

ELEMENTI DI PROBABILITÀ E STATISTICA

Andrea Carpignani

14 gennaio 2023



Indice

1	Elementi di calcolo delle probabilità	1
1.1	Gli spazi probabilizzati	1
1.1.1	Esperimenti aleatori, ripetibilità	1
1.1.2	La tribù degli eventi, misura del grado di fiducia	3
1.1.3	Proprietà generali degli spazi probabilizzati	8
1.1.4	Misure di probabilità su uno spazio discreto	10
1.1.5	Elementi di calcolo combinatorio	12
1.1.6	Probabilità condizionale, formula di Bayes	19
1.1.7	Indipendenza	23
	Esercizi	26
1.2	Le variabili aleatorie	32
1.2.1	Definizione di variabile aleatoria, legge, indipendenza	32
1.2.2	Variabili aleatorie discrete, leggi discrete	36
1.