzione  $g$  definita dappertutto salvo che su  $\mathcal{F}$ ; nei punti di  $\mathcal{F}$  definiamo  $g$  in modo arbitrario (per esempio dando valore 0). Otteniamo così una funzione, diciamola ancora  $g$ , definita su tutto  $\mathbb{R}$ , che è la nostra candidata per essere la densità di  $Y$ . Per verificarlo in modo rigoroso, dovremmo calcolare gli integrali del tipo  $\int_{-\infty}^t g(y) dy$  e controllare che *per ogni*  $t$  vale l'uguaglianza

$$\int_{-\infty}^t g(y) dy = G(t) = P(Y \leq t),$$

con il che la verifica sarebbe completa. Questa seconda parte (calcolo degli integrali) in genere si omette per brevità, ma è bene comunque sapere che esistono casi nei quali derivare  $G$  dove possibile non è sufficiente per garantire l'esistenza della densità, perché la funzione  $g$  risultante è nulla su tutto  $\mathbb{R}$  (e quindi non può essere una densità, perché come sappiamo ogni densità ha integrale 1 sulla retta). Tuttavia questi esempi sono (in un certo senso) patologici, e negli esercizi non si incontrano mai: la sola derivazione sarà sufficiente per ottenere la densità.

**Esercizio 4.6.5.** Calcolare con il metodo descritto sopra la densità della v. a.  $Y$  dell'Esempio 4.6.4.

**Esempio 4.6.6.** Sia  $X$  una v.a. avente legge  $\mathcal{N}(0, 1)$ . Vogliamo calcolare la densità di  $Y = X^2$  (si tratta della densità  $\chi^2(1)$ ). Essendo  $Y$  una v. a. non negativa, si ha intanto  $P(Y \leq t) = 0$  per ogni  $t < 0$ . Per ogni  $t \geq 0$  invece

$$\begin{aligned} P(Y \leq t) &= P(-\sqrt{t} \leq X \leq \sqrt{t}) = \int_{-\sqrt{t}}^{\sqrt{t}} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx \\ &= \Phi(\sqrt{t}) - \Phi(-\sqrt{t}) = 2\Phi(\sqrt{t}) - 1. \end{aligned}$$

Dunque la f.d.r. di  $Y$  è

$$G(t) = \begin{cases} 2\Phi(\sqrt{t}) - 1 & \text{per } t \geq 0 \\ 0 & \text{per } t < 0 \end{cases}$$

(disegnarla approssimativamente).

Se adesso vogliamo la densità di  $Y$ , dobbiamo derivare  $G$ . Per  $t < 0$  si ha  $G'(t) = 0$ . Ricordando che

$$\Phi'(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}$$

(perché?) e utilizzando la regola di derivazione delle funzioni composte, si ha, per  $t > 0$ ,

$$\frac{d}{dt}\Phi(\sqrt{t}) = \Phi'(\sqrt{t}) \cdot \frac{d}{dt}(\sqrt{t}) = \frac{1}{2\sqrt{2\pi t}} e^{-\frac{t}{2}},$$

e quindi, per  $t > 0$  si ha  $G'(t) = \frac{1}{\sqrt{2\pi t}} e^{-\frac{t}{2}}$ . La funzione trovata è definita per ogni  $t \neq 0$ . Possiamo estenderla ad una funzione definita su tutto  $\mathbb{R}$  dandole valore 0 per  $t = 0$  e quindi la (una) densità di  $Y$  è

$$g(t) = \begin{cases} 0 & \text{per } t \leq 0 \\ \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{t}} e^{-\frac{t}{2}} & \text{per } t > 0. \end{cases}$$

Ricordiamo che la densità  $\Gamma\left(\frac{1}{2}, \frac{1}{2}\right)$  è

$$h(t) = \begin{cases} 0 & \text{per } t \leq 0 \\ \frac{1}{\sqrt{2}\Gamma\left(\frac{1}{2}\right)} \frac{1}{\sqrt{t}} e^{-\frac{t}{2}} & \text{per } t > 0. \end{cases}$$

Confrontando  $g$  con  $h$ , si vede che esse coincidono per quanto riguarda la parte dipendente da  $t$ . D'altra parte sappiamo che sia  $g$  che  $h$  sono densità ( $g$  per i calcoli appena fatti,  $h$  per la teoria delle leggi Gamma), e dunque anche le costanti moltiplicative  $(\sqrt{2\pi})^{-1}$  e  $(\sqrt{2}\Gamma(1/2))^{-1}$  devono essere uguali (perché? osservare che gli integrali di  $g$  e  $h$  su  $\mathbb{R}$  valgono entrambi 1). Si ricava come sottoprodotto la relazione importante (v. Proprietà 4.2.12 (iii))

$$\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}.$$

## La statistica inferenziale

### 5.1 Introduzione ai problemi statistici

Nello studio scientifico di un fenomeno, di qualsiasi natura esso sia, normalmente si hanno a disposizione le risorse necessarie per esaminarne solo una piccolissima parte. Di qui la necessità di effettuare quello che si chiama un *rilevamento statistico*, cioè un'indagine (quantitativa o qualitativa) su un frammento del fenomeno stesso. Ogni rilevamento statistico produce *un campione di dati* (relativamente piccolo, appunto). La statistica descrittiva si occupa di organizzare e riassumere in modo significativo questi dati, e qui termina il suo compito. La statistica inferenziale, invece, utilizzando metodi e nozioni del calcolo delle probabilità, cerca di fare previsioni sul futuro, o di ottenere risultati estendibili all'intera popolazione (a partire solo dal piccolo campione effettivamente osservato).

Nella pratica, tipicamente si deve effettuare un esperimento che produce una variabile aleatoria  $X$  di cui non si conosce la legge, e si vogliono ricavare informazioni su di essa. Vediamo qualche esempio.

**Esempio 5.1.1.** Si deve decidere se una data moneta è truccata oppure no. In altre parole, se  $X$  è la v. a.

$$X = \begin{cases} 1 & \text{se esce } T \\ 0 & \text{se esce } C, \end{cases}$$

allora  $X \sim B(1, p)$ , ma  $p$  non è nota. Vedremo in seguito come si procede per ottenere informazioni su  $p$ ; intanto ci si può fare un'idea rileggendo l'Esempio 4.5.4.

**Esempio 5.1.2.** Si vuole sapere se in una certa coltura batterica i batteri si distribuiscono in modo uniforme oppure tendono a formare aggregazioni. Supponiamo che il numero di batteri nella coltura sia  $N$ .

Suddividiamo la coltura in  $n$  parti (per es.  $n \text{ mm}^3$ ), e sia  $X$  il numero di batteri presenti in una parte fissata. La coltura sarà omogenea se ogni batterio

sceglie a caso e indipendentemente la parte in cui impiantarsi; allora le v.a.

$$Y_k = \begin{cases} 1 & \text{se il } k\text{-esimo batterio sceglie la parte fissata} \\ 0 & \text{se no,} \end{cases} \quad (k = 1, \dots, N)$$

hanno densità  $B(1, 1/n)$  e sono tra loro indipendenti. Pertanto si ha

$$X = Y_1 + \dots + Y_N \sim B(N, 1/n).$$

Se  $N$  e  $n$  sono grandi, si può approssimare la legge di  $X$  con una Poisson di parametro  $\lambda = N/n$ . Concludiamo che la distribuzione dei batteri sarà omogenea se scopriremo che  $X$  ha densità di Poisson. In altri termini la domanda iniziale è diventata:  $X$  ha legge di Poisson oppure no?

Si osservi che in questo caso non è neppure noto il tipo di legge seguita da  $X$  (e non solo un parametro di essa).

## 5.2 Il concetto di stimatore

*Generalità.* Uno dei problemi che si presentano più frequentemente allo statista che voglia ottenere informazioni su una data v.a.  $X$  è quello di doverne decidere la legge: supponiamo che egli sappia che tale legge appartiene ad una data famiglia, dipendente da un parametro  $\theta$  non noto: ad esempio, in 5.1.1 si sa che la legge considerata è di tipo bernoulliano, ma non se ne conosce il parametro  $p$ .

*Convenzione.* Nel caso generale il parametro da studiare è indicato con  $\theta$ . In casi specifici il simbolo usato potrà essere diverso (per esempio, in 5.1.1 era  $p$ , in 5.1.2 era  $\lambda$ ).

Chiediamoci cosa può fare lo sperimentatore in questa situazione: la cosa più naturale è quella di “procurarsi delle osservazioni” del fenomeno di cui  $X$  è l'espressione, effettuando qualche tipo di esperimento, e decidere (in statistica si dice anche “fare inferenza”) in base ai risultati ottenuti.

Tipicamente come risultato del suo esperimento egli otterrà dei numeri  $x_1, \dots, x_n$ . Essi vanno pensati come valori assunti (nel corso dell'esperimento) da certe v.a.  $X_1, \dots, X_n$ , aventi una legge congiunta dipendente da  $\theta$ . In termini più precisi, le v.a.  $X_1, \dots, X_n$  vanno pensate definite su uno spazio di probabilità della forma  $(\Omega, \mathcal{A}, \{P^\theta; \theta \in \Theta\})$ , dove cioè la probabilità  $P^\theta$  dipende da un parametro  $\theta$ , che varia in un insieme di valori  $\Theta$ . Si dà allora la seguente:

**Definizione 5.2.1.** Una famiglia di spazi di probabilità del tipo

$$(\Omega, \mathcal{A}, \{P^\theta; \theta \in \Theta\})$$

si chiama *modello statistico parametrico*. L'insieme  $\Theta$  si chiama *insieme dei parametri*.

Chiameremo *osservazioni* di  $X$  le v. a.  $X_1, \dots, X_n$ . In corrispondenza i numeri  $x_1, \dots, x_n$  (che sono i valori assunti dalle osservazioni *dopo* che l'esperimento è stato effettuato) si chiameranno *valori osservati*.

Un caso molto frequente (ma non l'unico!) è quello in cui  $X_1, \dots, X_n$  sono tra loro indipendenti ed hanno tutte la stessa legge di  $X$ : è la formalizzazione matematica del caso in cui lo sperimentatore decide di ripetere  $n$  volte, in condizioni di indipendenza, *proprio* l'esperimento che produce  $X$ . Si dice allora che il vettore aleatorio  $(X_1, \dots, X_n)$  costituisce un *campione di numerosità (o taglia)  $n$  estratto* dalla legge di  $X$ .

**Esempio 5.2.2.** Torniamo al caso della moneta (Esempio. 5.1.1). In questo caso il parametro  $\theta$  varia nell'intervallo  $(0, 1)$  (almeno se non si hanno ulteriori informazioni che ci permettano di specificare meglio la natura di  $\theta$ ). Dunque  $\Theta = (0, 1)$  è il nostro insieme dei parametri. Supponiamo che lo sperimentatore lanci  $n$  volte la moneta. Ciò significa che egli si procura un campione di  $n$  osservazioni  $(X_1, \dots, X_n)$ , in cui le  $X_i$  sono indipendenti e tutte di legge  $\mathcal{B}(1, \theta)$ . Come sappiamo dal Calcolo delle Probabilità, la densità del vettore  $X$  è

$$P(X_1 = x_1, \dots, X_n = x_n) = q(x_1, \dots, x_n) = \theta^{\sum_{k=1}^n x_k} (1 - \theta)^{n - \sum_{k=1}^n x_k}.$$

Dunque, se, per ogni  $\theta$ , consideriamo lo schema di  $n$  prove indipendenti di parametro  $\theta$  (si veda il paragrafo 1.6), otteniamo una famiglia di spazi di probabilità  $((\Omega, \mathcal{A}, \{P^\theta; \theta \in (0, 1)\})$ , che costituisce il modello statistico di questo esempio.

**Osservazione 5.2.3.** Per noi la situazione più comune sarà quella in cui  $\theta \in \mathbb{R}$ , ma in generale  $\theta$  va pensato come un vettore (cioè la legge può dipendere da più di un parametro reale, come accade per esempio per la  $\mathcal{N}(\mu, \sigma^2)$ , quando sia la media  $\mu$  che la varianza  $\sigma^2$  non sono note).

Comunque sia, lo sperimentatore userà i numeri trovati per calcolare, a partire da essi, una stima del parametro incognito  $\theta$  (o, più in generale, di una sua funzione  $\psi(\theta) \in \mathbb{R}^k$ ); in altri termini sceglierà una opportuna (secondo lui) funzione  $t$  di  $n$  variabili reali e stimerà il parametro  $\theta$  con il numero  $t(x_1, \dots, x_n)$ . Ovviamente la funzione  $t$  andrà scelta non dipendente dal parametro incognito (dato che essa va usata appunto per stimarlo!). Queste considerazioni giustificano la seguente:

**Definizione 5.2.4.** Sia  $\psi : \Theta \rightarrow \mathbb{R}^k$  una funzione. Sia  $t : \mathbb{R}^n \rightarrow \mathbb{R}^k$  una funzione non dipendente da  $\theta$ .

(i) Si chiama *stimatore* di  $\psi(\theta)$  la v. a.  $T$  che ad ogni  $\omega \in \Omega$  associa il numero  $T(\omega) = t(X_1(\omega), \dots, X_n(\omega))$ , dove  $X_1, \dots, X_n$  sono  $n$  osservazioni. Per brevità ometteremo quasi sempre l'indicazione del simbolo  $\omega$ , scrivendo più semplicemente  $T = t(X_1, \dots, X_n)$ .

(ii) Si chiama *stima* di  $(\psi)\theta$  il numero  $t = t(x_1, \dots, x_n)$ , dove  $x_1, \dots, x_n$  sono gli  $n$  valori osservati (corrispondenti alle osservazioni del punto (i)).

**5.2.5. Stimatori corretti.** Siano  $X_1, \dots, X_n$   $n$  osservazioni aventi tutte la stessa legge, dipendente da un parametro  $\theta$ .

Sia  $T = t(X_1, \dots, X_n)$  uno stimatore di una funzione del parametro  $\psi(\theta)$ . La situazione “ideale” sarebbe che valesse l’uguaglianza

$$T(\omega) = \psi(\theta) \quad \forall \omega \in \Omega, \quad (5.2.6)$$

(o, in modo equivalente, se fosse  $T(x_1, \dots, x_n) = \psi(\theta)$  per ogni  $n$ -upla  $(x_1, \dots, x_n)$  di valori osservati), cioè che lo stimatore fornisse *sempre e esattamente* la quantità da stimare. Ciò non è ovviamente possibile; più ragionevole è chiedersi se l’uguaglianza (5.2.6) possa valere almeno *in media*; in effetti una buona proprietà di uno stimatore è la seguente:

**Definizione 5.2.7.** Lo stimatore  $T = t(X_1, \dots, X_n)$  di  $\psi(\theta)$  si dice *corretto* (o *non distorto*, *unbiased* in inglese) se vale la relazione

$$\mathbf{E}^\theta[T] = \psi(\theta) \quad \forall \theta \in \Theta.$$

*Note.* (i) In tutti gli esempi che seguono, indicheremo con  $\mu$  e  $\sigma^2$  rispettivamente la media e la varianza della comune legge delle  $X_i$ . Osservare che  $\mu$  e  $\sigma^2$  sono, naturalmente, due funzioni del parametro  $\theta$ , ma questo fatto non viene messo in risalto nella notazione usata, per motivi di brevità.

(ii) Nelle scritture del tipo  $\mathbf{E}^\theta$  o  $\mathbf{Var}^\theta$  o simili sottintenderemo il parametro  $\theta$ , (cioè scriveremo semplicemente  $\mathbf{E}$  o  $\mathbf{Var}$ ), naturalmente sempre che ciò non dia luogo ad equivoci.

(iii) Queste convenzioni di scrittura saranno tacitamente usate anche nei paragrafi successivi.

**Esempio 5.2.8.** (a) La media campionaria  $\bar{X}$  è uno stimatore corretto di  $\mu$ . Infatti

$$\mathbf{E}[\bar{X}] = \frac{1}{n} \sum_{i=1}^n \mathbf{E}[X_i] = \frac{n\mu}{n} = \mu.$$

(b) Se  $\mu$  è nota, lo stimatore

$$T = \frac{\sum_{i=1}^n (X_i - \mu)^2}{n} \quad (5.2.9)$$

è uno stimatore corretto di  $\sigma^2$ . Infatti

$$\mathbf{E} \left[ \frac{\sum_{i=1}^n (X_i - \mu)^2}{n} \right] = \frac{1}{n} \sum_{i=1}^n \mathbf{E}[(X_i - \mu)^2] = \frac{1}{n} \sum_{i=1}^n \sigma^2 = \sigma^2.$$

(c) Supponiamo ora in più che le osservazioni  $X_1, \dots, X_n$  (oltre ad essere equidistribuite) siano tra loro indipendenti, cioè costituiscano un campione.

Vogliamo trovare uno stimatore corretto di  $\sigma^2$  nel caso, molto frequente, che  $\mu$  non sia nota (in tale situazione la v. a. indicata in (5.2.9) non è uno stimatore, perché dipende da  $\mu$ ). L'idea è quella di sostituire  $\mu$  con il suo stimatore  $\bar{X}$  nella formula (5.2.9), cioè usare lo stimatore

$$Z = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n}. \quad (5.2.10)$$

Calcoliamo dunque  $\mathbf{E}[Z]$ . Cominciamo calcolando la media del numeratore della frazione in (5.2.10); sommando e sottraendo  $\mu$  all'interno della parentesi e svolgendo il quadrato si ha

$$\begin{aligned} \mathbf{E}\left[\sum_{i=1}^n (X_i - \bar{X})^2\right] &= \mathbf{E}\left[\sum_{i=1}^n ((X_i - \mu) - (\bar{X} - \mu))^2\right] \\ &= \sum_{i=1}^n \mathbf{E}[(X_i - \mu)^2] + \sum_{i=1}^n \mathbf{E}[(\bar{X} - \mu)^2] - 2\mathbf{E}\left[(\bar{X} - \mu) \sum_{i=1}^n (X_i - \mu)\right] \\ &= \sum_{i=1}^n \mathbf{E}[(X_i - \mu)^2] + n\mathbf{E}[(\bar{X} - \mu)^2] - 2n\mathbf{E}\left[(\bar{X} - \mu)^2\right] \\ &= \sum_{i=1}^n \mathbf{E}[(X_i - \mu)^2] - n\mathbf{E}[(\bar{X} - \mu)^2]. \end{aligned} \quad (5.2.11)$$

Osserviamo ora che  $\mathbf{E}[(X_i - \mu)^2] = \mathbf{Var}X_i = \sigma^2$ ; inoltre ricordando che  $\mathbf{E}[\bar{X}] = \mu$ ,

$$\mathbf{E}[(\bar{X} - \mu)^2] = \mathbf{Var}\bar{X} = \mathbf{Var}\left(\frac{\sum_{i=1}^n X_i}{n}\right) = \frac{1}{n^2} \sum_{i=1}^n \mathbf{Var}X_i = \frac{\sigma^2}{n} \quad (5.2.12)$$

(qui è stata usata l'indipendenza delle  $X_i$ ; dove?).

Usando la relazione (5.2.12) nella (5.2.11) si ottiene:

$$\mathbf{E}\left[\sum_{i=1}^n (X_i - \bar{X})^2\right] = n\sigma^2 - n\frac{\sigma^2}{n} = (n-1)\sigma^2.$$

Dunque

$$\mathbf{E}[Z] = \frac{n-1}{n}\sigma^2, \quad (5.2.13)$$

cioè  $Z$  è uno stimatore distorto! Tuttavia il calcolo appena fatto ci dice che è corretto lo stimatore (non molto diverso da  $Z$  per  $n$  grande)

$$S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}.$$

**Osservazione 5.2.14.** Abbiamo già incontrato la v.a.  $S^2$ . Dove?

**5.2.15. Confronto di stimatori.** Sostituendo il vero valore di  $\psi(\theta)$  con il suo stimatore  $T = t(X_1, \dots, X_n)$  si commette un errore, che è opportuno saper misurare. Si dà dunque la seguente:

**Definizione 5.2.16.** Si chiama *rischio quadratico (medio)* dello stimatore  $T$  la funzione  $\theta \mapsto R_T(\theta)$  definita su  $\Theta$  da

$$R_T(\theta) = \mathbf{E}^\theta[(T - \psi(\theta))^2].$$

**Osservazione 5.2.17.** La definizione data sopra si motiva come abbiamo fatto a suo tempo per la varianza (vedere le considerazioni fatte prima della Definizione 3.1.15).

**Osservazione 5.2.18.** Se lo stimatore  $T$  è corretto, si ha evidentemente

$$R_T(\theta) = \mathbf{Var}^\theta T.$$

In presenza di due stimatori  $S$  e  $T$  della quantità  $\psi(\theta)$ , preferiremo ovviamente lo stimatore con rischio più piccolo, *per ogni valore di  $\theta$*  (sempre che uno dei due realizzi questa richiesta). Cioè:

**Definizione 5.2.19.** (i) Si dice che  $S$  è *preferibile a* (o *non peggiore di*)  $T$  se  $R_S(\theta) \leq R_T(\theta) \forall \theta \in \Theta$ ; se in più  $\exists \theta_0 \in \Theta$  tale che  $R_S(\theta_0) < R_T(\theta_0)$ , allora si dice che  $S$  è *strettamente preferibile a* (o *migliore di*)  $T$ .

(ii) Uno stimatore è detto *ammissibile* in un'assegnata classe di stimatori  $\mathcal{C}$  se in  $\mathcal{C}$  non esistono stimatori ad esso strettamente preferibili.

**5.2.20. Stimatori consistenti.** Molte proprietà degli stimatori (ad esempio la correttezza) hanno bisogno solo di un numero finito di osservazioni  $X_1, \dots, X_n$  (cioè in questo momento  $n$  va pensato come un numero intero fissato).

Talvolta, però, è utile conoscere il comportamento *asintotico* di uno stimatore (cioè per  $n \rightarrow \infty$ ), e in tal caso bisogna immaginare di avere a disposizione una *successione* di osservazioni  $X_1, X_2, X_3, \dots$ .

Ad esempio, uno stimatore assegnato può non essere corretto per nessun valore finito di  $n$ , ma “tendere a diventare corretto” quando  $n \rightarrow \infty$ . In tal caso, se possibile e in pratica non troppo costoso, questo potrebbe indurci ad aumentare il numero delle osservazioni, in modo da avvicinarci in modo abbastanza soddisfacente alla correttezza. (Questo è quello che accade ad esempio per lo stimatore  $Z$  della varianza  $\sigma^2$  definito in (5.2.10): abbiamo visto che  $Z$  non è corretto, ma si ha (si veda (5.2.13))

$$\lim_{n \rightarrow \infty} \mathbf{E}[Z] = \lim_{n \rightarrow \infty} \frac{n-1}{n} \sigma^2 = \sigma^2,$$

e dunque  $Z$  è *asintoticamente corretto*.



Un'altra proprietà asintotica che può essere importante per uno stimatore  $T_n = t_n(X_1, \dots, X_n)$  è la “consistenza”. Essa riguarda la funzione di ripartizione di  $T_n$  e la sua variazione al crescere di  $n$  (per sottolineare il fatto che in questo momento stiamo parlando di una *successione* di stimatori, li indichiamo con  $T_n$  invece che semplicemente con  $T$ ). Al solito indichiamo con  $\psi(\theta)$  una funzione del parametro  $\theta$ .

**Definizione 5.2.21.** Una successione  $(T_n)$  di stimatori di  $\psi(\theta)$  si dice *fortemente consistente* se  $T_n$  converge verso  $\psi(\theta)$  quasi certamente.

**Esempio 5.2.22.** Per la Legge Forte dei Grandi Numeri, la media campionaria

$$\bar{X} = \frac{X_1 + \dots + X_n}{n}$$

è uno stimatore fortemente consistente della media. (Per essere precisi, dovremmo mettere un indice  $n$  a  $\bar{X}$ , poichè  $n$  varia, e meglio sarebbe dire “la successione delle medie campionarie”).

**Definizione 5.2.23.** Una successione  $(T_n)$  di stimatori di  $\psi(\theta)$  si dice *debolmente consistente* se  $T_n$  converge verso  $\psi(\theta)$  in probabilità.

Nella pratica, per trovare stimatori consistenti, è utile il seguente risultato, di semplice dimostrazione.

**Teorema 5.2.24.** Sia  $(T_n)$  una successione di stimatori fortemente consistenti del parametro  $\theta$ , e supponiamo che  $\theta \mapsto \psi(\theta)$  sia una funzione continua. Allora  $U_n = \psi(T_n)$  è una successione di stimatori fortemente consistenti di  $\psi(\theta)$ .

**Esempio 5.2.25 (tipico).** Sia  $X$  una v. a. avente densità

$$f_\alpha(x) = \begin{cases} \alpha x^{\alpha-1} & 0 < x < 1; \\ 0 & \text{altrove,} \end{cases}$$

dove  $\alpha > 0$  non è noto. Trovare una successione di stimatori consistenti per  $\alpha$ .

*Soluzione.* In questo, come in altri casi simili, il trucco è quello di considerare la legge dipendente *non* dal parametro  $\alpha$ , ma dal parametro  $\mu = \mathbf{E}[X]$ , nel modo seguente. Calcoliamo prima di tutto  $\mu$  (ovviamente in funzione di  $\alpha$ !). Si ha

$$\mu = \mathbf{E}[X] = \int_0^1 x \alpha x^{\alpha-1} = \frac{\alpha}{\alpha+1}. \quad (5.2.26)$$

Si osserva intanto che  $0 < \mu < 1$ . Inoltre, invertendo la relazione (5.2.26), si ottiene

$$\alpha = \frac{\mu}{1-\mu} := \psi(\mu).$$

Nell'intervallo aperto  $(0, 1)$ ,  $\mu \mapsto \psi(\mu)$  è una funzione continua. Sappiamo (Esempio 5.2.22) che  $T_n = \overline{X}$  è uno stimatore fortemente consistente di  $\mu$ . Posto allora

$$U_n = \psi(\overline{X}) = \frac{\overline{X}}{1 - \overline{X}}, \quad (5.2.27)$$

si ottiene dal Teorema 5.2.23 che  $(U_n)$  è uno stimatore fortemente consistente di  $\alpha$ .

**Osservazione 5.2.28.** L'espressione al secondo membro della relazione (5.2.27) non ha senso sull'evento  $\{\overline{X} = 1\}$ , ma si può dimostrare che questo evento ha probabilità nulla.

Come si sarà già intuito dopo aver letto il paragrafo sugli stimatori corretti, generalmente si adotta qualche particolare criterio per individuare uno stimatore adatto (ad esempio, appunto, la correttezza). Nelle due prossime sezioni vedremo altri due criteri importanti.

**5.2.29. Stimatori dei momenti.** Sia  $X$  una v. a. la cui legge dipende da un certo numero di parametri  $\theta_1, \dots, \theta_r$  (il “parametro”  $\theta$  è in generale un vettore,  $\theta = (\theta_1, \dots, \theta_r)$  appunto). Supponiamo che  $\theta_1, \dots, \theta_r$  non siano noti, e come al solito il nostro scopo è darne una stima dipendente dalle osservazioni  $(X_1, \dots, X_n)$ .

Ricordiamo che in (3.1.15) abbiamo dato la definizione di momento di ordine  $k$  di una v. a.  $X$ . Si tratta della quantità  $m_k = \mathbf{E}[X^k]$  (a patto che  $\mathbf{E}[|X|^k] < +\infty$ ).

**Osservazione 5.2.30.** Dato che la legge di  $X$  dipende da  $\theta_1, \dots, \theta_r$ , lo stesso accadrà per il momento teorico  $m_k$  (che, a sua volta, dipende solo dalla legge di  $X$ ); in altre parole esisterà una funzione  $f_k$  di  $r$  variabili tale che

$$m_k = f_k(\theta_1, \dots, \theta_r).$$

Supponiamo che la v. a.  $X$  ammetta i primi  $q$  momenti ( $q$  è un numero intero  $\geq 1$ ). Ciò significa che possiamo scrivere la relazione precedente  $\forall k = 1, \dots, q$ , ottenendo così il sistema

$$\begin{cases} m_1 = f_1(\theta_1, \dots, \theta_r) \\ m_2 = f_2(\theta_1, \dots, \theta_r) \\ \cdot \\ \cdot \\ m_q = f_q(\theta_1, \dots, \theta_r). \end{cases}$$

Si tratta evidentemente di un sistema di  $q$  equazioni nelle  $r$  incognite  $\theta_1, \dots, \theta_r$ , che si può cercare di risolvere. Se questo è possibile, otterremo  $r$  espressioni

del tipo seguente

$$\begin{cases} \theta_1 = g_1(m_1, m_2, \dots, m_q) \\ \theta_2 = g_2(m_1, m_2, \dots, m_q) \\ \cdot \\ \cdot \\ \cdot \\ \theta_r = g_r(m_1, m_2, \dots, m_q). \end{cases} \quad (5.2.31)$$

**Definizione 5.2.32.** Supponiamo che le osservazioni  $(X_1, \dots, X_n)$  siano tra loro indipendenti. Si definisce *momento empirico* di  $X$  la quantità (aleatoria!)

$$\hat{m}_k := \frac{\sum_{i=1}^n X_i^k}{n}.$$

Per la Legge dei Grandi Numeri, si ha

$$\hat{m}_k = \frac{\sum_{i=1}^n X_i^k}{n} \xrightarrow{P} \mathbf{E}[X^k] = m_k, \quad n \rightarrow \infty. \quad (5.2.33)$$

**Osservazione 5.2.34.** Questa relazione spiega i nomi di momento *teorico* e momento *empirico* dati alle due quantità sopra definite; ci dice anche che i momenti empirici sono stimatori consistenti di quelli teorici (rivedere la definizione di consistenza).

La relazione (5.2.33) suggerisce il procedimento seguente: dato che per  $n$  grande  $\hat{m}_k \simeq m_k$ , nel sistema (5.2.31) sostituiamo  $\hat{m}_k$  al posto di  $m_k$  per ogni  $k = 1, \dots, q$ , ottenendo le relazioni

$$\begin{cases} \theta_1 \simeq g_1(\hat{m}_1, \hat{m}_2, \dots, \hat{m}_q) \\ \theta_2 \simeq g_2(\hat{m}_1, \hat{m}_2, \dots, \hat{m}_q) \\ \cdot \\ \cdot \\ \cdot \\ \theta_r \simeq g_r(\hat{m}_1, \hat{m}_2, \dots, \hat{m}_q), \end{cases}$$

ovvero avremo “espresso” (approssimativamente!) ciascuno dei parametri in termini delle v. a.  $\hat{m}_1, \hat{m}_2, \dots, \hat{m}_q$  che sono note, perché dipendenti solo dalle osservazioni.

Dunque, a sua volta, ciascuna delle funzioni  $g_i(\hat{m}_1, \hat{m}_2, \dots, \hat{m}_q)$ , per ogni  $i = 1, \dots, r$ , dipende solo dalle osservazioni, ed è dunque uno stimatore  $\hat{\theta}_i = \hat{\theta}_i(X_1, \dots, X_n)$  di  $\theta_i$ .

**Definizione 5.2.35.** Il vettore di v. a.  $\hat{\theta} = (\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_r)$  si chiama *stimatore dei momenti* del parametro  $\theta = (\theta_1, \dots, \theta_r)$ .

**Osservazione 5.2.36.** Se si deve stimare con il metodo dei momenti una funzione del parametro  $\theta$ , sia essa  $\psi(\theta) = \psi(\theta_1, \dots, \theta_r)$ , si prende lo stimatore

$$\psi(\hat{\theta}_1, \dots, \hat{\theta}_r).$$

**Osservazione 5.2.37.** Seguendo la definizione (5.2.4), chiameremo *stime dei momenti* i numeri che si ottengono mettendo i valori osservati, cioè  $x_1, \dots, x_n$  al posto delle osservazioni  $X_1, \dots, X_n$  nelle espressioni, sopra definite,  $\theta_i(X_1, \dots, X_n)$ .

**Osservazione 5.2.38.** Spesso il numero  $q$  dei momenti che servono per fare i conti è uguale al numero  $r$  dei parametri da stimare, cioè si cerca di scrivere (e risolvere) un sistema di  $r$  equazioni in  $r$  incognite.

**Esempio 5.2.39.** Stimatore dei momenti del parametro della legge esponenziale basato sulle osservazioni  $(X_1, \dots, X_n)$ .

È noto che, se  $X \sim \mathcal{E}(\lambda)$ , allora

$$m_1 = \mathbf{E}[X] = \frac{1}{\lambda} \neq 0$$

e, risolvendo, si trova

$$\lambda = \frac{1}{m_1}.$$

Si pone allora

$$\hat{\lambda} = \frac{1}{\hat{m}_1} = \frac{n}{X_1 + \dots + X_n}.$$

Naturalmente  $\hat{\lambda}$  è definito solo sull'evento  $\{X_1 + \dots + X_n \neq 0\}$  (che però si può dimostrare avere probabilità nulla).

Gli stimatori dei momenti sono spesso distorti, ma consistenti (come si può intuire dalla consistenza dei momenti empirici, di cui essi sono funzione). Tuttavia in generale essi non sono dei buoni stimatori.

**5.2.40. Stimatori di massima verosimiglianza.** Stimatori con buone proprietà asintotiche si ottengono con il *metodo della massima verosimiglianza*, che vedremo in questa sezione.

**Esempio introduttivo 5.2.41.** Una moneta è truccata; della probabilità  $\theta$  che essa dia “testa” si sa soltanto che essa vale  $1/1000$  oppure  $999/1000$ . Un tizio deve stabilire quale dei due valori è quello vero, basandosi su qualche tipo di osservazione, a sua scelta. Egli allora decide di effettuare 100 lanci, e ottiene in ciascun lancio la faccia “testa”. A questo punto, come è facile capire, egli è propenso a credere che il vero valore di  $\theta$  sia  $999/1000$ . Ritene infatti che il risultato ottenuto (100 volte “testa” in 100 lanci) potrebbe sì verificarsi

anche se la moneta fosse truccata nell'altro modo ( $\theta = 1/1000$ ), ma con una probabilità inferiore.

Cerchiamo di formalizzare quello che è accaduto. Indichiamo con  $X$  il risultato di un generico lancio:

$$X = \begin{cases} 1 & \text{se esce "testa"} \\ 0 & \text{se esce "croce"} \end{cases}.$$

Allora i 100 lanci effettuati costituiscono un campione di 100 osservazioni del fenomeno, che indicheremo, come sempre, con  $(X_1, \dots, X_{100})$ . Sulla legge di  $X$  l'informazione è la seguente:

$$X \sim B(1, \theta), \text{ con } \theta \in \Theta = \left\{ \frac{1}{1000}, \frac{999}{1000} \right\}.$$

Allora la probabilità di ottenere 100 volte "croce" in 100 lanci è

$$P^\theta(X_1 = 1, X_2 = 1, \dots, X_{100} = 1) = \theta^{100} = \begin{cases} \left(\frac{1}{1000}\right)^{100} & \text{se } \theta = \frac{1}{1000}, \\ \left(\frac{999}{1000}\right)^{100} & \text{se } \theta = \frac{999}{1000}. \end{cases}$$

Il tizio ha dunque deciso di considerare vero il valore di  $\theta$  per il quale il risultato effettivamente osservato è il più probabile. In altre parole, egli ha calcolato

$$\max_{\theta \in \Theta} P^\theta(X_1 = 1, X_2 = 1, \dots, X_{100} = 1) = \max_{\theta \in \Theta} \theta^{100},$$

ed ha deciso per il valore del parametro in cui tale massimo è raggiunto, cioè per il *punto di massimo* della funzione

$$\theta \mapsto P^\theta(X_1 = 1, X_2 = 1, \dots, X_{100} = 1) = \theta^{100}.$$

Questo esempio suggerisce il procedimento che ora descriveremo (metodo per il calcolo dello *stimatore di massima verosimiglianza*).

Sia  $X$  una v.a. la cui legge dipende da un parametro  $\theta$ . Disponiamo di  $n$  osservazioni di  $X$  (non necessariamente indipendenti, come invece capitava nell'esempio), che indichiamo con  $(X_1, \dots, X_n)$ . Per semplicità, in questo momento supporremo che la legge congiunta delle osservazioni sia discreta, e indicheremo con  $P^\theta$  la loro densità congiunta (in seguito toglieremo questa restrizione). Dopo avere effettuato l'esperimento, il campione di osservazioni avrà prodotto un campione di  $n$  valori osservati (ricordare la distinzione che abbiamo fatto tra i termini *osservazione* e *valore osservato*),  $(x_1, \dots, x_n)$ . La probabilità (in funzione di  $\theta \in \Theta$ ) che il risultato sia quello *effettivamente* osservato è

$$\theta \mapsto P^\theta(X_1 = x_1, \dots, X_n = x_n), \quad \theta \in \Theta.$$

Essa è evidentemente una funzione di  $\theta$  dipendente dai valori osservati. Nel caso generale (cioè se il vettore di osservazioni non è necessariamente discreto) essa si chiama *funzione di verosimiglianza* (*likelihood function* in inglese) e il simbolo usato sarà piuttosto

$$\theta \mapsto L(\theta; x_1, \dots, x_n), \quad \theta \in \Theta.$$

(In questa scrittura, un po' ambigua, la variabile è  $\theta$ , mentre i valori osservati  $x_1, \dots, x_n$  devono essere pensati come dei numeri assegnati.) Vedremo fra poco come calcolare  $L$  in alcuni casi importanti.

Supponiamo di essere riusciti a trovare (in qualche modo) il massimo della funzione di verosimiglianza (al variare di  $\theta \in \Theta$ ); il corrispondente punto di massimo dipenderà anch'esso da  $x_1, \dots, x_n$ ; indichiamolo, come d'uso, con  $\hat{\theta} = \hat{\theta}(x_1, \dots, x_n)$ .

**Definizione 5.2.42.** Il numero  $\hat{\theta}(x_1, \dots, x_n)$  si chiama *stima di massima verosimiglianza* di  $\theta$  (in corrispondenza dei valori osservati  $x_1, \dots, x_n$ ).

È evidente dalla definizione che la stima di massima verosimiglianza è funzione delle variabili reali  $x_1, \dots, x_n$ . Dunque:

**Definizione 5.2.43.** La v.a.  $\hat{\theta}(X_1, \dots, X_n)$  (in cui semplicemente abbiamo messo le v.a.  $X_1, \dots, X_n$  al posto dei numeri  $x_1, \dots, x_n$ ) si chiama *stimatore di massima verosimiglianza* di  $\theta$ .

**Osservazione 5.2.44.** Normalmente lo stimatore di massima verosimiglianza è ancora indicato con  $\hat{\theta}$ , anche se la notazione può generare confusione tra stimatore e stima.

Occupiamoci ora un po' più in dettaglio di alcuni metodi per il calcolo della funzione di verosimiglianza. Una situazione semplice (del resto già vista sopra) è quella in cui il vettore delle osservazioni ha densità (congiunta) discreta. Un caso particolare di questa situazione si ha quando le osservazioni costituiscono un campione, cioè sono tra loro indipendenti. Supponiamo cioè che la v.a.  $X$  abbia densità discreta  $p_\theta(x) = P^\theta(X = x)$ . Allora

$$P^\theta(X_1 = x_1, \dots, X_n = x_n) = p_\theta(x_1) \cdots p_\theta(x_n) := L(\theta; x_1, \dots, x_n).$$

Per analogia, se  $X$  è assolutamente continua, con densità  $f_\theta(x)$  (e naturalmente le osservazioni sono ancora tra loro indipendenti), la funzione di verosimiglianza è definita dalla formula

$$L(\theta; x_1, \dots, x_n) := f_\theta(x_1) \cdots f_\theta(x_n).$$

Per concludere, diamo un rapido elenco di alcuni metodi usati per trovare il punto di massimo di  $L$ .

(i) Come nell'esempio della moneta fatto all'inizio, supponiamo che l'insieme  $\Theta$  sia costituito da un numero finito di elementi  $\theta_1, \dots, \theta_M$ . Se  $M$  non è troppo grande, un modo del tutto elementare consiste nel calcolare  $L(\theta_i)$   $\forall i = 1, \dots, M$  e scegliere il valore di  $\theta_i$  per cui  $L(\theta_i)$  risulta massimo. Se  $M$  è grande, bisognerà ricorrere a qualche trucco, da vedere caso per caso.

(ii) Se  $\Theta$  è un intervallo della retta (eventualmente non limitato), si possono applicare i metodi studiati in “Analisi I” per trovare i punti di massimo e minimo di una funzione: in particolare, la stima di massima verosimiglianza, se interna all'intervallo  $\Theta$ , è soluzione dell'equazione

$$\frac{d}{d\theta} L(\theta; x_1, \dots, x_n) = 0.$$

Attenzione agli (eventuali) estremi dell'intervallo, per i quali non si può applicare il criterio dei punti stazionari.

(iii) Nel caso che la stima di massima verosimiglianza sia stata ottenuta con il solo criterio dei punti stazionari, è necessario ricordare che il fatto che un punto sia stazionario non garantisce automaticamente che esso sia di massimo, né che sia di massimo assoluto (il criterio della derivata prima fornisce punti di solo *estremo relativo*). Di regola, bisognerebbe proseguire nell'indagine con i metodi imparati in Analisi. Noi ci accontenteremo quasi sempre della sola stazionarietà, per non appesantire i calcoli (in alcuni casi esistono criteri specifici, che garantiscono la correttezza del procedimento).

(iv) Spesso il metodo del punto (ii) può essere semplificato passando al logaritmo prima di effettuare la derivazione di  $L$ ; si deriva cioè *non* la funzione  $\theta \mapsto L(\theta; x_1, \dots, x_n)$ , ma  $\theta \mapsto \log L(\theta; x_1, \dots, x_n)$ . Il passaggio al logaritmo non cambia infatti i punti stazionari, né la loro natura (perché?). In questi casi, dunque, la stima di massima verosimiglianza viene trovata come soluzione dell'equazione

$$\frac{d}{d\theta} \log L(\theta; x_1, \dots, x_n) = 0,$$

che è nota come *equazione di verosimiglianza*. Di essa esiste una versione multidimensionale (che si utilizza quando il parametro  $\theta$  è un vettore), ma la trattazione va oltre gli scopi di questi appunti (in particolare, il caso multidimensionale ha bisogno dei metodi dell’“Analisi II”).

**Osservazione 5.2.45.** Per stimare con il metodo della massima verosimiglianza una funzione  $\psi(\theta)$  del parametro  $\theta$ , si utilizza la quantità  $\psi(\hat{\theta})$ , dove  $\hat{\theta}$  è la stima (o lo stimatore) di massima verosimiglianza di  $\theta$ . Non daremo una giustificazione teorica di questa regola, che va sotto il nome di *proprietà di invarianza*; provare a farne la dimostrazione nell'ipotesi ulteriore che  $\psi$  sia strettamente monotona.

**Esempio 5.2.46.** Calcolare la stima e lo stimatore di massima verosimiglianza del parametro  $\theta$  della legge bernoulliana  $B(1, \theta)$ , basato sul campione  $(X_1, \dots, X_n)$ . Si suppone che  $\theta \in (0, 1)$ .

Sia  $X$  una v. a. di legge  $B(1, \theta)$ ; la sua densità, come è noto, è

$$P^\theta(X = x) = \theta^x(1 - \theta)^{1-x}, \quad x \in \{0, 1\}.$$

Dunque la funzione di verosimiglianza è

$$\begin{aligned} \theta \mapsto L(\theta; x_1, \dots, x_n) &= \theta^{x_1}(1 - \theta)^{1-x_1} \dots \theta^{x_n}(1 - \theta)^{1-x_n} \\ &= \theta^{x_1 + \dots + x_n} (1 - \theta)^{n - (x_1 + \dots + x_n)}, \end{aligned}$$

per  $(x_1, \dots, x_n) \in \{0, 1\}^n$ . Per semplicità poniamo

$$\alpha = x_1 + \dots + x_n.$$

L'equazione di verosimiglianza è

$$0 = \frac{d}{d\theta} \log L(\theta; x_1, \dots, x_n) = \frac{d}{d\theta} (\alpha \log \theta + (n - \alpha) \log(1 - \theta)) = \frac{\alpha}{\theta} - \frac{n - \alpha}{1 - \theta},$$

la cui unica soluzione è

$$\hat{\theta} = \frac{\alpha}{n} = \frac{x_1 + \dots + x_n}{n} = \bar{x}$$

(media del campione di valori osservati). È semplice vedere che il valore così trovato è effettivamente un punto di massimo (verifica per esercizio), e si conclude che esso è la stima di massima verosimiglianza di  $\theta$  cercata. Lo stimatore di massima verosimiglianza è dunque

$$\hat{\theta} = \frac{X_1 + \dots + X_n}{n} = \bar{X},$$

ovvero la media campionaria delle osservazioni.

**Esempio 5.2.47.** Calcolare la stima e lo stimatore di massima verosimiglianza del parametro  $\theta$  della legge esponenziale  $\mathcal{E}(\theta)$ , basato sul campione  $(X_1, \dots, X_n)$ . Si suppone che  $\theta \in (0, +\infty)$ .

Sia  $X$  una v. a. di legge  $\mathcal{E}(\theta)$ ; la sua densità (assolutamente continua), come è noto, è

$$f_\theta(x) = \begin{cases} \theta e^{-\theta x} & \text{per } x > 0 \\ 0 & \text{altrove.} \end{cases}$$



Dunque la funzione di verosimiglianza è

$$\begin{aligned}\theta \mapsto L(\theta; x_1, \dots, x_n) &= \begin{cases} \theta e^{-\theta x_1} \dots \theta e^{-\theta x_n} & x_i > 0 \ \forall i, \\ 0 & \text{altrove} \end{cases} \\ &= \begin{cases} \theta^n e^{-\theta(x_1 + \dots + x_n)} & x_i > 0 \ \forall i, \\ 0 & \text{altrove} . \end{cases}\end{aligned}$$

È chiaro che il punto di massimo di questa funzione si troverà nel quadrante  $\{x_1 > 0, \dots, x_n > 0\}$  (dove la funzione è strettamente positiva). Ponendo di nuovo, come nell'esempio precedente,

$$\alpha = x_1 + \dots + x_n.$$

l'equazione di verosimiglianza è

$$0 = \frac{d}{d\theta} (\log(\theta^n e^{-\theta\alpha})) = \frac{d}{d\theta} (n \log \theta - \theta\alpha) = \frac{n}{\theta} - \alpha,$$

da cui si ricava

$$\hat{\theta} = \frac{n}{\alpha} = \frac{n}{x_1 + \dots + x_n} = \frac{1}{\bar{x}}.$$

Tralasciamo anche qui la verifica che questo è effettivamente il punto di massimo della funzione nell'intervallo considerato  $(0, +\infty)$ , e dunque la stima di massima verosimiglianza. Corrispondentemente, lo stimatore sarà

$$\hat{\theta} = \frac{1}{\overline{X}}.$$

Supponiamo ora di dover stimare la media dell'esponenziale. Dato che essa è uguale a  $\psi(\theta) = 1/\theta$ , per il criterio dell'Osservazione 5.2.45 si ottiene lo stimatore

$$\psi(\hat{\theta}) = \overline{X}.$$

Dunque, in questo caso, lo stimatore di massima verosimiglianza della media è la media campionaria.

## 5.3 I quantili

Sia  $F$  la f.d.r. di una assegnata v.a.  $X$ . Come sappiamo, essa fornisce, per ogni numero reale  $x$ , la probabilità che  $X$  assuma un valore non superiore a  $x$ . Spesso accade di dover procedere “al contrario”; più precisamente, assegnato un numero reale  $\alpha \in (0, 1)$ , ci interessa conoscere qual è il valore di  $x$  tale che sia esattamente uguale ad  $\alpha$  la probabilità che  $X$  assuma un valore non superiore a  $x$ . In formula, si tratta di risolvere rispetto a  $x$  l'equazione

$$F(x) = \alpha.$$

**Esempio 5.3.1.** Sia  $X$  una v. a. avente legge normale standard. Per  $\alpha = 0,85$ , l'equazione precedente diventa in questo caso

$$\Phi(x) = 0.85,$$

e un'occhiata alle tavole della f.d.r. della normale standard ci dice che  $x \approx 1,04$ . (Fare attenzione al fatto che per rispondere le tavole della legge normale standard vanno lette “da dentro a fuori”, non viceversa).

Naturalmente, come per ogni altro tipo di equazione, anche l'equazione  $F(x) = \alpha$  può non avere soluzioni, oppure averne più di una.

**Esempio 5.3.2.** (i) Consideriamo la f.d.r.  $F$  definita da

$$F(x) = \begin{cases} 0 & \text{per } x < 0 \\ x & \text{per } 0 \leq x < 1/2 \\ 1 & \text{per } x \geq 1/2. \end{cases}$$

Dato che  $F$  assume tutti e soli i valori appartenenti all'insieme  $[0, 1/2) \cup \{1\}$  (fare il grafico di  $F$ !), l'equazione  $F(x) = 0,7$  (ad esempio) non ha alcuna soluzione.

(ii) Consideriamo la f.d.r.  $F$  definita da

$$F(x) = \begin{cases} 0 & \text{per } x < 0 \\ 1/3 & \text{per } 0 \leq x < 1 \\ 1 & \text{per } x \geq 1 \end{cases}$$

(di quale v.a questa è la f.d.r.?). L'equazione  $F(x) = 1/3$  ha infinite soluzioni: tutti i numeri  $x$  appartenenti all'intervallo  $[0, 1)$ .

Gli esempi precedenti suggeriscono che l'esistenza e l'unicità della soluzione dell'equazione  $F(x) = \alpha$  dipendono essenzialmente dalle proprietà di  $F$ . In particolare: nell'esempio (i) la soluzione non esiste perché  $F$  non assume tutti i valori compresi tra 0 e 1, in quanto non *continua*; nell'esempio (ii) la soluzione non è unica perché  $F$  è costante sull'intervallo  $[0, 1)$  (cioè non è *strettamente crescente*). In effetti, si può dimostrare che se  $F$  è continua e strettamente crescente, allora l'equazione  $F(x) = \alpha$  ammette una e una sola soluzione. In realtà basta un po' meno; più precisamente:

**Definizione 5.3.3.** Sia  $F$  una f.d.r. continua su  $\mathbb{R}$ , non nulla su un intervallo  $I$  (eventualmente non limitato) contenuto in  $\mathbb{R}$ , e strettamente crescente su  $I$ , e sia  $\alpha \in (0, 1)$  un numero reale assegnato. Allora (si può dimostrare che) l'equazione

$$F(x) = \alpha$$

ammette una e una sola soluzione. Tale soluzione si chiama *quantile di ordine*  $\alpha$  della funzione  $F$  (o anche della legge di cui  $F$  è la f.d.r.). In contesti generali, il quantile di  $F$  verrà indicato con  $x_\alpha$ .

**Osservazione 5.3.4.** Per quanto ovvio, è utile sottolineare che la funzione  $\alpha \mapsto x_\alpha$  non è altro che la funzione inversa di  $F$ . In altre parole, si hanno le identità  $x_\alpha = F^{-1}(\alpha)$  e  $F(x_\alpha) = F(F^{-1}(\alpha)) = \alpha$ .

Esistono dei simboli ormai standard per indicare i quantili delle principali leggi della statistica, ai quali noi ci adegueremo. Precisamente indicheremo

- (i) con  $\phi_\alpha$  i quantili della legge normale standard (tavola a p. 144);
- (ii) con  $t_{\alpha,n}$  i quantili della  $t$  di Student a  $n$  gradi di libertà (tavola a p. 146);
- (iii) con  $\chi^2_{\alpha,n}$  i quantili della  $\chi^2$  a  $n$  gradi di libertà (tavola a p. 145),

dove, in tutti i casi,  $\alpha$  = ordine del quantile.

**Osservazione 5.3.5.** Sia  $F$  la f.d.r. di una v.a. continua, simmetrica (Definizione 4.4.1) e soddisfacente le ipotesi della Definizione 5.3.3. Allora per ogni  $\alpha \in (0, 1)$  si ha

$$x_\alpha = -x_{1-\alpha}.$$

*Dimostrazione.* Basta dimostrare che  $F(x_\alpha) = F(-x_{1-\alpha})$  (perché?). Per l'Osservazione 5.3.4 si ha  $F(x_\alpha) = \alpha$ ; d'altra parte, per la simmetria di  $X$ , per ogni  $t$  reale vale la relazione  $F(-t) = 1 - F(t)$ , e quindi  $F(-x_{1-\alpha}) = 1 - F(x_{1-\alpha}) = 1 - (1 - \alpha) = \alpha$  (abbiamo usato un'altra volta l'Osservazione 5.3.4).  $\square$

## 5.4 Intervalli di fiducia (o di confidenza)

*Motivazione.* Spesso nei problemi di stima di un parametro incognito  $\theta$  (o di una sua funzione  $\psi(\theta)$ ) piuttosto che dare un valore “vicino” a quello reale (cioè cercare uno stimatore), è preferibile trovare un intervallo di valori (con estremi dipendenti dalle osservazioni) tale che si possa ritenere che il vero valore della quantità da stimare vi appartenga con una probabilità non troppo bassa.

*Nota.* In questo paragrafo, riguardante gli intervalli di fiducia, e nei successivi, in cui parleremo di test, ogni funzione delle osservazioni, non dipendente dal parametro incognito  $\theta$ , sarà detta *statistica*. Dunque il concetto di statistica non è sostanzialmente diverso da quello di *stimatore*, che abbiamo dato in precedenza (Definizione 5.2.4). Il nome cambia a causa del differente punto di vista in cui ci mettiamo: ora non si tratta solo di dare un valore approssimato del parametro, ma, come vedremo, di utilizzare la nostra funzione delle osservazioni per confronti più specifici.

**Definizione 5.4.1.** Sia  $\alpha \in (0, 1)$  un numero fissato. Date due statistiche  $T_1 = t_1(X_1, \dots, X_n)$  e  $T_2 = t_2(X_1, \dots, X_n)$ , si dice che  $I = [T_1, T_2]$  è un *intervallo di fiducia* (o *di confidenza*) per  $\psi(\theta)$  di *livello* (maggiore o uguale

a)  $1 - \alpha$  se,  $\forall \theta \in \Theta$  si ha

$$P^\theta(I \text{ contiene } \psi(\theta)) = P^\theta(\psi(\theta) \in I) \geq 1 - \alpha.$$

**Osservazione 5.4.2.** La scrittura  $P^\theta(\psi(\theta) \in I)$  è un po' ambigua perchè fa pensare che  $\psi(\theta)$  sia una quantità aleatoria; è al contrario  $I$  ad essere casuale, dato che i suoi estremi  $T_1$  e  $T_2$  dipendono dalle osservazioni.

**Osservazione 5.4.3.** Tipicamente il valore di  $\alpha$  è piccolo ( $\alpha = 0.05$ ;  $\alpha = 0.01, \dots$ ).

**Osservazione 5.4.4.** Il significato della definizione (5.4.2) è il seguente: in base alle osservazioni che abbiamo fatto, possiamo dire che  $T_1 \leq \psi(\theta) \leq T_2$  con probabilità almeno  $1 - \alpha$ , e questo  $\forall \theta$ .

**Esempio 5.4.5.** Costruire un intervallo di fiducia di livello 0.95 per il parametro  $\theta$  dell'esponenziale, basato su una sola osservazione  $X$ .

Significa che si devono trovare due funzioni  $t_1(X)$  e  $t_2(X)$  tali che

$$P^\theta(t_1(X) \leq \theta \leq t_2(X)) \geq 0.95.$$

Partiamo da questa semplice osservazione: se  $X \sim \mathcal{E}(\theta)$ , allora la variabile  $Y = \theta X \sim \mathcal{E}(1)$ . Infatti, per  $t > 0$  si ha

$$P(Y \leq t) = P(\theta X \leq t) = P(X \leq t/\theta) = 1 - e^{-t}.$$

Di conseguenza,  $\forall a, b > 0$  con  $a < b$  si ottiene

$$P^\theta(a \leq Y \leq b) = (1 - e^{-b}) - (1 - e^{-a}) = e^{-a} - e^{-b},$$

il che equivale a

$$P^\theta\left(\frac{a}{X} \leq \theta \leq \frac{b}{X}\right) = e^{-a} - e^{-b}.$$

Allora poniamo

$$t_1(X) = \frac{a}{X}, \quad t_2(X) = \frac{b}{X},$$

dove le costanti  $a$  e  $b$  sono scelte in modo che  $e^{-a} - e^{-b} = 0.95$ .

**Osservazione 5.4.6.** Il metodo qui seguito, e che useremo sistematicamente, è quello della *quantità pivotale*, che consiste nel determinare una funzione di  $X_1, \dots, X_n$  e del parametro  $\theta$ , monotona in  $\theta$  (dunque invertibile), che indichiamo genericamente con  $Q(X_1, \dots, X_n, \theta)$  ( $\theta X$  nell'esempio), in modo che la legge  $P^\theta(Q(X_1, \dots, X_n, \theta) \in A)$  non dipenda da  $\theta$  (e, aggiungiamo, tale sia in qualche modo possibile farci dei conti).

*Nota.* Gli intervalli che costruiremo saranno tutti di livello  $1 - \alpha$  fissato.

Nei paragrafi 5.4.7, 5.4.12, 5.4.13 e 5.4.14 ( $X_1, \dots, X_n$ ) sarà un campione di numerosità  $n$  di legge  $\mathcal{N}(\mu, \sigma^2)$ .

#### 5.4.7. Intervalli di fiducia per la media della normale con varianza

**nota.** Si parte dall'osservazione che

$$Y = \frac{\bar{X} - \mu}{\sigma} \sqrt{n} \sim \mathcal{N}(0, 1).$$

(a) *Intervallo bilaterale.* Dalla relazione

$$P^\mu(a \leq Y \leq b) = \Phi(b) - \Phi(a)$$

si ricava

$$\begin{aligned} & \Phi(b) - \Phi(a) \\ &= P^\mu\left(a \leq \frac{\bar{X} - \mu}{\sigma} \sqrt{n} \leq b\right) = P^\mu\left(\bar{X} - \frac{\sigma}{\sqrt{n}}b \leq \mu \leq \bar{X} - \frac{\sigma}{\sqrt{n}}a\right). \end{aligned}$$

Pertanto basterà trovare  $a$  e  $b$  in modo che

$$\Phi(b) - \Phi(a) = 1 - \alpha.$$

Siano  $\beta$  e  $\gamma$  due numeri reali  $\in [0, 1]$  tali che  $b = \phi_\beta$ ,  $a = \phi_\gamma$ ; allora si ha

$$\Phi(b) - \Phi(a) = \beta - \gamma.$$

Dobbiamo dunque scegliere  $\beta$  e  $\gamma$  in modo che risulti  $\beta - \gamma = 1 - \alpha$ . Una scelta possibile è  $\beta = 1 - \alpha/2$ ,  $\gamma = \alpha/2$ , (cioè  $b = \phi_{1-\alpha/2}$ ,  $a = \phi_{\alpha/2} = -\phi_{1-\alpha/2}$ , dove l'ultima uguaglianza segue dalla nota proprietà dei quantili della legge normale standard).

L'intervallo risultante è allora

$$\left(\bar{X} - \frac{\sigma}{\sqrt{n}}\phi_{1-\alpha/2}, \bar{X} + \frac{\sigma}{\sqrt{n}}\phi_{1-\alpha/2}\right). \quad (5.4.8)$$

**Osservazione 5.4.9.** La scelta fatta per  $\beta$  e  $\gamma$  non è ovviamente l'unica possibile: ad esempio  $\beta = 1 - \alpha/3$ ,  $\gamma = \alpha(2/3)$  va ancora bene. Tuttavia l'intervallo (5.4.8) è quello di ampiezza minima, (cioè dà la stima migliore possibile, al livello  $1 - \alpha$  assegnato); la dimostrazione è lasciata per esercizio al lettore.

(b) *Intervallo unilaterale destro.* Il termine significa che si vuole trovare una limitazione per  $\mu$  solo dal basso, cioè del tipo  $H < \mu$  (il termine “destro” si spiega osservando che in tal caso  $\mu \in (H, +\infty)$ , semiretta destra).

Questa volta partiamo dalla relazione

$$P^\mu(Y \leq b) = \Phi(b),$$

che equivale a

$$P^\mu\left(\bar{X} - \frac{\sigma}{\sqrt{n}}b \leq \mu\right) = \Phi(b).$$

Se  $b = \phi_\beta$ , allora basterà che

$$1 - \alpha = \Phi(b) = \beta,$$

ovvero, semplicemente,  $b = \phi_{1-\alpha}$ , e l'intervallo è

$$\left( \bar{X} - \frac{\sigma}{\sqrt{n}} \phi_{1-\alpha}, +\infty \right).$$

(c) *Intervallo unilaterale sinistro*. Non ripeteremo i calcoli, che sono analoghi ai precedenti (ma farli, per esercizio!). Si trova l'intervallo

$$\left( -\infty, \bar{X} - \frac{\sigma}{\sqrt{n}} \phi_\alpha \right) = \left( -\infty, \bar{X} + \frac{\sigma}{\sqrt{n}} \phi_{1-\alpha} \right),$$

ricordando la solita relazione  $-\phi_\alpha = \phi_{1-\alpha}$ .

**Osservazione 5.4.10.** La scelta del tipo di intervallo da considerare è in genere legata alla situazione pratica (se si deve avere una stima sia da destra che da sinistra calcoleremo un intervallo bilaterale, se invece occorre stimare il parametro solo dal basso cercheremo un intervallo unilaterale destro, e così via).

**Osservazione 5.4.11.** Volendo confrontare le stime unilaterali di  $\mu$  con la stima bilaterale, tutte di livello  $1 - \alpha$  fissato, si ha (per quanto riguarda per esempio la stima da sinistra)

$$\bar{X} - \frac{\sigma}{\sqrt{n}} \phi_{1-\alpha/2} \leq \bar{X} - \frac{\sigma}{\sqrt{n}} \phi_{1-\alpha}$$

(dimostrazione per esercizio). Trovare l'analogia relazione fra la stima bilaterale e quella da destra.

**5.4.12. Intervalli di fiducia per la media della normale con varianza non nota.** Nella pratica gli intervalli del paragrafo 5.4.7 sono di scarsa utilità, perché nelle formule che li definiscono interviene la varianza  $\sigma^2$ , che in genere non si conosce.

In questo caso si può sostituire  $\sigma^2$  con

$$S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}$$

(che ne è uno stimatore corretto), e applicare di nuovo il metodo della quantità pivotale partendo dalla v. a.

$$Z = \frac{\bar{X} - \mu}{S} \sqrt{n} \sim t(n-1),$$

(teorema di Cochran). Osservando che l'unica proprietà della legge normale standard che abbiamo usato nel paragrafo 5.4.7 è stata la simmetria (a proposito dei quantili, nella relazione  $-\phi_\alpha = \phi_{1-\alpha}$ ), e ricordando che anche la legge  $t$  di Student è simmetrica, risulta chiaro che tutto ciò che abbiamo detto nel paragrafo 5.4.7 si può ripetere, semplicemente sostituendo  $\sigma$  con  $S$  e i quantili della  $\mathcal{N}(0, 1)$  con quelli della  $t(n-1)$ . Per comodità di chi legge, riportiamo comunque le formule finali.

(a) *Intervallo bilaterale.*

$$\left( \bar{X} - \frac{S}{\sqrt{n}} t_{1-\frac{\alpha}{2}, n-1}, \bar{X} + \frac{S}{\sqrt{n}} t_{1-\frac{\alpha}{2}, n-1} \right).$$

(b) *Intervallo unilaterale destro.*

$$\left( \bar{X} - \frac{S}{\sqrt{n}} t_{1-\alpha, n-1}, +\infty \right).$$

(c) *Intervallo unilaterale sinistro.*

$$\left( -\infty, \bar{X} + \frac{S}{\sqrt{n}} t_{1-\alpha, n-1} \right).$$

**5.4.13. Intervalli di fiducia per la varianza della normale con media nota.** Qui si parte ricordando che la v. a.

$$Y = \frac{\sum_{i=1}^n (X_i - \mu)^2}{\sigma^2}$$

ha legge  $\chi^2(n)$  (Osservazione 4.3.16). Posto

$$U^2 = \frac{\sum_{i=1}^n (X_i - \mu)^2}{n},$$

(che, come sappiamo, è uno stimatore corretto della varianza) si può scrivere

$$Y = \frac{nU^2}{\sigma^2}.$$

(a) *Intervallo bilaterale.* Indichiamo con  $F_n$  la funzione di ripartizione della  $\chi^2(n)$ . Allora si ha

$$\begin{aligned} F_n(b) - F_n(a) &= P^{\sigma^2}(a \leq Y \leq b) \\ &= P^{\sigma^2}\left(a \leq \frac{nU^2}{\sigma^2} \leq b\right) = P^{\sigma^2}\left(\frac{nU^2}{b} \leq \sigma^2 \leq \frac{nU^2}{a}\right). \end{aligned}$$

Quindi, se al solito poniamo  $b = \chi_{\beta,n}^2$ ,  $a = \chi_{\gamma,n}^2$ , avremo

$$1 - \alpha = F_n(b) - F_n(a) = \beta - \gamma.$$

Una scelta possibile è  $\beta = 1 - \alpha/2$ ,  $\gamma = \alpha/2$ , e si ottiene l'intervallo

$$\left( \frac{nU^2}{\chi_{1-\frac{\alpha}{2},n}^2}, \frac{nU^2}{\chi_{\frac{\alpha}{2},n}^2} \right).$$

Gli intervalli unilaterali si ottengono in modo analogo. Riportiamo solo le formule finali.

(b) *Intervallo unilaterale destro.*

$$\left( \frac{nU^2}{\chi_{1-\alpha,n}^2}, +\infty \right).$$

(c) *Intervallo unilaterale sinistro.*

$$\left( 0, \frac{nU^2}{\chi_{\alpha,n}^2} \right).$$

**5.4.14. Intervalli di fiducia per la varianza della normale con media non nota.** Poiché normalmente la media non è nota, nella relazione del Paragrafo 5.4.13 che definisce  $U^2$  si può cercare di sostituire  $\mu$  con il suo stimatore  $\bar{X}$ , usando, al posto di  $U^2$ , la v. a.

$$S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}$$

(che è ancora uno stimatore corretto della varianza). Quindi applicheremo il metodo della quantità pivotale a partire dalla v. a.

$$Z = \frac{(n-1)S^2}{\sigma^2},$$

che, dal teorema di Cochran, sappiamo avere legge  $\chi^2(n-1)$ . Dunque, per avere i tre nuovi intervalli basterà sostituire nelle formule del paragrafo (5.4.13)  $n-1$  al posto di  $n$  (e i quantili della  $\chi^2(n-1)$  al posto di quelli della  $\chi^2(n)$ ). Si ottengono così le espressioni che seguono:

(a) *Intervallo bilaterale.*

$$\left( \frac{(n-1)S^2}{\chi_{1-\frac{\alpha}{2},n-1}^2}, \frac{(n-1)S^2}{\chi_{\frac{\alpha}{2},n-1}^2} \right).$$



(b) *Intervallo unilaterale destro.*

$$\left( \frac{(n-1)S^2}{\chi_{1-\alpha, n-1}^2}, +\infty \right).$$

(c) *Intervallo unilaterale sinistro.*

$$\left( 0, \frac{(n-1)S^2}{\chi_{\alpha, n-1}^2} \right).$$

**5.4.15. Intervalli di fiducia per il parametro della bernoulliana (per una proporzione).** Sia  $X_1, \dots, X_n$  un campione estratto dalla legge  $B(1, p)$ . Osservando che, per  $n$  abbastanza grande, la v.a.

$$Y = \frac{\bar{X} - p}{\sqrt{p(1-p)}} \sqrt{n}$$

è (approssimativamente!) di legge  $\mathcal{N}(0, 1)$  per il Teorema Limite Centrale (verifica per esercizio), si può pensare di ripetere i calcoli fatti per campioni gaussiani (cioè a proposito della stima della media della legge normale). Per la stima bilaterale, per esempio, si troverebbe l'intervallo

$$\left( \bar{X} - \frac{\sqrt{p(1-p)}}{\sqrt{n}} \phi_{1-\alpha/2}, \bar{X} + \frac{\sqrt{p(1-p)}}{\sqrt{n}} \phi_{1-\alpha/2} \right),$$

in cui, semplicemente, abbiamo sostituito  $\sqrt{p(1-p)}$  al posto di  $\sigma$  (o  $S$ ).

Purtroppo però l'intervallo risultante dipende dal parametro  $p$ , che è appunto la quantità incognita da stimare, e dunque non è un buon intervallo di fiducia (rivedere la definizione).

L'idea è allora quella di sostituire  $p$  con il suo stimatore  $\bar{X}$ . La cosa può essere giustificata rigorosamente (ma noi non lo faremo). Riportiamo soltanto le formule finali, ancora per comodità.

(a) *Intervallo bilaterale.*

$$\left( \bar{X} - \frac{\sqrt{\bar{X}(1-\bar{X})}}{\sqrt{n}} \phi_{1-\alpha/2}, \bar{X} + \frac{\sqrt{\bar{X}(1-\bar{X})}}{\sqrt{n}} \phi_{1-\alpha/2} \right).$$

(b) *Intervallo unilaterale destro.*

$$\left( \bar{X} - \frac{\sqrt{\bar{X}(1-\bar{X})}}{\sqrt{n}} \phi_{1-\alpha}(n-1), 1 \right).$$

(c) *Intervallo unilaterale sinistro.*

$$\left( 0, \bar{X} + \frac{\sqrt{\bar{X}(1-\bar{X})}}{\sqrt{n}} \phi_{1-\alpha} \right).$$

## 5.5 Test statistici (parametrici)

*Motivazione.* Un problema tipico della statistica è quello di stabilire se il parametro incognito  $\theta \in \Theta$  è “di un certo tipo” oppure no. Ovvero, per motivi legati alla situazione pratica da esaminare, l’insieme dei parametri  $\Theta$  viene suddiviso in due sottoinsiemi  $\Theta_0$  e  $\Theta_1$ , con  $\Theta_0 \cup \Theta_1 = \Theta$ ,  $\Theta_0 \cap \Theta_1 = \emptyset$  (una partizione di  $\Theta$ , dunque); si deve poi decidere se  $\theta \in \Theta_0$  oppure  $\theta \in \Theta_1$ . In seguito alla decisione presa, poi si opera in un modo oppure nel modo contrario. Vediamo qualche esempio per chiarire meglio.

**Esempio 5.5.1 (sperimentazione farmacologica).** Si sperimenta un nuovo farmaco contro il colesterolo. È noto che, per gli individui sani, il tasso di colesterolo nel sangue è una v. a. di legge  $\mathcal{N}(\mu_0, \sigma_0^2)$  ( $\mu_0$  e  $\sigma_0^2$  sono quantità note). Indichiamo con  $X$  il tasso di colesterolo rilevato nel sangue di un generico individuo dopo la somministrazione del farmaco. Ipotizziamo che  $X$  sia una v. a. avente legge  $\mathcal{N}(\mu, \sigma_0^2)$  dove  $\mu$  adesso non può più essere considerata nota (per semplicità supporremo invece che  $\sigma_0^2$  non sia cambiata rispetto all’individuo sano).

In questo caso il parametro da valutare è ovviamente  $\mu$ . Sarà ragionevole dire che il farmaco è risultato efficace se  $\mu \leq \mu_0$ . Cioè: in generale sappiamo che  $\mu \in (0, +\infty) := \Theta$ , e ci interessa discriminare fra le due alternative  $\mu \in (0, \mu_0]$  e  $\mu \in (\mu_0, +\infty)$ .

Se decideremo che  $\mu \leq \mu_0$ , ciò significherà che abbiamo stabilito di considerare il farmaco come efficace, e di conseguenza lo metteremo in commercio. Non lo faremo invece se avremo preso la decisione contraria.

**Osservazione 5.5.2.** Nella teoria dei test alcune notazioni sono ormai standard; per il momento non indichiamo i due sottoinsiemi  $(0, \mu_0]$  e  $(\mu_0, +\infty)$  con i simboli  $\Theta_0$  e  $\Theta_1$  usati all’inizio, perchè tali simboli avranno in seguito un significato particolare, che in questo momento non abbiamo ancora precisato. Questa osservazione vale anche per l’esempio che segue.

**Esempio 5.5.3 (controllo di qualità).** Una fabbrica vuole controllare la qualità dei pezzi prodotti. Secondo il suo standard di qualità, la proporzione di pezzi difettosi non deve superare un valore fissato  $p_0$  (ovviamente noto). Poniamo

$$X = \begin{cases} 1 & \text{se il pezzo generico è difettoso,} \\ 0 & \text{in caso contrario.} \end{cases}$$

Consideriamo  $X$  come una v. a. avente legge  $B(1, p)$ , dove  $p$  non è nota. Sarà ragionevole dire che lo standard di qualità è rispettato se  $p \leq p_0$ . In questo caso, dunque, si ha  $p \in (0, 1) = \Theta$ , e vogliamo discriminare fra  $p \in (0, p_0]$  e  $p \in (p_0, 1)$ .

Supponiamo che la ditta decida che  $p \in (p_0, 1)$ , cioè che lo standard non è rispettato. Questa decisione indurrà a cambiare qualcosa nel processo di produzione. Niente verrà cambiato invece se la decisione sarà quella contraria.

*Generalità.* Vediamo allora come si procede in generale. Quello che viene in mente di fare è procurarsi delle osservazioni  $X_1, \dots, X_n$ , una statistica  $T = t(X_1, \dots, X_n)$  opportuna e decidere, in base al valore assunto da questa, tra le due possibilità in esame. Di esse, una viene chiamata *ipotesi* (o *ipotesi nulla* nei test a maggior connotazione applicativa), e indicata con  $H_0$ , l'altra *alternativa* (o *ipotesi alternativa*), e indicata con  $H_1$  (chiariremo fra poco il motivo di questi simboli). Nell'Esempio 5.5.1 potrebbe essere  $H_0 : \mu \leq \mu_0$ ;  $H_1 : \mu > \mu_0$ , oppure al contrario. Analogamente, nell'Esempio 5.5.3 potremmo scrivere  $H_0 : p \leq p_0$ ;  $H_1 : p > p_0$ , oppure il contrario.

Va però osservato fin da ora che la scelta (tra quale possibilità indicare con  $H_0$  e quale con  $H_1$ ) non è arbitraria, per motivi che vedremo in seguito.

Riprendiamo le nostre considerazioni sulla statistica  $T$ . Ci saranno dei valori appartenenti ad un sottoinsieme  $D \subset \mathbb{R}$  che  $T$  può assumere e che ci faranno propendere per  $H_1$ , altri invece ci faranno preferire  $H_0$ . Supponiamo che il sottoinsieme di  $\Omega$   $\{T \in D\}$  sia un evento, così da poterne calcolare ogni probabilità del tipo  $P^\theta(T \in D)$ ,  $\theta \in \Theta$ .

**Definizione 5.5.4.** Si chiama *regione critica* (o *di rigetto*) del test l'evento  $\{T \in D\}$ , dove  $D$  è l'insieme dei valori che ci inducono a respingere l'ipotesi nulla  $H_0$ . La *regione di accettazione* del test è l'evento complementare  $\{T \in D^c\}$ .

**Osservazione 5.5.5.** In generale, l'insieme  $D$  è di qualche tipo particolare, suggerito dalla situazione (ad esempio una semiretta destra della retta, oppure un intervallo). I test che studieremo chiariranno come ci si orienta nella scelta. (In realtà c'è una teoria matematica, dovuta agli statistici J. Neyman e E. Pearson, che indica come cercare le "buone" regioni critiche, ma la sua trattazione va al di là degli scopi di questi appunti).

La regione critica decide evidentemente il risultato del test (e, come abbiamo visto, a seconda che si accetti o si respinga  $H_0$  si procederà in un modo oppure nel modo contrario). Occorre dunque dare un criterio "prudente" per determinarla; ciò che diremo metterà in luce il fatto che, diversamente da quanto può sembrare, *non* c'è simmetria nella scelta di  $H_0$  e  $H_1$ .

Tra le due possibilità in esame, in genere ce ne è una che reputiamo sfavorevole, (cioè che, se respinta nel caso che sia vera, comporta maggiori danni). Questa è assunta come  $H_0$ , e lo scopo del test è quello di rifiutarla (se possibile, naturalmente). Indicheremo allora con  $\Theta_0$  il sottoinsieme di valori del parametro corrispondenti a  $H_0$ .

Ripensiamo, per chiarire, all'Esempio 5.5.1: stabilire che un farmaco è efficace, quando non è vero, comporta come conseguenza dei rischi per la salute di chi se ne servirà, perchè avremo messo in commercio un farmaco inutile. In questo caso, dunque, lo sperimentatore agirà in modo corretto (dal punto di vista etico, almeno!) se stabilirà che  $H_0 : \mu > \mu_0$  (e dunque  $\Theta_0 = \{\mu > \mu_0\}$ ). In sostanza, un buon sperimentatore *deve* essere pessimista!

**Osservazione 5.5.6.** Si capisce da quello che abbiamo detto sopra che la scelta di  $H_0$  è comunque in qualche modo arbitraria, e legata alle convinzioni dello sperimentatore: per esercizio, provare a descrivere le conseguenze della scelta contraria, cioè  $H_0 : \mu \leq \mu_0$  nell'esempio del farmaco. Quale tipo di sperimentatore dovremmo ipotizzare in questo secondo caso?

**Esempio 5.5.7.** Discutere anche l'esempio del controllo di qualità:

(a) quali conseguenze comporta respingere a torto l'ipotesi  $p \leq p_0$ ? E quali respingere a torto  $p > p_0$ ?

(b) quale tipo di sperimentatore sceglierebbe come ipotesi nulla  $p \leq p_0$ ? E quale  $p > p_0$ ?

**Esempio 5.5.8.** Ancora a proposito dell'Osservazione 5.5.6, discutere l'esempio di un test sul DNA eseguito nello stato americano del Texas per decidere se un imputato è o no colpevole di omicidio (nel Texas per questo tipo di reato è prevista la pena di morte): se l'ipotesi  $H_0$  corrisponde all'affermazione “l'imputato è innocente”, quale conseguenza ha il fatto di accettare l'ipotesi se essa è falsa? e quale il fatto di respingerla se è vera? in quale caso, secondo il lettore, si produce il danno più grave?

**Osservazione 5.5.9.** Il simbolo usato  $H_0$  e il termine degli statistici *ipotesi nulla* si spiegano pensando che un'ipotesi a noi sfavorevole *vanifica*, *rende nulle* le nostre speranze (ripensare ancora all'esempio del farmaco).

La discussione fatta avrà chiarito che nell'eseguire un test si possono commettere due tipi di errore:

(i) *Errore di prima specie*: consiste nel respingere a torto  $H_0$  (cioè quando in realtà essa è vera); questa situazione si presenta quando  $\theta \in \Theta_0$  ma  $T \in D$ .

(ii) *Errore di seconda specie*: consiste nell'accettare a torto  $H_0$  (cioè quando in realtà essa è falsa) (si potrebbe dire anche: “respingere a torto  $H_1$ ”, ma si preferisce quasi sempre esprimersi in termini dell'ipotesi anziché dell'alternativa); questo accade quando  $\theta \in \Theta_1$  ma  $T \in D^c$ .

Dalle considerazioni svolte sopra, risulta anche che i due tipi di errore non possono essere considerati della stessa gravità: l'errore di prima specie è più grave di quello di seconda.

Tutti e due i tipi di errore possono verificarsi, ed è importante calcolarne le probabilità; da quanto abbiamo detto sopra si ricava che la probabilità di un errore di prima specie è data da  $P^\theta(T \in D)$ ,  $\theta \in \Theta_0$ , mentre la probabilità di errore di seconda specie è  $P^\theta(T \in D^c) = 1 - P^\theta(T \in D)$ ,  $\theta \in \Theta_1$ .

Poichè, come abbiamo detto, l'errore di primo tipo è considerato più grave, lo sperimentatore dovrà cautelarsi contro di esso, imponendo che la probabilità che esso si verifichi non sia troppo alta; in altre parole egli fisserà un valore

$\alpha \in (0, 1)$  non troppo alto (valori tipici per  $\alpha$  sono 0.05, 0.01 ecc.) ed imporrà che risulti

$$\alpha^* := \sup_{\theta \in \Theta_0} P^\theta(T \in D) \leq \alpha. \quad (5.5.10)$$

La relazione (5.5.10) significa che, qualunque sia il valore di  $\theta \in \Theta_0$ , la probabilità di errore di prima specie non supera il valore assegnato  $\alpha$ .

**Definizione 5.5.11.** La quantità  $\alpha^*$  definita in (5.5.10) si chiama *livello* del test. Più genericamente, assegnato  $\alpha \in (0, 1)$ , si dice che il test è *di livello*  $\alpha$  se vale la (5.5.10).

Nella (5.5.10) l'incognita del problema è  $D$ ; dunque tale relazione determina  $D$  (in funzione di  $\alpha$ ). I test del paragrafo seguente chiariranno meglio questa affermazione.

Come abbiamo visto, l'errore di seconda specie è  $1 - P^\theta(T \in D)$ ,  $\theta \in \Theta_1$ ; dunque, per  $\theta \in \Theta_1$ , la quantità  $P^\theta(T \in D)$  rappresenta la probabilità di prendere la decisione corretta (di respingere a ragione  $H_0$ ). Questo motiva la seguente

**Definizione 5.5.12.** La funzione  $\theta \in \Theta_1 \mapsto P^\theta(T \in D)$  si chiama *potenza del test*.

Per un assegnato test, una volta fissato il livello (cioè una volta determinato l'insieme  $D$  in funzione di  $\alpha$ ), la funzione potenza è determinata. Può dunque capitare che per alcuni valori di  $\theta \in \Theta_1$  la potenza sia bassa, il che non è soddisfacente (per esempio, se per  $\theta_0 \in \Theta_1$  si trovasse  $P^{\theta_0}(T \in D) = 0.25$ , questo significherebbe che la probabilità di respingere  $H_0$  nel caso in cui  $\theta = \theta_0$ , cioè con  $H_0$  falsa, è solo del 25%, il che è veramente poco!). Una situazione del genere si può verificare ad esempio per valori di  $n$  troppo piccoli (abbiamo fatto poche osservazioni), oppure per valori di  $\alpha$  troppo bassi (pretendiamo un livello del test troppo basso).

**Osservazione 5.5.13.** Aumentando  $n$  oppure  $\alpha$  si può migliorare la potenza del test; questo naturalmente ha un costo pratico: aumentare  $\alpha$  significa infatti rischiare di più (probabilità di errore di prima specie più alta), mentre aumentare il numero di osservazioni può essere dispendioso dal punto di vista economico, o magari proprio impossibile, per motivi tecnici.

*Nota.* In tutti i test che considereremo nei successivi paragrafi, si intende fissato il valore  $\alpha$ .

## 5.6 Test per campioni gaussiani

In questo paragrafo considereremo il caso di osservazioni  $(X_1, \dots, X_n)$  che costituiscano un campione di legge normale.

**5.6.1. Test per la media di una normale con varianza nota (Test di Student).** Supponiamo che la legge delle  $(X_i)$  sia  $\mathcal{N}(\mu, \sigma_0^2)$ , dove  $\sigma_0^2$  è una quantità nota. Lo scopo dei test che vedremo è quello di confrontare  $\mu$  (che non conosciamo) con un valore noto  $\mu_0$ .

Useremo sempre la statistica

$$T = \frac{\bar{X} - \mu_0}{\sigma_0} \sqrt{n}. \quad (5.6.2)$$

(a) *Test bilaterale.* Si tratta del test

$$H_0 : \mu = \mu_0; \quad H_1 : \mu \neq \mu_0.$$

Cerchiamo anzitutto di capire di che tipo dovrà essere l'insieme dei valori critici  $D$ . Ricordando che, per la Legge dei Grandi Numeri,  $\bar{X} \simeq \mu$  quando  $n$  è grande e guardando la (5.6.2), è ragionevole supporre che, se  $\mu \neq \mu_0$ , la statistica  $T$  tenda ad assumere valori di modulo “grande” (il segno di  $T$  dipende da quello di  $\mu - \mu_0$ , e dunque non possiamo determinarlo a partire dalla sola condizione  $\mu \neq \mu_0$ ): in altre parole cerchiamo una regione critica del tipo  $\{|T| > z\}$ , con  $z$  valore da determinare (in funzione di  $\alpha$ , come abbiamo detto in precedenza, vedere quanto detto dopo la Definizione 5.5.11).

Sia  $\alpha^*$  il livello del test. Poichè in questo caso si ha  $\Theta_0 = \{\mu_0\}$ , dalla formula 5.5.10 si ricava

$$\alpha^* = \sup_{\mu=\mu_0} P^\mu(|T| > z) = P^{\mu_0}(|T| > z).$$

D'altra parte, se  $\mu = \mu_0$ , si ha  $T \sim \mathcal{N}(0, 1)$ , e dunque il test sarà di livello  $\alpha$  se

$$\begin{aligned} \alpha &\geq \alpha^* = P^{\mu_0}(|T| > z) = P^{\mu_0}(T > z) + P^{\mu_0}(T < -z) \\ &= 1 - \Phi(z) + \Phi(-z) = 2(1 - \Phi(z)). \end{aligned}$$

Risolvendo per semplicità l'uguaglianza  $2(1 - \Phi(z)) = \alpha$  (anziché la disuguaglianza) rispetto all'incognita  $z$ , si trova (eseguire i calcoli per esercizio)

$$z = \phi_{1-\alpha/2}.$$

La regione critica sarà dunque

$$\{|T| > \phi_{1-\alpha/2}\}.$$

Cerchiamo ora la potenza del test. Si tratta di calcolare il valore della funzione

$$\Pi : \mu \mapsto P^\mu(|T| > \phi_{1-\alpha/2}), \quad \mu \neq \mu_0.$$

Prima di tutto si può scrivere

$$P^\mu(|T| > \phi_{1-\alpha/2}) = P^\mu(T > \phi_{1-\alpha/2}) + P^\mu(T < -\phi_{1-\alpha/2}).$$

Per il primo addendo, si ha

$$\begin{aligned} P^\mu(T > \phi_{1-\alpha/2}) &= P^\mu\left(\frac{\bar{X} - \mu_0}{\sigma_0}\sqrt{n} > \phi_{1-\alpha/2}\right) \\ &= P^\mu\left(\frac{\bar{X} - \mu}{\sigma_0}\sqrt{n} > \phi_{1-\alpha/2} - \frac{\mu - \mu_0}{\sigma_0}\sqrt{n}\right) \\ &= 1 - \Phi\left(\phi_{1-\alpha/2} - \frac{\mu - \mu_0}{\sigma_0}\sqrt{n}\right), \end{aligned}$$

dove l'ultima uguaglianza segue dal fatto che  $\mu$  (non  $\mu_0$ !) è la media di  $X$  e quindi

$$\frac{\bar{X} - \mu}{\sigma_0}\sqrt{n} \sim \mathcal{N}(0, 1).$$

In modo analogo si trova

$$\begin{aligned} P^\mu(T < -\phi_{1-\alpha/2}) &= P^\mu\left(\frac{\bar{X} - \mu}{\sigma_0}\sqrt{n} < -\phi_{1-\alpha/2} - \frac{\mu - \mu_0}{\sigma_0}\sqrt{n}\right) \\ &= \Phi\left(-\phi_{1-\alpha/2} - \frac{\mu - \mu_0}{\sigma_0}\sqrt{n}\right), \end{aligned}$$

e infine

$$\Pi(\mu) = 1 - \Phi\left(\phi_{1-\alpha/2} - \frac{\mu - \mu_0}{\sigma_0}\sqrt{n}\right) + \Phi\left(-\phi_{1-\alpha/2} - \frac{\mu - \mu_0}{\sigma_0}\sqrt{n}\right).$$

(b) *Test unilaterale*  $H_0 : \mu \leq \mu_0$ ;  $H_1 : \mu > \mu_0$ . Ragionando come nel caso (a), decidiamo che, per  $\mu > \mu_0$  e  $n$  grande  $T$  tenderà ad assumere valori grandi e positivi, e quindi cerchiamo una regione critica del tipo

$$\{T > z\},$$

e al solito determineremo  $z$  in funzione di  $\alpha$ . Il livello del test è in questo caso

$$\alpha^* = \sup_{\mu \leq \mu_0} P^\mu(T > z), \quad (5.6.3)$$

e, volendo calcolare esplicitamente il secondo membro della relazione precedente, abbiamo bisogno di conoscere la legge di  $T$ . Purtroppo la statistica  $T$  segue una legge nota (precisamente la  $\mathcal{N}(0, 1)$ ) solo se  $\mu = \mu_0$ ; mentre noi dobbiamo fare il calcolo per ogni  $\mu \leq \mu_0$ . La teoria matematica (dei test a

rapporto di verosimiglianza monotona) ci direbbe come fare, ma non è alla nostra portata; comunque possiamo cavarcela come segue: osserviamo che, se  $\mu \leq \mu_0$ , si ha

$$\frac{\bar{X} - \mu}{\sigma_0} \sqrt{n} \geq \frac{\bar{X} - \mu_0}{\sigma_0} \sqrt{n} = T,$$

e quindi si ha la seguente inclusione tra eventi:

$$\{T > z\} = \left\{ \frac{\bar{X} - \mu_0}{\sigma_0} \sqrt{n} > z \right\} \subseteq \left\{ \frac{\bar{X} - \mu}{\sigma_0} \sqrt{n} > z \right\},$$

da cui, per l'isotonia di  $P^\mu$ , si deduce

$$P^\mu(T > z) = P^\mu\left(\frac{\bar{X} - \mu_0}{\sigma_0} \sqrt{n} > z\right) \leq P^\mu\left(\frac{\bar{X} - \mu}{\sigma_0} \sqrt{n} > z\right) = 1 - \Phi(z), \quad (5.6.4)$$

dove l'ultima uguaglianza segue dal fatto che, essendo  $\mu$  il *vero* valore della media delle  $X_i$ , la v. a.

$$\frac{\bar{X} - \mu}{\sigma_0} \sqrt{n}$$

ha legge  $\mathcal{N}(0, 1)$ , come abbiamo detto sopra.

Dalle (5.6.3) e (5.6.4) si conclude allora che

$$\alpha^* \leq 1 - \Phi(z).$$

Dunque, per ottenere un test di livello  $\alpha$ , basterà che sia  $1 - \Phi(z) = \alpha$ , e cioè  $z = \phi_{1-\alpha}$ . La regione critica del test è allora

$$\{T > \phi_{1-\alpha}\}.$$

Per la formula della potenza, si ragiona esattamente come in (a) e si trova per  $\mu > \mu_0$ ,

$$\begin{aligned} \Pi(\mu) &= P^\mu(T > \phi_{1-\alpha}) = P^\mu\left(\frac{\bar{X} - \mu}{\sigma_0} \sqrt{n} > \phi_{1-\alpha} - \frac{\mu - \mu_0}{\sigma_0} \sqrt{n}\right) \\ &= 1 - \Phi\left(\phi_{1-\alpha} - \frac{\mu - \mu_0}{\sigma_0} \sqrt{n}\right). \end{aligned}$$

(c) *Test unilaterale*  $H_0 : \mu \geq \mu_0$ ;  $H_1 : \mu < \mu_0$ . I ragionamenti e i calcoli sono analoghi a quelli dell'altro test unilaterale. Si ottiene la regione critica

$$\{T < \phi_\alpha\}.$$

La potenza è

$$\begin{aligned} \Pi(\mu) &= P^\mu(T < \phi_\alpha) = P^\mu\left(\frac{\bar{X} - \mu}{\sigma_0} \sqrt{n} < \phi_\alpha - \frac{\mu - \mu_0}{\sigma_0} \sqrt{n}\right) \\ &= \Phi\left(\phi_\alpha - \frac{\mu - \mu_0}{\sigma_0} \sqrt{n}\right). \end{aligned}$$



**5.6.5. Test per la media di una normale con varianza non nota (Test di Student).** Quando la varianza non è nota (è il caso più frequente, naturalmente), si può mettere  $S^2$  al posto di  $\sigma^2$ : in questo caso cioè la statistica usata è

$$T = \frac{\bar{X} - \mu_0}{S} \sqrt{n};$$

ripetendo esattamente le considerazioni fatte a proposito degli intervalli di fiducia con varianza non nota (5.4.12), si trovano le seguenti formule (in cui ovviamente i quantili della  $t(n-1)$  hanno sostituito quelli della normale standard).

(a) *Test bilaterale*  $H_0 : \mu = \mu_0$ ;  $H_1 : \mu \neq \mu_0$ .

Regione critica

$$\{|T| > t_{1-\frac{\alpha}{2}, n-1}\}.$$

(b) *Test unilaterale*  $H_0 : \mu \leq \mu_0$ ;  $H_1 : \mu > \mu_0$ .

Regione critica

$$\{T > t_{1-\alpha, n-1}\}.$$

(c) *Test unilaterale*  $H_0 : \mu \geq \mu_0$ ;  $H_1 : \mu < \mu_0$ .

Regione critica

$$\{T < t_{\alpha, n-1}\}.$$

Tralasciamo di scrivere le formule della potenza, perchè troppo complicate da giustificare.

**Osservazione 5.6.6.** Un caso particolare del test di Student si ha quando l'esperimento consiste nell'effettuare  $n$  misurazioni di una certa quantità prima e dopo un dato intervento dello sperimentatore (le generiche misurazioni prima e dopo l'intervento sono due v. a., indicate con  $U$  e  $V$  rispettivamente); si può pensare per esempio al caso della sperimentazione di un farmaco contro il diabete, in cui a  $n$  individui viene misurato il tasso di glicemia prima e dopo la somministrazione del farmaco in questione. È importante saper confrontare le due medie  $\mu_U$  e  $\mu_V$ : ancora nel caso del farmaco contro il diabete, domandarsi ad esempio se  $\mu_U > \mu_V$  equivale a chiedersi se il farmaco è risultato efficace. Basta allora considerare la v.a.  $X = U - V$  (differenza tra la generiche misurazioni prima e dopo, avente media  $\mu_X = \mu_U - \mu_V$ ) ed eseguire su  $X$  il test di Student con  $\mu_0 = 0$ . Si suppone naturalmente che  $X$  sia una v. a. di legge normale e si utilizzano le osservazioni  $X_i = U_i - V_i$ , dove  $(U_i)_{i=1, \dots, n}$  (risp.  $(V_i)_{i=1, \dots, n}$ ) sono i valori della quantità in osservazione misurati prima (risp. dopo) l'intervento (nell'esempio sono gli  $n$  valori della glicemia rispettivamente prima e dopo la somministrazione).

Il test qui descritto è noto sotto il nome di test di *confronto tra due medie per campioni accoppiati* (*paired comparison* in inglese). Si usa in generale in situazioni, come nell'esempio, in cui si misura una certa quantità

prima e dopo un particolare trattamento (un farmaco, una qualità di concime per piante, un tipo di mangime per animali, ecc...) di cui si vogliono studiare gli effetti, e le misurazioni sono effettuate *sugli stessi individui* sia prima che dopo. In particolare, i due campioni di dati raccolti hanno la stessa numerosità.

### 5.6.7. Test di confronto tra due medie per campioni indipendenti.

Diverso dal precedente è il caso in cui i due campioni di dati provengono da due popolazioni differenti; in particolare adesso può succedere che le loro taglie siano differenti. Una situazione di questo tipo si ha per esempio quando si deve decidere se due diversi metodi (diciamoli  $A$  e  $B$ ) di produzione di un certo manufatto danno lo stesso standard di qualità; tale standard viene valutato misurando una certa quantità caratteristica del manufatto e denotata con  $X$  (risp.  $Y$ ) per i manufatti prodotti con il metodo  $A$  (risp.  $B$ ). Si suppone che  $X$  e  $Y$  siano due v. a. aventi legge normale di medie  $\mu_X$  e  $\mu_Y$  rispettivamente, e il confronto delle produzioni  $A$  e  $B$  si traduce nel confronto tra  $\mu_X$  e  $\mu_Y$ . A questo scopo, ci si procurano un primo campione  $X_1, \dots, X_n$  di misurazioni ottenute da  $n$  manufatti prodotti con il metodo  $A$  e un secondo campione  $Y_1, \dots, Y_m$  di misurazioni ottenute da  $m$  manufatti prodotti con il metodo  $B$ . Supporremo che le v.a.  $X_1, \dots, X_n, Y_1, \dots, Y_m$  siano indipendenti (come si interpreta questa ipotesi?). Vediamo come si opera nel caso (poco realistico in verità, per i motivi esposti in altre occasioni) in cui le varianze  $\sigma_X^2$  e  $\sigma_Y^2$  siano note.

(a) *Test bilaterale di confronto fra due medie, varianze note.* L'ipotesi è  $H_0 : \mu_X = \mu_Y$ , l'alternativa  $H_1 : \mu_X \neq \mu_Y$ .

Si utilizza la statistica

$$T = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}}.$$

Ragioniamo come al solito per decidere il tipo di regione critica. Se è vera  $H_1$ , allora  $\bar{X} - \bar{Y} \sim \mu_X - \mu_Y \neq 0$ , e, se  $m$  e  $n$  sono abbastanza grandi, la quantità  $T$  è grande in valore assoluto. Dunque pensiamo ad una regione critica del tipo  $\{|T| > k\}$ , con  $k$  da determinare in funzione del livello  $\alpha$ , osservando che, sotto  $H_0$ , la v. a.  $T$  ha legge  $\mathcal{N}(0, 1)$  (dettagli per esercizio) e quindi

$$\alpha^* = \sup_{\{\mu_X = \mu_Y\}} P^{\{\mu_X = \mu_Y\}}(|T| > k) = 2(1 - \Phi(k)) \leq \alpha,$$

da cui, come al solito, si ricava  $k = \phi_{1-\alpha/2}$ .

**Esempio 5.6.8.** Svolgere le considerazioni, analoghe alle precedenti, per impostare i due test unilaterali di confronto fra due medie, nel caso di varianze note.

(b) *Test bilaterale di confronto fra due medie, varianze non note ma uguali.* Nel caso, più realistico, che le varianze non siano note, e ragionando come abbiamo sempre fatto, si può pensare di utilizzare la statistica

$$T = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{S_X^2}{n} + \frac{S_Y^2}{m}}},$$

dove le quantità non note  $\sigma_X^2$  e  $\sigma_Y^2$  sono state sostituite dai rispettivi stimatori

$$S_X^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}, \quad S_Y^2 = \frac{\sum_{i=1}^m (Y_i - \bar{Y})^2}{m-1}.$$

Tuttavia il resto della discussione poggia sulla conoscenza della legge di  $T$  sotto l'ipotesi  $H_0$ , e sul fatto che tale legge non dipenda da quantità incognite (in quale momento della trattazione si ha necessità di questi fatti?). Purtroppo questa legge in realtà dipende da  $\sigma_X^2$  e  $\sigma_Y^2$ , non note, ed è oltretutto molto complicata. L'unico caso che possiamo trattare con i nostri strumenti è quello in cui si sa almeno che  $\sigma_X^2 = \sigma_Y^2 = \sigma^2$  (non note ma uguali). Infatti, sotto  $H_0$  (cioè sapendo che  $\mu_X = \mu_Y$ ), la v. a.  $\bar{X} - \bar{Y}$  ha legge  $\mathcal{N}(0, \frac{\sigma^2}{n} + \frac{\sigma^2}{m})$ , e quindi, standardizzando,

$$\frac{\bar{X} - \bar{Y}}{\sigma \sqrt{\frac{1}{n} + \frac{1}{m}}} \sim \mathcal{N}(0, 1). \quad (5.6.9)$$

Inoltre, dal teorema di Cochran sappiamo che

$$\begin{aligned} \frac{(n-1)S_X^2}{\sigma^2} &\sim \chi^2(n-1) = \Gamma\left(\frac{n-1}{2}, \frac{1}{2}\right), \\ \frac{(m-1)S_Y^2}{\sigma^2} &\sim \chi^2(m-1) = \Gamma\left(\frac{m-1}{2}, \frac{1}{2}\right); \end{aligned}$$

queste due v. a. sono indipendenti (la prima è funzione delle sole  $X_i$ , la seconda delle  $Y_j$ ), e, ricordando il teorema sulla somma di due v. a. Gamma indipendenti, si ottiene che

$$\frac{(n-1)S_X^2}{\sigma^2} + \frac{(m-1)S_Y^2}{\sigma^2} \sim \Gamma\left(\frac{m+n-2}{2}, \frac{1}{2}\right) = \chi^2(m+n-2). \quad (5.6.10)$$

Infine,  $S_X^2$  è indipendente da  $\bar{X} - \bar{Y}$  (naturalmente è indipendente da  $\bar{Y}$ , e, per il teorema di Cochran è indipendente anche da  $\bar{X}$ ). Analogamente,  $S_Y^2$  è indipendente da  $\bar{X} - \bar{Y}$ , e quindi anche  $\frac{(n-1)S_X^2}{\sigma^2} + \frac{(m-1)S_Y^2}{\sigma^2}$  lo è. Per le (5.6.9), (5.6.10) e la definizione di legge di Student, si conclude allora che la v. a.

$$\begin{aligned} U &= \sqrt{n+m-2} \cdot \frac{\bar{X} - \bar{Y}}{\sigma \sqrt{\frac{1}{n} + \frac{1}{m}}} : \sqrt{\frac{(n-1)S_X^2}{\sigma^2} + \frac{(m-1)S_Y^2}{\sigma^2}} \\ &= \frac{\bar{X} - \bar{Y}}{\sqrt{\left[\frac{1}{n} + \frac{1}{m}\right] [(n-1)S_X^2 + (m-1)S_Y^2]}} \cdot \sqrt{n+m-2} \sim t(n+m-2), \end{aligned}$$

sotto l'ipotesi  $H_0$ . Utilizzando allora la statistica  $U$ , con i ragionamenti abituali (svolgerli!) si trova che la regione critica di livello  $\alpha$  del nostro test è data da

$$\{|U| > t_{1-\frac{\alpha}{2}, n+m-2}\}.$$

Per i test unilaterali, i ragionamenti sono simili (dettagli per esercizio).

**5.6.11. Test per la varianza di una normale, media nota (Test di Fisher-Snedecor).** Qui facciamo l'ipotesi che le osservazioni  $X_1, \dots, X_n$  seguano una legge  $\mathcal{N}(\mu_0, \sigma^2)$ , dove  $\mu_0$  è una quantità nota.  $\sigma^2$  non è nota, e lo scopo dei test che seguiranno è quello di confrontarla con un valore assegnato  $\sigma_0^2$ .

Si utilizzerà la statistica

$$T = \frac{\sum_{i=1}^n (X_i - \mu_0)^2}{\sigma_0^2} = \frac{U^2}{\sigma_0^2} n,$$

dove si è posto

$$U^2 = \frac{\sum_{i=1}^n (X_i - \mu_0)^2}{n}.$$

(a) *Test bilaterale.* È il test

$$H_0 : \sigma^2 = \sigma_0^2, H_1 : \sigma^2 \neq \sigma_0^2.$$

Facciamo le solite considerazioni per capire come è fatta la regione critica. Osserviamo che, per la Legge dei Grandi Numeri, si ha

$$U^2 \rightarrow \sigma^2, \quad n \rightarrow \infty.$$

Dunque sotto  $H_0$  la quantità  $T = nU^2/\sigma_0^2$  prenderà valori maggiori (risp. minori) di  $n$  quando  $\sigma^2 > \sigma_0^2$  (risp.  $\sigma^2 < \sigma_0^2$ ). In altre parole ci aspettiamo una regione critica del tipo

$$\begin{aligned} & \{T < z_1\} \cup \{T > z_2\} \\ &= \left\{ \frac{\sum_{i=1}^n (X_i - \mu_0)^2}{\sigma_0^2} < z_1 \right\} \cup \left\{ \frac{\sum_{i=1}^n (X_i - \mu_0)^2}{\sigma_0^2} > z_2 \right\}, \end{aligned}$$

con  $z_1 < z_2$ , in cui  $z_1$  e  $z_2$  sono da determinare (in funzione di  $\alpha$ ). Calcoliamo il livello del test:

$$\begin{aligned} \alpha^* &= \sup_{\sigma^2 = \sigma_0^2} \left\{ P^{\sigma^2} \left( \frac{\sum_{i=1}^n (X_i - \mu_0)^2}{\sigma_0^2} < z_1 \right) + P^{\sigma^2} \left( \frac{\sum_{i=1}^n (X_i - \mu_0)^2}{\sigma_0^2} > z_2 \right) \right\} \\ &= P^{\sigma_0^2} \left( \frac{\sum_{i=1}^n (X_i - \mu_0)^2}{\sigma_0^2} < z_1 \right) + P^{\sigma_0^2} \left( \frac{\sum_{i=1}^n (X_i - \mu_0)^2}{\sigma_0^2} > z_2 \right) \\ &= F_n(z_1) + 1 - F_n(z_2); \end{aligned} \tag{5.6.12}$$

nella relazione (5.6.12) qui sopra  $F_n$  rappresenta la funzione di ripartizione di una  $\chi^2(n)$ , che è appunto, come si ricorderà, la legge della variabile che compare in (5.6.12), cioè

$$\frac{\sum_{i=1}^n (X_i - \mu_0)^2}{\sigma_0^2}.$$

Riprendiamo le considerazioni sul livello del nostro test. Dalla (5.6.12) segue che esso sarà pari ad  $\alpha$  se vale la relazione  $F_n(z_1) + 1 - F_n(z_2) = \alpha$ , una scelta possibile (ma non l'unica!) è  $z_1 = \chi_{\frac{\alpha}{2}, n}^2$  e  $z_2 = \chi_{1-\frac{\alpha}{2}, n}^2$ , e la regione critica è

$$\{T < \chi_{\frac{\alpha}{2}, n}^2\} \cup \{T > \chi_{1-\frac{\alpha}{2}, n}^2\}$$

(b) *Test unilaterale*  $H_0 : \sigma^2 \leq \sigma_0^2$ ,  $H_1 : \sigma^2 > \sigma_0^2$ .

Anche qui cerchiamo di intuire di che tipo deve essere la regione critica. Con le stesse considerazioni del punto (a), questa volta siamo indotti a cercare una regione critica del tipo

$$\{T > z\} = \left\{ \frac{\sum_{i=1}^n (X_i - \mu_0)^2}{\sigma_0^2} > z \right\}.$$

Dobbiamo ora determinare  $z$ , come sempre in funzione di  $\alpha$ . La procedura è la stessa di prima. Calcolando esplicitamente il livello del test si trova

$$\begin{aligned} \alpha^* &= \sup_{\sigma^2 \leq \sigma_0^2} P^{\sigma^2} \left( \frac{\sum_{i=1}^n (X_i - \mu_0)^2}{\sigma_0^2} > z \right) \\ &= \sup_{\sigma^2 \leq \sigma_0^2} P^{\sigma^2} \left( \frac{\sum_{i=1}^n (X_i - \mu_0)^2}{\sigma^2} > z \frac{\sigma_0^2}{\sigma^2} \right) = \sup_{\sigma^2 \leq \sigma_0^2} (1 - F_n(z \sigma_0^2 / \sigma^2)) \\ &= 1 - F_n(z). \end{aligned} \tag{5.6.13}$$

In questo caso, ad avere legge  $\chi^2(n)$  è la v. a.

$$\frac{\sum_{i=1}^n (X_i - \mu_0)^2}{\sigma^2},$$

dato che il vero valore della varianza del campione è appunto  $\sigma^2$  e non, come nel caso (a),  $\sigma_0^2$ . L'ultima uguaglianza in (5.6.13) segue dalla crescenza e continuità a destra di  $F_n$ : per  $\sigma^2 \leq \sigma_0^2$  si ha  $z \sigma_0^2 / \sigma^2 \geq z$  e quindi

$$\inf_{\sigma^2 \leq \sigma_0^2} F_n(z \sigma_0^2 / \sigma^2) = \lim_{t \downarrow z} F_n(t) = F_n(z).$$

Dalla (5.6.13) segue che il livello del test sarà pari ad  $\alpha$  se vale la relazione  $1 - F_n(z) = \alpha$ , cioè per  $z = \chi_{1-\alpha, n}^2$ , e la regione critica è

$$\{T > \chi_{1-\alpha, n}^2\}.$$

(c) *Test unilaterale*  $H_0 : \sigma^2 \geq \sigma_0^2$ ,  $H_1 : \sigma^2 < \sigma_0^2$ .

Non ripeteremo i conti per la terza volta. Diamo solo la formula finale della regione critica, che è

$$\{T < \chi_{\alpha, n}^2\}.$$

**5.6.14. Test per la varianza, media non nota (Test di Fisher-Snedecor).** Se la media  $\mu_0$  non è nota, la v. a.  $T$  utilizzata precedentemente nel caso di media nota non è una statistica, perchè dipende da  $\mu_0$ . L'idea è la stessa di quella usata a proposito degli intervalli di fiducia: sostituire  $\mu_0$  con il suo stimatore  $\bar{X}$ , cioè utilizzare la statistica

$$R = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{\sigma_0^2} = \frac{S^2}{\sigma_0^2} (n-1),$$

dove, come sempre,  $S^2$  indica la varianza campionaria (Paragrafo 5.2.8, esempio (c)).

Come sappiamo dal teorema di Cochran, la v. a.

$$\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{\sigma^2} = \frac{S^2}{\sigma^2} (n-1)$$

ha legge  $\chi^2(n-1)$ , anziché  $\chi^2(n)$  come invece accadeva per la v. a.

$$\frac{\sum_{i=1}^n (X_i - \mu_0)^2}{\sigma^2},$$

che abbiamo usato sopra nel caso della media nota. Quindi si possono ripetere tutti i ragionamenti precedenti, sostituendo semplicemente  $n-1$  al posto di  $n$  e i quantili della  $\chi^2(n-1)$  al posto di quelli della  $\chi^2(n)$  nelle formule trovate per il caso di media nota. Riportiamo comunque le tre regioni critiche, per comodità.

(a) *Test bilaterale*  $H_0 : \sigma^2 = \sigma_0^2$ ,  $H_1 : \sigma^2 \neq \sigma_0^2$ .

Regione critica

$$\{R < \chi_{\frac{\alpha}{2}, n-1}^2\} \cup \{R > \chi_{1-\frac{\alpha}{2}, n-1}^2\}$$

(b) *Test unilaterale*  $H_0 : \sigma^2 \leq \sigma_0^2$ ,  $H_1 : \sigma^2 > \sigma_0^2$ .

Regione critica

$$\{R > \chi_{1-\alpha, n-1}^2\}$$

(c) *Test unilaterale*  $H_0 : \sigma^2 \geq \sigma_0^2$ ,  $H_1 : \sigma^2 < \sigma_0^2$ .

Regione critica

$$\{R < \chi_{\alpha, n-1}^2\}.$$

**Esercizio 5.6.15.** Per le tre forme del test di Fisher-Snedecor, calcolare la formula della potenza, sia nel caso di media nota che in quello di media non nota.

**5.6.16. Test per il parametro della bernoulliana.** Qui supporremo che la legge delle  $(X_i)$  sia  $B(1, p)$ , con  $p$  non nota. Scopo del test è confrontare  $p$  con un valore assegnato  $p_0$ . Il campione di osservazioni non è gaussiano; tuttavia vedremo che, per  $n$  abbastanza grande, potremo servirci del Teorema Limite Centrale, riconducendoci al caso dei test per campioni gaussiani visti sopra.

In tutti i casi che studieremo la statistica sarà

$$T = \frac{\bar{X} - p_0}{\sqrt{p_0(1-p_0)}} \sqrt{n}.$$

(a) *Test bilaterale*  $H_0 : p = p_0, H_1 : p \neq p_0$ .

Se il campione è abbastanza grande e se  $H_0$  è vera, allora  $T$  ha approssimativamente legge  $\mathcal{N}(0, 1)$  (ricordare che la media e la varianza di una  $B(1, p_0)$  sono uguali rispettivamente a  $p_0$  e  $p_0(1-p_0)$ ). Dunque, per  $n$  abbastanza grande ( $n \geq 25 \div 50$ ), si possono ripetere esattamente le considerazioni fatte per il test bilaterale di Student con media nota (5.6.1)(a), e si trova dunque la regione critica

$$\{|T| > \phi_{1-\alpha/2}\}.$$

(b) *Test unilaterale*  $H_0 : p \leq p_0, H_1 : p > p_0$ .

Ragionando esattamente come nel caso (5.6.1)(b), si vede che la regione critica è ancora del tipo  $\{T > z\}$ , con  $z$  da determinare in funzione di  $\alpha$ . Si può dimostrare (ma noi tralascieremo i conti) che  $\forall z \in \mathbb{R}$  la funzione

$$p \mapsto P^p(T > z)$$

è crescente e continua. Quindi, calcolando la taglia del test, si trova, per  $n$  grande,

$$\alpha^* = \sup_{p \leq p_0} P^p(T > z) = P^{p_0}(T > z) \simeq 1 - \Phi(z),$$

dato che, per  $p = p_0$ , la statistica  $T$  è approssimativamente una  $\mathcal{N}(0, 1)$ , come abbiamo detto sopra.

Come nel caso (5.6.1)(b) si trova allora che la regione critica è

$$\{T > \phi_{1-\alpha}\}.$$

(c) *Test unilaterale*  $H_0 : p \geq p_0, H_1 : p < p_0$ .

Diamo solo le conclusioni, che si ottengono, ancora per  $n$  grande, usando il teorema limite centrale (e la decrescenza e continuità della funzione  $p \mapsto P^p(T < z)$ ). La regione critica di livello  $\alpha$  è

$$\{T < \phi_\alpha\}.$$

**Osservazione 5.6.17.** È facile, a questo punto, capire che le formule (approssimate) della potenza si ottengono, in ciascun caso, dalle relative formule del test di Student con varianza nota, semplicemente sostituendo  $p_0$  e  $p_0(1 - p_0)$  al posto di  $\mu_0$  e  $\sigma_0^2$ . Scriverle per esercizio.



## Il test del chi-quadro per l'adattamento (goodness-for-fit test)

### 6.1 Motivazione

In certe situazioni si tratta di decidere se una data variabile aleatoria  $X$  segue una certa legge oppure no.

**Esempio 6.1.1.** Un esempio può essere quello della coltura di batteri che abbiamo visto in (5.1.2), in cui ci si chiedeva se il numero di batteri in una porzione di coltura seguiva o meno una legge di Poisson.

Nell'Esempio 5.1.2 abbiamo già messo in evidenza l'importanza pratica di questa domanda (significa che ci stiamo chiedendo se la distribuzione dei batteri nella coltura è omogenea oppure ci sono aggregazioni di batteri dovute a motivi da indagare).

### 6.2 Descrizione del test del $\chi^2$

**6.2.1. Primo caso.** Vedremo prima di tutto la situazione tipica in cui applicare il test del  $\chi^2$ .

**Esempio introduttivo 6.2.2.** Da un'urna contenente palline di cinque colori diversi, denotati con 1, 2, 3, 4, 5, si eseguono 1000 estrazioni con rimpiazzo, ottenendo i risultati riportati nella tabella seguente:

<i>n° colore</i>	<i>n° risultati</i>
1	225
2	211
3	172
4	182
5	210

Si può ritenere che la proporzione di palline di ciascun colore sia la stessa?

Cerchiamo, al solito, di formalizzare la situazione, nel modo seguente. La generica osservazione  $X$  (risultato di un'estrazione nel caso dell'esempio) è una v. a. a valori in un insieme finito (numerico o no) che indichiamo con  $\{x_1, x_2, \dots, x_m\}$  ( $\{1, 2, 3, 4, 5\}$  per l'estrazione dall'urna). La legge di  $X$  è

$$v_i = P(X = x_i), \quad i = 1, \dots, m.$$

In altre parole, la legge di  $X$  è determinata dal parametro  $\theta = (v_1, \dots, v_m)$  (si tratta di un parametro *vettoriale* questa volta!);  $\theta$  varia nell'insieme  $\Theta$  di tutti i vettori  $(v_1, \dots, v_m)$  tali che  $v_i \geq 0$  per ogni  $i = 1, \dots, m$  e  $\sum_{i=1}^m v_i = 1$ :

$$\Theta = \{(v_1, \dots, v_m) : v_i \geq 0; \sum_{i=1}^m v_i = 1\}.$$

Ci chiediamo se l'osservazione segue una legge data da un particolare vettore  $\theta_0 \in \Theta$ : indicheremo le componenti di  $\theta_0$  con  $(p_1, \dots, p_m)$ .

Per esempio, nel caso dell'urna contenente palline di cinque colori, ci chiediamo se

$$P(X = 1) = P(X = 2) = \dots = P(X = 5) = \frac{1}{5},$$

ovvero se

$$\theta_0 = \left(\frac{1}{5}, \frac{1}{5}, \dots, \frac{1}{5}\right).$$

Per effettuare il test, si fanno al solito  $n$  osservazioni, che supponiamo tra loro indipendenti e tutte con la stessa legge di  $X$  (nel caso dell'urna, ad esempio, si eseguono 1000 estrazioni). Indichiamole con  $X_1, X_2, \dots, X_n$  e poniamo, per ogni  $i = 1, 2, \dots, m$ ,

$$O_i = \text{card}\{k \leq n : X_k = x_i\}.$$

$O_i$  non è altro che il numero di osservazioni che hanno dato valore  $x_i$ , e viene chiamato *effettivo empirico* di  $x_i$  (il termine *empirico* sta ad indicare che  $O_i$  è il valore *realmente* ottenuto nell'esperimento).

Poniamo inoltre

$$\tilde{p}_i = \frac{O_i}{n}. \quad (6.2.3)$$

Le quantità  $\tilde{p}_i$  sono le proporzioni di risultati  $x_i$ , e si chiamano *probabilità* (o *frequenze relative*) *empiriche*. In corrispondenza, i numeri  $p_i$  vengono detti *probabilità teoriche*, e si chiamano *effettivi teorici* i numeri

$$E_i = np_i. \quad (6.2.4)$$

Questi ultimi indicano il numero di volte nelle quali, in teoria, dovremmo aspettarci il risultato  $x_i$  se le probabilità fossero quelle teoriche, cioè le  $p_i$ .

È comunemente accettata la regola pratica di ritenere attendibile il test se

$$E_i = np_i \geq 5 \quad \forall i.$$

Il test del  $\chi^2$  discrimina tra:

$$\begin{aligned} H_0 : X &\text{ segue la legge teorica, ovvero } \theta = \theta_0; \\ H_1 : X &\text{ non segue la legge teorica, ovvero } \theta \neq \theta_0, \end{aligned}$$

e fa uso della statistica (detta *di Pearson* dal nome dello statistico che ne ha studiato per primo le proprietà)

$$T = \sum_{i=1}^m \frac{(O_i - E_i)^2}{E_i}.$$

Grazie alle relazioni (6.2.3) e (6.2.4) è facile vedere che essa è uguale a

$$\sum_{i=1}^m \frac{(n\tilde{p}_i - np_i)^2}{np_i} = n \sum_{i=1}^m \frac{(\tilde{p}_i - p_i)^2}{p_i}.$$

Basandosi sulla statistica  $T$  sopra definita, costruiamo ora una regione critica per il test.

Ragioniamo così. Se  $X$  non segue la legge data, allora esiste almeno un indice  $j$  tale che  $v_j = P(X = x_j) \neq p_j$ . D'altra parte, per la Legge dei Grandi Numeri, risulta, per  $n \rightarrow \infty$ ,

$$\tilde{p}_i \longrightarrow v_i.$$

Pertanto, nella sommatoria

$$R := \sum_{i=1}^m \frac{(\tilde{p}_i - p_i)^2}{p_i}$$

almeno il termine

$$\frac{(v_j - p_j)^2}{p_j}$$

risulterà strettamente positivo, e dunque, per  $n$  abbastanza grande, la statistica  $T = nR$  tenderà a prendere valori strettamente positivi e grandi.

Questo ci fa pensare che una regione critica ragionevole deve essere del tipo

$$\{T > k\};$$

se  $\alpha$  è il livello fissato, ricordando che l'ipotesi nulla è  $\theta = \theta_0$ , potremo determinare  $k$  sotto la condizione

$$\alpha^* = P^{\theta_0} \{T > k\} \leq \alpha,$$

e il primo termine della precedente disuguaglianza potrà essere calcolato a patto di conoscere la legge di  $T$  quando  $\theta = \theta_0$  (cioè nel caso che le osservazioni seguano effettivamente la legge teorica determinata dal vettore  $\theta_0$ ). Ci viene in aiuto il seguente risultato, che non dimostriamo:

**Teorema 6.2.5 (di Pearson).** *Sia  $(X_n)_{n \geq 1}$  una successione di v.a. indipendenti, aventi tutte la legge determinata dal vettore  $\theta_0$ , cioè tali che, per ogni  $n$ , risulti, per  $i = 1, \dots, m$*

$$P(X_n = x_i) = p_i.$$

*Allora, per ogni  $x \in \mathbb{R}$  si ha*

$$\lim_{n \rightarrow \infty} P(T \leq x) = F_{m-1}(x),$$

*dove con  $F_{m-1}$  si è indicata la funzione di ripartizione di una v. a. avente legge  $\chi^2(m-1)$ .*

Dal teorema di Pearson si deduce allora che, per  $n$  abbastanza grande, approssimativamente sarà

$$P^{\theta_0}(T > k) \simeq 1 - F_{m-1}(k);$$

la condizione da imporre a  $k$  sarà dunque

$$F_{m-1}(k) = 1 - \alpha,$$

ovvero  $k$  dovrà essere il quantile di ordine  $1 - \alpha$  della legge  $\chi^2(m-1)$ ; in simboli

$$k = \chi_{1-\alpha, m-1}^2,$$

e la regione critica del test sarà

$$\{T > \chi_{1-\alpha, m-1}^2\}.$$

**Esempio 6.2.6.** Ritorniamo all'esempio dell'urna, costruendo la tabella seguente, che contiene anche i valori delle altre quantità utili per il calcolo della statistica  $T$  (il cui valore osservato  $t$  è riportato in basso a destra).

$x_i$	$O_i$	$E_i$	$O_i - E_i$	$(O_i - E_i)^2 / E_i$
1	225	200	25	3.17
2	211	200	11	0.65
3	172	200	28	3.92
4	182	200	-18	1.62
5	210	200	10	0.50
				$t = 9.86$

Si può dire che la proporzione di palline di ciascun colore è la stessa?

Osserviamo che

$$E_i = np_i = 1000 \times \frac{1}{5} = 200$$

e che la regola pratica  $E_i \geq 5, \forall i$  in questo caso è valida e consente di ritenere attendibile il risultato del test.

Sia  $\alpha = 0.05$  il livello del test. Poiché  $\chi_{0.95,4}^2 = 9.48$ , il valore assunto dalla statistica ( $t = 9.86$ ) cade nella regione critica e l'ipotesi è respinta: al livello  $\alpha = 0.05$  le proporzioni dei vari colori sono da ritenere diverse.

**Osservazione 6.2.7.** Però il valore della statistica  $T$  ottenuto è *di poco* superiore al quantile: qualche dubbio resta.

**6.2.8. Secondo caso.** Il test del  $\chi^2$  può essere adattato al caso di variabili aleatorie continue oppure discrete ma con un numero infinito, o comunque molto grande, di valori, come ora vedremo.

Sia  $X$  una v. a. del tipo sopra detto, ma con legge non nota, e sia  $X_1, \dots, X_n$  un campione di osservazioni con la stessa legge di  $X$ . Sia  $F$  una assegnata funzione di ripartizione: l'ipotesi nulla è

$$H_0 : X \text{ ammette } F \text{ come funzione di ripartizione.}$$

Vediamo come si procede.

Suddividiamo  $E = Im(X)$  in sottoinsiemi  $I_1, I_2, \dots, I_m$  (che generalmente sono intervalli) tali che

$$I_i \cap I_j = \emptyset \text{ per } i \neq j, \quad \cup_{k=1}^m I_k = E.$$

Diremo che  $X$  “assume il valore  $x_i$ ” se  $X \in I_i$ .

Supponiamo che sia  $I_i = (a_i, b_i]$ ; sotto l'ipotesi  $H_0$  si ha allora

$$p_i = P(X \in I_i) = F(b_i) - F(a_i);$$

queste sono dunque le nostre probabilità teoriche.

Gli effettivi teorici si calcolano a questo punto nel modo consueto; notiamo inoltre che

$$O_i = \text{card}\{k \leq n : X_k \in I_i\};$$

le probabilità empiriche sono ancora definite come  $\tilde{p}_i = O_i/n$ .

**Osservazione 6.2.9.** Di regola il numero di intervalli  $I_i$  in cui  $E$  è suddiviso dovrebbe essere abbastanza grande, altrimenti c'è il rischio di non poter distinguere tra leggi poco diverse tra loro. Tuttavia, se la suddivisione è troppo fine, può capitare che  $p_i = P(X \in I_i)$  sia piccola (almeno per qualche indice  $i$ ) e che quindi risulti  $np_i \leq 5$ , violando così la condizione di attendibilità del test. In genere dunque bisogna ricorrere a dei compromessi, da valutare caso per caso.

**Osservazione 6.2.10.** Può capitare che, invece di dover dire se  $X$  segue una legge nota, si debba decidere se  $X$  segue una legge appartenente ad una data famiglia, dipendente da  $r$  parametri incogniti. In tal caso si esegue il test

precedente, stimando prima i parametri con stime ottenute dai dati (in genere si utilizzano le stime di massima verosimiglianza). In questo caso però la legge del  $\chi^2$  da utilizzare non è a  $m - 1$ , bensì a  $m - 1 - r$  gradi di libertà (ogni parametro stimato riduce di uno i gradi di libertà).

Quello appena descritto è il classico *test di adattamento* (*goodness-for-fit test* in inglese).

**Esempio 6.2.11 (Adattamento alla legge  $\mathcal{N}(0, 1)$ ).** Nella tabella sottostante sono riportati 63 numeri aleatori. Si può affermare che il vettore  $X_1, \dots, X_{63}$  è un campione estratto dalla legge  $\mathcal{N}(0, 1)$ ?

−0.69; 1.52; −0.42; −0.39; −0.82; 0.36; −0.45; 1.25; −2.05; −0.97; −0.15; 1.41; 0.27; −1.12; 0.42; −0.98; 0.99; −0.59; 1.56; 0.70; −0.72; −0.09; 3.11; −0.43; −0.73; 0.51; −0.79; 0.72; −1.25; 2.01; −0.95; 0.52; −1.02; 2.07; 0.74; 0.40; 3.05; 0.75; 0.03; 0.63; 0.29; 0.99; 0.05; 0.58; 0.32; 0.12; 0.88; −1.28; 0.19; 0.73; −0.91; 0.79; −1.13; −0.63; −0.83; −0.17; 0.87; −0.02; 3.14; −0.21; 0.59; 0.92; −0.81.

Si tratta di scegliere la suddivisione in intervalli  $I_1, I_2, \dots, I_m$  in modo opportuno. Poniamo

$$I_1 = (-\infty, a_1], I_2 = (a_1, a_2], \dots, I_{m-1} = (a_{m-2}, a_{m-1}], I_m = (a_{m-1}, +\infty);$$

dunque avremo, per  $i = 1, \dots, m$ ,

$$\begin{aligned} p_1 &= \Phi(a_1), \quad p_2 = \Phi(a_2) - \Phi(a_1), \dots, \\ p_{m-1} &= \Phi(a_{m-1}) - \Phi(a_{m-2}), \quad p_m = 1 - \Phi(a_{m-1}). \end{aligned}$$

Per semplicità (ma non sarebbe necessario) faremo in modo che sia  $p_i = 1/m$  per ogni  $i$ , con  $m$  da determinare. Dunque

$$\begin{aligned} p_1 &= \Phi(a_1) = 1/m \Rightarrow a_1 = \phi_{1/m}; \\ p_2 &= \Phi(a_2) - \Phi(a_1) = 1/m \Rightarrow \Phi(a_2) = 2/m \Rightarrow a_2 = \phi_{2/m} \\ &\vdots \\ p_{m-1} &= \Phi(a_{m-1}) - \Phi(a_{m-2}) = 1/m \\ &\Rightarrow \Phi(a_{m-1}) = (m-1)/m \Rightarrow a_{m-1} = \phi_{(m-1)/m}. \end{aligned}$$

In altri termini abbiamo fatto la scelta

$$a_i = \phi_{i/m}, \quad i = 1, \dots, m-1.$$

Al solito, per l'applicabilità del test deve essere  $np_i \geq 5$ , ovvero, in questo caso,  $63/m \geq 5$ , da cui  $m \leq 63/5 = 12.6$  (come si vede, il numero  $m$  delle classi in cui si effettua la suddivisione non può essere troppo grande). Prendiamo

allora il numero  $m$  più grande possibile, cioè  $m = 12$ , e calcoliamo con le tavole  $\phi_{i/12}$  per ogni  $i = 1, 2, \dots, 11$ . Si ha ( $1/12 = 0.083$ )

$$\begin{aligned}\phi_{1/12} &= -\phi_{11/12} = -\phi_{0.91} = -1.34; \\ \phi_{2/12} &= -\phi_{10/12} = -\phi_{0.83} = -0.96; \\ \phi_{3/12} &= -\phi_{9/12} = -\phi_{0.75} = -0.68; \\ \phi_{4/12} &= -\phi_{8/12} = -\phi_{0.66} = -0.41; \\ \phi_{5/12} &= -\phi_{7/12} = -\phi_{0.58} = -0.20; \\ \phi_{6/12} &= \phi_{0.5} = 0.\end{aligned}$$

Inoltre per ogni  $i$  si ha  $E_i = 63/12 = 5.25$ .

Otteniamo allora la tabella seguente:

$x_i$	$O_i$	$E_i$	$O_i - E_i$	$(O_i - E_i)^2 / E_i$
$\leq -1.47$	1	5.25	-4.25	18.06
$-1.34 \leftrightarrow -0.96$	7	5.25	1.75	0.43
$-0.96 \leftrightarrow -0.68$	9	5.25	3.75	1.56
$-0.68 \leftrightarrow -0.41$	5	5.25	-0.25	0.01
$-0.41 \leftrightarrow -0.20$	2	5.25	-3.25	5.28
$-0.20 \leftrightarrow 0.00$	4	5.25	-1.25	0.39
$0.00 \leftrightarrow 0.20$	4	5.25	-1.25	0.39
$0.20 \leftrightarrow 0.41$	5	5.25	-0.25	0.01
$0.41 \leftrightarrow 0.68$	6	5.25	-0.75	0.09
$0.68 \leftrightarrow 0.96$	9	5.25	-3.75	1.56
$0.96 \leftrightarrow 1.34$	3	5.25	-2.25	1.68
$\geq 1.34$	8	5.25	2.75	0.94
				$t = 30.94$

Prendendo  $\alpha = 0.05$  come livello per il test, risulta  $\chi_{0.95,11}^2 = 19.675$  e  $H_0$  è respinta.

**Osservazione 6.2.12.** Può capitare che la suddivisione in intervalli sia suggerita dal tipo di problema (e non costruita a posteriori come è accaduto nell'esempio precedente). In tal caso, se gli effettivi teorici risultano troppo bassi, si usa il trucco di raggruppare alcune classi (in modo ragionevole!) come mostra l'esempio seguente (nel quale si ha anche la necessità di stimare un parametro).

**Esempio 6.2.13** (dal testo [6]). Si è effettuata un'indagine in un parco nazionale inglese, con lo scopo di studiare la distribuzione del numero di tane di volpe. Sono stati ispezionati 78 ettari di bosco, e il rilevamento è il seguente:

$n^\circ \text{ tane} = x_i$	$n^\circ \text{ ettari} = O_i$
0	16
1	46
2	12
3	2
4	2

Un'ipotesi di distribuzione "a caso" equivale all'affermazione che il numero di tane ha legge di Poisson (per motivi che già conosciamo, si veda (5.1.2)). In altri termini, se  $X$  rappresenta il numero di tane in un ettaro, l'ipotesi nulla è

$$H_0 : X \text{ ha legge } \Pi_\lambda.$$

Poiché il parametro  $\lambda$  non è noto esso deve essere stimato. In generale in queste situazioni si usano le stime di massima verosimiglianza; in questo caso, si può mostrare che la stima di massima verosimiglianza di  $\lambda$  è la media campionaria  $\bar{x}$  (verificarlo per esercizio):

$$\bar{x} = \frac{0 \times 16 + 1 \times 46 + 2 \times 12 + 3 \times 2 + 4 \times 2}{78} = 1.0769 = \hat{\lambda}.$$

Calcoliamo ora le probabilità e gli effettivi teorici

$$\begin{aligned} p_0 &= P(X = 0) = e^{-\lambda} = 0.3406; & E_0 &= 78 \times 0.3406 = 25.56 \\ p_1 &= P(X = 1) = \frac{\lambda^1}{1!} e^{-\lambda} = 0.3668; & E_1 &= 78 \times 0.3668 = 28.61; \\ p_2 &= P(X = 2) = \frac{\lambda^2}{2!} e^{-\lambda} = 0.1975; & E_2 &= 78 \times 0.1975 = 15.40; \\ p_3 &= P(X = 3) = \frac{\lambda^3}{3!} e^{-\lambda} = 0.071; & E_3 &= 78 \times 0.071 = 5.52; \\ p_4 &= P(X \geq 4) = 1 - \sum_{k=1}^3 p_k = 0.0241; & E_4 &= 78 \times 0.0241 = 1.87. \end{aligned}$$

Poiché l'ultimo effettivo teorico è risultato inferiore a 5, non è garantita l'attendibilità del test. Allora un modo per risolvere il problema può essere quello di conglobare le due classi finali in una sola, ottenendo un effettivo teorico pari a  $5.52 + 1.87 = 7.39 \geq 5$ ; così la tabella del rilevamento sarà la seguente (la classe finale, denotata con  $\geq 3$ , ha ora un effettivo empirico pari alla somma dei due effettivi empirici della tabella iniziale):

$n^\circ \text{ tane} = x_i$	$n^\circ \text{ ettari} = O_i$
0	16
1	46
2	12
$\geq 3$	4

Abbiamo ora  $p_3 = P(X = 3) + P(X \geq 4) = 0.071 + 0.0241 = 0.095$ .



Il test del  $\chi^2$  si effettua allora con  $m = 4$  risultati possibili e  $r = 1$  parametri stimati, quindi i gradi di libertà saranno  $m - r - 1 = 2$ .

Lo schema che si ottiene per il calcolo della statistica  $T$  di Pearson è il seguente:

$x_i$	$O_i$	$E_i$	$O_i - E_i$	$(O_i - E_i)^2 / E_i$
0	16	26.57	-10.57	4.20
1	46	28.61	17.39	10.57
2	12	15.40	-3.40	0.75
$\geq 3$	4	7.39	3.41	1.57
				$t = 17.099$

Al livello  $\alpha = 0.05$  la regione critica del test è

$$\{T \geq \chi^2_{1-\alpha,2}\} = \{T \geq \chi^2_{0.95,2}\} = \{T \geq 5.991\}.$$

Poiché il valore osservato della statistica di Pearson (17.099) cade nella regione critica,  $H_0$  è respinta: c'è una forte evidenza che le volpi non scelgono a caso il sito per la tana. Anzi, riguardando i dati, si osserva che l'effettivo empirico più alto corrisponde ad un numero di tane pari a 1; esso è molto superiore a quello che avremmo se la legge fosse di Poisson. Osservando poi l'ultima colonna della tabella, si vede che questo termine dà il contributo più alto al valore della statistica (10.57, più della metà!). Possiamo cioè affermare che, con grande evidenza, le volpi tendono ad evitare luoghi già occupati.

Anche il caso di 0 tane è interessante: fornisce il secondo contributo alla statistica, in ordine di grandezza (4.20).

**Osservazione 6.2.14.** Si può pensare che le volpi non si insediano volentieri in siti vuoti, ma questa ipotesi contraddice la precedente: servono dunque ulteriori indagini.

Una quantità spesso usata nella teoria dei test è il cosiddetto *livello di significatività* (detto anche *valore  $p$  dei dati*,  *$p$ -value* in inglese). Diamone la definizione. Supponiamo che il test che stiamo usando impieghi la statistica  $T$  e, per fissare le idee, che l'insieme dei valori  $D \subseteq \mathbb{R}$  che definisce la regione critica sia del tipo  $[k, +\infty)$  (in altri termini, la nostra regione critica sarà della forma  $\{T \geq k\}$ ); supponiamo poi che, in seguito alle nostre osservazioni,  $T$  abbia assunto il valore  $t$ . Si chiama *valore  $p$  dei dati* la quantità

$$\beta^* = \beta^*(t) = \sup_{\theta \in \Theta_0} P^\theta(T \geq t).$$

Sia  $\alpha^*$  il livello del test. È facile mostrare che:

- (i) se  $\alpha^* > \beta^*(t)$ , allora  $t \geq k$  e quindi l'ipotesi viene respinta;
- (ii) se  $\alpha^* < \beta^*(t)$ , allora  $t < k$  e quindi l'ipotesi non viene respinta.

(Provare a fare le dimostrazioni, per assurdo.)

In altri termini, possiamo dire che  $\beta^*(t)$  è l'estremo inferiore dei livelli  $\alpha^*$  per le quali l'ipotesi può essere respinta se l'osservazione ha fornito il valore osservato  $t$ .

Dunque, se il livello  $\alpha$  scelto è superiore a  $\beta^*(t)$ , l'ipotesi viene respinta; se invece il livello è inferiore a  $\beta^*(t)$ , essa viene accettata; per spiegarci meglio, supponiamo di aver trovato che  $\beta^*(t) = 0.0003$ . Questo è un valore molto piccolo, e quindi *a tutti i livelli usuali* ( $\alpha = 0.05$ , oppure  $\alpha = 0.01$  ecc.) l'ipotesi viene respinta (con il valore osservato  $t$ ). È quindi chiaro che  $H_0$  è (quasi sicuramente) falsa.

**Esercizio 6.2.15.** Trovare un metodo di calcolo per il valore  $p$  dei dati per i test gaussiani visti nel capitolo 5.

**Osservazione 6.2.16.** Il calcolo del  $p$ -value nell'esempio delle volpi dà:

$$\begin{aligned}\beta^* &= \beta^*(17.099) = \sup_{\theta=\theta_0} P^\theta(T \geq 17.099) = P^{\theta_0}(T \geq 17.099) = \\ &= 1 - F_2(17.099)\end{aligned}$$

da cui, invertendo con la funzione quantile e guardando le tavole

$$\chi^2_{1-\beta^*,2} = 17.099 \Rightarrow 1 - \beta^* \gg 0.99 \Rightarrow \beta^* \ll 0.01.$$

Ciò vuol dire che, ai livelli comunemente usati (maggiori di 0.01),  $H_0$  è *sempre* rifiutata.

## 6.3 Altre applicazioni del test del $\chi^2$

Il test del  $\chi^2$  può essere adattato ad una serie di situazioni di notevole interesse pratico, come mostreremo negli esempi che seguono.

**6.3.1. Test del  $\chi^2$  per l'indipendenza (o test di omogeneità).** Si tratta in pratica di un caso di test di adattamento con alcuni parametri stimati, ma è opportuno trattarlo nei dettagli.

Siano  $X$  e  $Y$  due variabili aleatorie (numeriche o no), che assumono rispettivamente i valori  $x_1, \dots, x_r$  e  $y_1, \dots, y_s$ , dunque la variabile aleatoria bivariata  $(X, Y)$  assume i valori  $(x_i, y_j)$ ,  $i = 1, \dots, r$ ;  $j = 1, \dots, s$ .

Si vuole verificare l'ipotesi

$$H_0 : X \text{ e } Y \text{ sono indipendenti.}$$

Poniamo

$$\begin{aligned}p_i &= P(X = x_i) & i &= 1, \dots, r; \\ q_j &= P(Y = y_j) & j &= 1, \dots, s.\end{aligned}$$

L'ipotesi  $H_0$  può essere allora espressa dicendo che la legge congiunta di  $(X, Y)$  coincide con il prodotto (tensoriale) delle leggi marginali, ovvero, per ogni coppia  $(i, j)$  si ha

$$P(X = x_i, Y = y_j) = P(X = x_i)P(Y = y_j) = p_i q_j.$$

Queste probabilità (teoriche) non sono note, e devono quindi essere stimate; si eseguono  $n$  osservazioni  $(X_1, Y_1), \dots, (X_n, Y_n)$  (tra loro indipendenti e con legge uguale a quella di  $(X, Y)$ ); per ogni coppia  $(i, j) \in \{1, \dots, r\} \times \{1, \dots, s\}$  poniamo

$$O_{i,j} = \text{effettivo empirico di } (x_i, y_j).$$

È evidente che le quantità

$$O_{i\cdot} \equiv \sum_{j=1}^s O_{i,j},$$

$$O_{\cdot j} \equiv \sum_{i=1}^r O_{i,j}$$

non sono altro che gli effettivi empirici di  $x_i$  e  $y_j$  rispettivamente.

Come già sappiamo, le probabilità teoriche possono essere stimate per mezzo delle stime di massima verosimiglianza; si può far vedere che tali stime sono date dai numeri

$$\hat{p}_i = \frac{O_{i\cdot}}{n}; \quad \hat{q}_j = \frac{O_{\cdot j}}{n};$$

le probabilità teoriche (stimate!) per la coppia  $(X, Y)$  sono allora

$$\hat{p}_i \hat{q}_j = \frac{O_{i\cdot} O_{\cdot j}}{n^2},$$

mentre gli effettivi teorici (stimati!) saranno

$$E_{i,j} = n \hat{p}_i \hat{q}_j = \frac{O_{i\cdot} O_{\cdot j}}{n}.$$

La statistica di Pearson vale dunque

$$T = \sum_{\substack{1 \leq i \leq r \\ 1 \leq j \leq s}} \frac{(O_{i,j} - E_{i,j})^2}{E_{i,j}} = \sum_{\substack{1 \leq i \leq r \\ 1 \leq j \leq s}} \frac{(O_{i,j} - (O_{i\cdot} O_{\cdot j})/n)^2}{(O_{i\cdot} O_{\cdot j})/n}.$$

Resta ora da stabilire qual è il numero di gradi di libertà della  $\chi^2$  da usare, e per questo è necessario contare quanti parametri sono stati stimati. I parametri  $p_i$  sono stati stimati per  $i = 1, \dots, r-1$  ( $p_r$  in realtà non lo è stato, poiché esso è determinato dalla relazione  $p_r = 1 - (p_1 + \dots + p_{r-1})$ ). Pertanto i parametri  $p_i$  stimati sono in numero di  $r-1$ . Per lo stesso motivo i parametri  $q_j$  stimati sono in numero di  $s-1$ , e quindi si ha un totale di  $k = r + s - 2$ .

Il numero  $m$  di valori assunti dalla variabile  $(X, Y)$  è evidentemente  $rs$ , e quindi la  $\chi^2$  avrà un numero di gradi di libertà pari a

$$m - 1 - k = rs - 1 - (r + s - 2) = (r - 1)(s - 1).$$

**Esempio di applicazione 6.3.2.** La tabella sottostante (detta *tabella di contingenza*) riporta i risultati ottenuti da due scuole,  $U$  e  $V$ , in un certo tipo di esame (uguale per entrambe) (*legenda*:  $P$  = promossi,  $A$  = ammessi all'orale,  $R$  = respinti):

$X \backslash Y$	$P$	$A$	$R$	$Tot.$	
$U$	75	53	32	160	$O_U$ .
$V$	140	62	38	240	$O_V$ .
$Tot.$	215	115	70	<b>400</b>	
	$O_{.P}$	$O_{.A}$	$O_{.R}$		

L'ipotesi nulla è

$H_0$  : i risultati sono indipendenti dalla scuola di appartenenza

(cioè le v. a.  $X$  = “scuola di appartenenza di un generico studente” e  $Y$  = “esito dell'esame di un generico studente” sono tra loro *indipendenti*).

Alcune osservazioni sulla lettura della tabella:

(i) la quantità  $O_U$ . (risp.  $O_V$ .) rappresenta il numero di studenti della scuola  $U$  (risp.  $V$ ) che hanno sostenuto l'esame, o, in altre parole, l'effettivo empirico di  $U$  (risp.  $V$ ). Esso si calcola semplicemente sommando i valori della tabella che si trovano sulla prima (risp. seconda) riga;

(ii) considerazioni analoghe (sulle colonne invece che sulle righe) si possono fare per le quantità  $O_{.P}$ ,  $O_{.A}$  e  $O_{.R}$ ;

(iii) il numero totale di osservazioni (cioè di studenti che hanno sostenuto l'esame) si può calcolare sommando gli effettivi  $O_U$ . e  $O_V$ . oppure gli effettivi  $O_{.P}$ ,  $O_{.A}$  e  $O_{.R}$ , e vale 400.

Osserviamo inoltre che l'ipotesi  $H_0$  può essere espressa nel modo seguente:

$H_0$  : i risultati ottenuti nelle due scuole sono fra loro omogenei.

Questo spiega il nome di *test di omogeneità* con cui talvolta si indica il test di indipendenza.

La tabella degli effettivi teorici (stimati) è la seguente:

	$P$	$A$	$R$	$Tot.$
$U$	86	46	28	160
$V$	129	69	42	240
$Tot.$	215	115	70	<b>400</b>

ad esempio:

$$E_{UP} = \frac{O_U \cdot O_P}{n} = \frac{160 \times 215}{400} = 86,$$

e così via per gli altri valori. Infine, ad esempio, la probabilità teorica (stimata)  $\hat{p}_U$  è calcolata come segue:

$$\hat{p}_U = \frac{O_U}{n} = \frac{160}{400} = 0.4;$$

nello stesso modo le altre. Inoltre i totali sono calcolati per controllo: ovviamente devono risultare uguali a quelli della tabella di contingenza data all'inizio.

La statistica di Pearson vale dunque

$$T = \frac{(75 - 86)^2}{86} + \frac{(53 - 46)^2}{46} + \frac{(32 - 28)^2}{28} + \frac{(140 - 129)^2}{129} + \frac{(62 - 69)^2}{69} + \frac{(38 - 42)^2}{42} = 5.07.$$

Il numero di gradi di libertà è  $(r - 1)(s - 1) = (2 - 1)(3 - 1) = 2$ . Vediamo cosa accade a due diversi livelli di significatività.

(i)  $\alpha = 0.1$ .

Si ha

$$\chi_{1-\alpha,2}^2 = \chi_{0.9,2}^2 = 4.60$$

e dunque  $H_0$  è respinta.

(ii)  $\alpha = 0.05$ .

Si ha

$$\chi_{1-\alpha,2}^2 = \chi_{0.95,2}^2 = 5.99$$

e dunque  $H_0$  non è respinta.

Confermiamo i risultati con il calcolo del  $p$ -value. Si tratta di risolvere rispetto all'incognita  $\beta^*$  l'equazione

$$\chi_{1-\beta^*}^2(2) = 5.07.$$

Lo facciamo per interpolazione lineare (vedi tabella sottostante):

0.90	$1 - \beta^*$	0.95
4.60	5.07	5.99

Il calcolo è il seguente

$$1 - \beta^* = \frac{0.95 - 0.90}{5.99 - 4.60}(5.07 - 4.60) + 0.90 = 0.917,$$

ovvero  $\beta^* = 0.083$ .

Poiché è

$$0.05 < 0.083 < 0.1,$$

$H_0$  deve essere respinta nel caso (i), mentre non può esserlo nel caso (ii).

**6.3.3. Test di simmetria.** Anche questo è un caso di test di adattamento con un certo numero di parametri stimati. La situazione è la seguente: si studia un campione di taglia  $n$  di una coppia di caratteri (v.a.)  $X$  e  $Y$ , entrambi a valori nell'insieme (numerico o no)  $\{x_1, x_2, \dots, x_k\}$ , e si ottiene la seguente tabella di contingenza (dove con il simbolo  $O_{i,j}$  si è indicato come sempre l'effettivo empirico della coppia  $(x_i, x_j)$ ).

$X \backslash Y$	$x_1$	$x_2$	$x_3$	$\dots$	$x_k$
$x_1$	$O_{11}$	$O_{12}$	$O_{13}$	$\dots$	$O_{1k}$
$x_2$	$O_{21}$				
$x_3$	$O_{31}$				
$\vdots$	$\vdots$				
$x_k$	$O_{k1}$				

Siamo interessati all'ipotesi

$$H_0 : \text{la legge congiunta di } (X, Y) \text{ è simmetrica}$$

o, ciò che è lo stesso

$$H_0 : p_{i,j} = P(X = x_i, Y = y_j) = P(X = x_j, Y = y_i) = p_{j,i} \quad \forall (i, j).$$

Dobbiamo ora costruire la tabella “teorica”, e per questo occorre stimare le probabilità teoriche  $p_{i,j}$ , che non sono note. Al solito per la Legge dei Grandi

Numeri si ha,

$$\lim_{n \rightarrow \infty} \frac{O_{i,j}}{n} = p_{i,j}.$$

Poiché le probabilità teoriche sono simmetriche, (cioé  $p_{i,j} = p_{j,i}$ ), anche le loro stime dovranno esserlo. Dunque è del tutto naturale utilizzare le quantità

$$\hat{p}_{i,j} = \frac{O_{i,j} + O_{j,i}}{2n},$$

esse infatti sono simmetriche e, per  $n \rightarrow \infty$ , convergono a

$$\frac{p_{i,j} + p_{j,i}}{2} = \frac{2p_{i,j}}{2} = p_{i,j}.$$

I parametri che verranno stimati sono allora i seguenti:

$$p_{1,1}, p_{1,2}, \dots, p_{1,k}$$

e di conseguenza anche

$$p_{2,1}, \dots, p_{k,1}$$

(si tratta di  $k$  parametri)

$$p_{2,2}, p_{2,3}, \dots, p_{2,k}$$

e di conseguenza anche

$$p_{3,2}, \dots, p_{k,2}$$

(si tratta di  $k-1$  parametri), e così via fino ai parametri  $p_{k-1,k-1}$  e  $p_{k-1,k}$  (2 parametri). Il parametro  $p_{k,k}$  non è stimato perché

$$p_{k,k} = 1 - \sum_{(i,j) \neq (k,k)} p_{i,j}.$$

In totale si ha dunque un numero di parametri stimati pari a

$$k + (k-1) + \dots + 2 = \frac{k^2 + k - 2}{2} := r. \quad (6.3.4)$$

Dunque gli effettivi teorici (stimati) sono

$$E_{i,j} = \hat{p}_{i,j}n = \frac{O_{i,j} + O_{j,i}}{2}$$

e la statistica di Pearson vale

$$\begin{aligned} T &= \sum_{i,j} \frac{(O_{i,j} - E_{i,j})^2}{E_{i,j}} = \sum_{i,j} \frac{(O_{i,j} - (O_{i,j} + O_{j,i})/2)^2}{(O_{i,j} + O_{j,i})/2} \\ &= \sum_{i,j} \frac{(O_{i,j} - O_{j,i})^2}{2(O_{i,j} + O_{j,i})} = \sum_{i < j} \frac{(O_{i,j} - O_{j,i})^2}{O_{i,j} + O_{j,i}}, \end{aligned}$$

dove l'ultima uguaglianza è dovuta al fatto che  $\sum_{i=j} = 0$  e che l'espressione  $\frac{(O_{i,j}-O_{j,i})^2}{2(O_{i,j}+O_{j,i})}$  è simmetrica nella coppia  $(i, j)$ .

Ricordando la (6.3.4), si ottiene che, sotto l'ipotesi nulla,  $T$  ha densità  $\chi^2(k^2 - 1 - r) = \chi^2((k(k-1)/2))$ , e la regione critica di livello  $\alpha$  per il test è

$$\{T > \chi^2_{1-\alpha, \frac{k(k-1)}{2}}\}.$$

**Esempio 6.3.5** (dal testo [3]). Si misura la capacità visiva di ciascun occhio in 7477 soggetti di sesso femminile e di età compresa fra i 30 e i 40 anni. La capacità visiva viene classificata in 4 gruppi, denotati con 1, 2, 3, 4. I dati sono i seguenti:

$X \backslash Y$	1	2	3	4
1	1520	266	124	66
2	234	1512	432	78
3	117	362	1772	205
4	36	82	179	492

Consideriamo i due caratteri (o variabili aleatorie)

$X$  = capacità visiva dell'occhio sinistro,

$Y$  = capacità visiva dell'occhio destro;

al livello  $\alpha = 0.05$  l'ipotesi che  $(X, Y)$  abbia legge simmetrica è accettata o respinta? (si potrebbe esprimere l'ipotesi di simmetria dicendo che la capacità visiva complessiva dei due occhi non dipende dalla lateralità).

Si lasciano le conclusioni per esercizio.



## Appendice A

---

### Richiami di calcolo combinatorio

**A.1.** Dati due insiemi finiti  $A$  e  $B$ , con  $\text{card}A = m$ ,  $\text{card}B = n$ , si ha

$$\text{card}(A \times B) = m \cdot n.$$

Possiamo anche dire che il numero di scelte possibili di un elemento di  $A$  e di un elemento di  $B$  è  $m \cdot n$ .

**A.2.** Siano  $A_1, A_2, \dots, A_n$   $n$  insiemi finiti, con  $\text{card}A_i = m_i$ . Si deduce facilmente per induzione da (A.1) che

$$\text{card}(A_1 \times A_2 \times \dots \times A_n) = m_1 \cdot m_2 \cdot \dots \cdot m_n.$$

**A.3.** È importante il caso particolare di (A.2) in cui sia

$$A_1 = A_2 = \dots = A_n = A.$$

Posto  $m = \text{card}A$ , si trova

$$\text{card}(A^n) = (\text{card}A)^n = m^n.$$

**A.4.** Siano dati due insiemi  $X$  e  $A$ , con  $\text{card}X = n$ ,  $\text{card}A = m$ . Allora il numero delle applicazioni di  $X$  in  $A$  è dato da  $m^n$ .

Infatti l'insieme delle applicazioni di  $X$  in  $A$  può essere messo in corrispondenza biunivoca con il prodotto cartesiano di  $n$  copie dell'insieme  $A$  (in che modo?), e quindi si applica (A.3).

**A.5.** Ogni elemento del prodotto cartesiano  $A^n = A \times A \times \dots \times A$  ( $n$  volte) può essere visto come una  $n$ -upla di elementi di  $A$ , dei quali, eventualmente, qualcuno sia ripetuto.

Ad esempio, se  $A = \{a_1, a_2, a_3, a_4\}$  e  $n = 3$ , gli elementi di  $A^3$  sono:  $(a_1, a_1, a_1), (a_1, a_1, a_2), (a_1, a_1, a_3), (a_1, a_1, a_4), (a_2, a_3, a_1) \dots$

Le  $n$ -uple di elementi di  $A$  si chiamano anche *sequenze con ripetizione* (di elementi di  $A$ ) di lunghezza  $n$ . Il numero di tali sequenze è dunque, come abbiamo appena visto  $(\text{card}A)^n$ .

**A.6.** Dato un insieme  $A$  formato da  $m$  oggetti, siamo interessati anche al numero di *sequenze senza ripetizione* (di elementi di  $A$ ) di lunghezza  $n$  ( $0 < n \leq m$ ). Esse vengono chiamate anche *disposizioni* di  $m$  oggetti a  $n$  a  $n$ . È facile vedere che questo numero è dato da

$$(m)_n := m \cdot (m-1) \cdot (m-2) \cdot \dots \cdot (m-n+1).$$

**A.7.** Siano  $X$  e  $A$  due insiemi come in (A.4), e supponiamo  $0 < n \leq m$ . Il numero delle applicazioni iniettive di  $X$  in  $A$  è dato da  $(m)_n$ .

**A.8.** È importante il caso particolare di (A.6) in cui sia  $n = m$ ;  $(m)_m$  è il numero di modi diversi in cui si possono disporre gli  $m$  elementi di  $A$ , o, come si dice, il numero delle *permutazioni* degli  $m$  oggetti che formano l'insieme  $A$ .  $(m)_m$  si indica più frequentemente con il simbolo  $m!$  (da leggere “ $m$  fattoriale”), e vale

$$m! = 1 \cdot 2 \cdot 3 \cdot \dots \cdot m.$$

**A.9.** Sia  $A$  come in (A.6). Due sequenze senza ripetizione di elementi di  $A$  possono differire per gli oggetti che le compongono, oppure anche soltanto per l'ordine in cui questi oggetti vengono presi; in molte situazioni, tuttavia, basterà contare il numero delle sequenze che differiscono per la natura degli elementi (identificando cioè quelle formate dagli stessi oggetti, presi in ordine diverso); ci interessa dunque trovare il numero di gruppi, costituiti da  $n$  elementi, che si possono formare a partire da un insieme di cardinalità  $m$ . Questi gruppi si chiamano *combinazioni* di  $m$  oggetti a  $n$  a  $n$ .

Non è difficile vedere che il numero che cerchiamo vale

$$\frac{(m)_n}{n!} = \frac{m \cdot (m-1) \cdot (m-2) \cdot \dots \cdot (m-n+1)}{n!} = \frac{m!}{n!(m-n)!} := \binom{m}{n}.$$

$\binom{m}{n}$  si chiama *coefficiente binomiale*  $m, n$ . Osserviamo che

$$\binom{m}{n} = \binom{m}{m-n}.$$

**A.10.** È necessario dare un senso anche al simbolo  $0!$  (il caso  $m = n = 0$  non rientra nella situazione considerata in (A.8)). Possiamo ragionare così: se in (A.9) prendiamo  $n = m$ , il coefficiente binomiale risultante dovrà essere il numero dei gruppi costituiti da  $m$  oggetti, formati a partire da un insieme di cardinalità  $m$ . È chiaro che di questi gruppi ne potremo formare uno solo

(l'insieme dato) e quindi dovremo avere

$$\binom{m}{m} = 1;$$

d'altra parte, sviluppando formalmente la formula che definisce il coefficiente binomiale, si trova

$$\binom{m}{m} = \frac{m!}{m!0!} = \frac{1}{0!}.$$

Dovremo pertanto porre  $0! = 1$ .

**A.11.** Sia  $A$  come in (A.6), e siano  $n_1, n_2, \dots, n_r$  degli interi  $\geq 0$ , tali che  $n_1 + n_2 + \dots + n_r = m$ . In quanti modi si possono formare  $r$  gruppi (a partire dagli elementi di  $A$ )  $G_1, G_2, \dots, G_r$ , in modo che il gruppo  $G_i$  contenga  $n_i$  elementi?

Per rispondere alla domanda, possiamo pensare di procedere così: formiamo prima un gruppo di  $n_1$  elementi, poi, con gli  $m - n_1$  elementi restanti, un gruppo di  $n_2$ , e così via. Il numero di gruppi possibili è quindi dato da

$$\binom{m}{n_1} \cdot \binom{m - n_1}{n_2} \cdot \dots \cdot \binom{m - n_1 - \dots - n_{r-1}}{n_r}.$$

Si vede facilmente che la quantità scritta sopra è uguale a

$$\frac{m!}{n_1! n_2! \cdot \dots \cdot n_r!}$$

e si indica con il simbolo

$$\binom{m}{n_1, n_2, \dots, n_r},$$

detto *coefficiente multinomiale*  $m, n_1, \dots, n_r$ .

Per  $r = 2$  si ha

$$\binom{m}{n_1, n_2} = \binom{m}{n_1} = \binom{m}{n_2}.$$

**A.12.** Se  $1 \leq n \leq m - 1$  si ha

$$\binom{m}{n} = \binom{m-1}{n-1} + \binom{m-1}{n}.$$

La dimostrazione di questa nota formula può essere fatta algebricamente sviluppando i coefficienti binomiali che vi compaiono, oppure, più semplicemente, ricordando il significato combinatorio di tali coefficienti: basta osservare che, a partire da un insieme di cardinalità  $m$ , si possono formare tutti i gruppi di  $n$  elementi togliendo uno degli oggetti, e prendendo poi dagli  $m - 1$  rimasti ogni gruppo costituito da  $n$  elementi ed ogni gruppo costituito da  $n - 1$  elementi (a questi ultimi sarà poi riaggiunto l'oggetto tolto all'inizio).

**A.13. Il binomio di Newton.** Siano  $a, b$  due numeri reali,  $n$  un intero non negativo. Si ha allora

$$(a + b)^n = \sum_{k=0}^n \binom{n}{k} a^k b^{n-k}.$$

La dimostrazione di questa famosa formula può essere fatta per induzione su  $n$ , sfruttando il risultato di (A.12) oppure, più semplicemente, osservando che moltiplicando  $(a + b)$  per se stesso  $n$  volte, il termine  $a^k b^{n-k}$  compare tante volte quanti sono i modi di formare gruppi di  $k$  a partire da un insieme di  $n$  oggetti ( $a^k b^{n-k}$  è il prodotto di  $n$  fattori:  $k$  volte il fattore  $a$ ,  $n - k$  volte il fattore  $b$ ).

**A.14.** Un caso particolare della formula precedente: prendendo  $a = b = 1$ , si trova

$$\sum_{k=0}^n \binom{n}{k} = 2^n.$$

Ricordando il significato combinatorio dei coefficienti binomiali  $\binom{n}{k}$ , la formula scritta sopra si può interpretare dicendo che il numero di sottoinsiemi che si possono formare a partire da un insieme di  $n$  elementi è pari a  $2^n$ .

## Appendice B

---

### Grafici e tavole

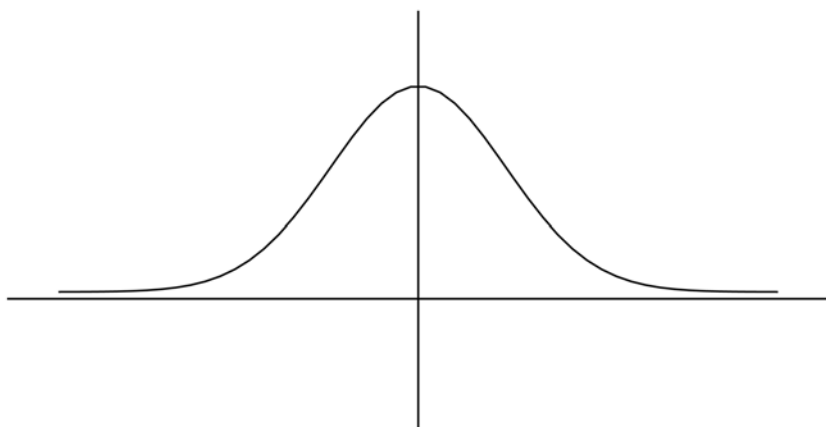


Grafico della densità  $N(0, 1)$

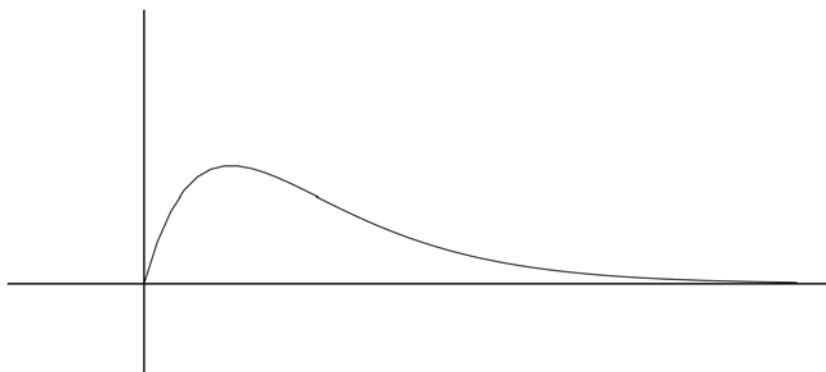


Grafico della densità  $\chi^2(n)$

Tavola dei quantili della distribuzione  $N(0, 1)$ 

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.50000	0.50399	0.50798	0.51197	0.51595	0.51994	0.52392	0.52790	0.53188	0.53586
0.1	0.53983	0.54380	0.54776	0.55172	0.55567	0.55962	0.56356	0.56749	0.57142	0.57535
0.2	0.57926	0.58317	0.58706	0.59095	0.59483	0.59871	0.60257	0.60642	0.61026	0.61409
0.3	0.61791	0.62172	0.62552	0.62930	0.63307	0.63683	0.64058	0.64431	0.64803	0.65173
0.4	0.65542	0.65910	0.66276	0.66640	0.67003	0.67364	0.67724	0.68082	0.68439	0.68793
0.5	0.69146	0.69497	0.69847	0.70194	0.70540	0.70884	0.71226	0.71566	0.71904	0.72240
0.6	0.72575	0.72907	0.73237	0.73565	0.73891	0.74215	0.74537	0.74857	0.75175	0.75491
0.7	0.75804	0.76115	0.76424	0.76730	0.77035	0.77337	0.77637	0.77935	0.78230	0.78524
0.8	0.78814	0.79103	0.79389	0.79673	0.79955	0.80234	0.80511	0.80785	0.81057	0.81327
0.9	0.81594	0.81859	0.82121	0.82381	0.82639	0.82894	0.83147	0.83398	0.83646	0.83891
1.0	0.84134	0.84375	0.84614	0.84849	0.85083	0.85314	0.85543	0.85769	0.85993	0.86214
1.1	0.86433	0.86650	0.86864	0.87076	0.87286	0.87493	0.87698	0.87900	0.88100	0.88298
1.2	0.88493	0.88686	0.88877	0.89065	0.89251	0.89435	0.89617	0.89796	0.89973	0.90147
1.3	0.90320	0.90490	0.90658	0.90824	0.90988	0.91149	0.91308	0.91466	0.91621	0.91774
1.4	0.91924	0.92073	0.92220	0.92364	0.92507	0.92647	0.92785	0.92922	0.93056	0.93189
1.5	0.93319	0.93448	0.93574	0.93699	0.93822	0.93943	0.94062	0.94179	0.94295	0.94408
1.6	0.94520	0.94630	0.94738	0.94845	0.94950	0.95053	0.95154	0.95254	0.95352	0.95449
1.7	0.95543	0.95637	0.95728	0.95818	0.95907	0.95994	0.96080	0.96164	0.96246	0.96327
1.8	0.96407	0.96485	0.96562	0.96638	0.96712	0.96784	0.96856	0.96926	0.96995	0.97062
1.9	0.97128	0.97193	0.97257	0.97320	0.97381	0.97441	0.97500	0.97558	0.97615	0.97670
2.0	0.97725	0.97778	0.97831	0.97882	0.97932	0.97982	0.98030	0.98077	0.98124	0.98169
2.1	0.98214	0.98257	0.98300	0.98341	0.98382	0.98422	0.98461	0.98500	0.98537	0.98574
2.2	0.98610	0.98645	0.98679	0.98713	0.98745	0.98778	0.98809	0.98840	0.98870	0.98899
2.3	0.98928	0.98956	0.98983	0.99010	0.99036	0.99061	0.99086	0.99111	0.99134	0.99158
2.4	0.99180	0.99202	0.99224	0.99245	0.99266	0.99286	0.99305	0.99324	0.99343	0.99361
2.5	0.99379	0.99396	0.99413	0.99430	0.99446	0.99461	0.99477	0.99492	0.99506	0.99520
2.6	0.99534	0.99547	0.99560	0.99573	0.99585	0.99598	0.99609	0.99621	0.99632	0.99643
2.7	0.99653	0.99664	0.99674	0.99683	0.99693	0.99702	0.99711	0.99720	0.99728	0.99736
2.8	0.99744	0.99752	0.99760	0.99767	0.99774	0.99781	0.99788	0.99795	0.99801	0.99807
2.9	0.99813	0.99819	0.99825	0.99831	0.99836	0.99841	0.99846	0.99851	0.99856	0.99861
3.0	0.99865	0.99869	0.99874	0.99878	0.99882	0.99886	0.99889	0.99893	0.99896	0.99900
3.1	0.99903	0.99906	0.99910	0.99913	0.99916	0.99918	0.99921	0.99924	0.99926	0.99929
3.2	0.99931	0.99934	0.99936	0.99938	0.99940	0.99942	0.99944	0.99946	0.99948	0.99950
3.3	0.99952	0.99953	0.99955	0.99957	0.99958	0.99960	0.99961	0.99962	0.99964	0.99965
3.4	0.99966	0.99968	0.99969	0.99970	0.99971	0.99972	0.99973	0.99974	0.99975	0.99976
3.5	0.99977	0.99978	0.99978	0.99979	0.99980	0.99981	0.99981	0.99982	0.99983	0.99983
3.6	0.99984	0.99985	0.99985	0.99986	0.99986	0.99987	0.99987	0.99988	0.99988	0.99989
3.7	0.99989	0.99990	0.99990	0.99990	0.99991	0.99991	0.99992	0.99992	0.99992	0.99992
3.8	0.99993	0.99993	0.99993	0.99994	0.99994	0.99994	0.99994	0.99995	0.99995	0.99995
3.9	0.99995	0.99995	0.99996	0.99996	0.99996	0.99996	0.99996	0.99996	0.99997	0.99997
4.0	0.99997	0.99997	0.99997	0.99997	0.99997	0.99997	0.99998	0.99998	0.99998	0.99998
4.1	0.99998	0.99998	0.99998	0.99998	0.99998	0.99998	0.99998	0.99998	0.99999	0.99999
4.2	0.99999	0.99999	0.99999	0.99999	0.99999	0.99999	0.99999	0.99999	0.99999	0.99999
4.3	0.99999	0.99999	0.99999	0.99999	0.99999	0.99999	0.99999	0.99999	0.99999	0.99999
4.4	0.99999	0.99999	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000

Tavola dei quantili della distribuzione  $T(n)$

n	Valore della funzione di ripartizione									
	0.75	0.8	0.85	0.9	0.95	0.975	0.99	0.995	0.999	0.9995
1	1.0000	1.3764	1.9626	3.0777	6.3137	12.7062	31.8210	63.6559	318.2888	636.5776
2	0.8165	1.0607	1.3862	1.8856	2.9200	4.3027	6.9645	9.9250	22.3285	31.5998
3	0.7649	0.9785	1.2498	1.6377	2.3534	3.1824	4.5407	5.8408	10.2143	12.9244
4	0.7407	0.9410	1.1896	1.5332	2.1318	2.7765	3.7469	4.6041	7.1729	8.6101
5	0.7267	0.9195	1.1558	1.4759	2.0150	2.5706	3.3649	4.0321	5.8935	6.8685
6	0.7176	0.9057	1.1342	1.4398	1.9432	2.4469	3.1427	3.7074	5.2075	5.9587
7	0.7111	0.8960	1.1192	1.4149	1.8946	2.3646	2.9979	3.4995	4.7853	5.4081
8	0.7064	0.8889	1.1081	1.3968	1.8595	2.3060	2.8965	3.3554	4.5008	5.0414
9	0.7027	0.8834	1.0997	1.3830	1.8331	2.2622	2.8214	3.2498	4.2969	4.7809
10	0.6998	0.8791	1.0931	1.3722	1.8125	2.2281	2.7638	3.1693	4.1437	4.5868
11	0.6974	0.8755	1.0877	1.3634	1.7959	2.2010	2.7181	3.1058	4.0248	4.4369
12	0.6955	0.8726	1.0832	1.3562	1.7823	2.1788	2.6810	3.0545	3.9296	4.3178
13	0.6938	0.8702	1.0795	1.3502	1.7709	2.1604	2.6503	3.0123	3.8520	4.2209
14	0.6924	0.8681	1.0763	1.3450	1.7613	2.1448	2.6245	2.9768	3.7874	4.1403
15	0.6912	0.8662	1.0735	1.3406	1.7531	2.1315	2.6025	2.9467	3.7329	4.0728
16	0.6901	0.8647	1.0711	1.3368	1.7459	2.1199	2.5835	2.9208	3.6861	4.0149
17	0.6892	0.8633	1.0690	1.3334	1.7396	2.1098	2.5669	2.8982	3.6458	3.9651
18	0.6884	0.8620	1.0672	1.3304	1.7341	2.1009	2.5524	2.8784	3.6105	3.9217
19	0.6876	0.8610	1.0655	1.3277	1.7291	2.0930	2.5395	2.8609	3.5793	3.8833
20	0.6870	0.8600	1.0640	1.3253	1.7247	2.0860	2.5280	2.8453	3.5518	3.8496
21	0.6864	0.8591	1.0627	1.3232	1.7207	2.0796	2.5176	2.8314	3.5271	3.8193
22	0.6858	0.8583	1.0614	1.3212	1.7171	2.0739	2.5083	2.8188	3.5050	3.7922
23	0.6853	0.8575	1.0603	1.3195	1.7139	2.0687	2.4999	2.8073	3.4850	3.7676
24	0.6848	0.8569	1.0593	1.3178	1.7109	2.0639	2.4922	2.7970	3.4668	3.7454
25	0.6844	0.8562	1.0584	1.3163	1.7081	2.0595	2.4851	2.7874	3.4502	3.7251
26	0.6840	0.8557	1.0575	1.3150	1.7056	2.0555	2.4786	2.7787	3.4350	3.7067
27	0.6837	0.8551	1.0567	1.3137	1.7033	2.0518	2.4727	2.7707	3.4210	3.6895
28	0.6834	0.8546	1.0560	1.3125	1.7011	2.0484	2.4671	2.7633	3.4082	3.6739
29	0.6830	0.8542	1.0553	1.3114	1.6991	2.0452	2.4620	2.7564	3.3963	3.6595
30	0.6828	0.8538	1.0547	1.3104	1.6973	2.0423	2.4573	2.7500	3.3852	3.6460
40	0.6807	0.8507	1.0500	1.3031	1.6839	2.0211	2.4233	2.7045	3.3069	3.5510
50	0.6794	0.8489	1.0473	1.2987	1.6759	2.0086	2.4033	2.6778	3.2614	3.4960
60	0.6786	0.8477	1.0455	1.2958	1.6706	2.0003	2.3901	2.6603	3.2317	3.4602
70	0.6780	0.8468	1.0442	1.2938	1.6669	1.9944	2.3808	2.6479	3.2108	3.4350
80	0.6776	0.8461	1.0432	1.2922	1.6641	1.9901	2.3739	2.6387	3.1952	3.4164
90	0.6772	0.8456	1.0424	1.2910	1.6620	1.9867	2.3685	2.6316	3.1832	3.4019
100	0.6770	0.8452	1.0418	1.2901	1.6602	1.9840	2.3642	2.6259	3.1738	3.3905
120	0.6765	0.8446	1.0409	1.2886	1.6576	1.9799	2.3578	2.6174	3.1595	3.3734
140	0.6762	0.8442	1.0403	1.2876	1.6558	1.9771	2.3533	2.6114	3.1495	3.3613
200	0.6757	0.8434	1.0391	1.2858	1.6525	1.9719	2.3451	2.6006	3.1315	3.3398
$\infty$	0.6745	0.8416	1.0364	1.2816	1.6449	1.9600	2.3263	2.5758	3.0902	3.2905





---

## Bibliografia

- [1] Baldi P., *Calcolo delle Probabilità e Statistica*, McGraw-Hill, 1998.
- [2] Caroti Ghelli F., *Calcolo delle Probabilità e Statistica*, ETS, 1997.
- [3] Dacunha-Castelle D., Duflo M., *Probabilités et statistiques*, Voll. 1–2, Hermann, 1975.
- [4] Dall’Aglio G., *Calcolo delle Probabilità*, Zanichelli, 2003.
- [5] Letta G., *Probabilità Elementare*, Zanichelli, 1994.
- [6] Moore P., Cobby, J., *Introductory Statistics for Environmentalists*, Prentice Hall, 1998.
- [7] Ross S.M., *Probabilità e statistica per l’ingegneria e le scienze*, Apogeo, 2008.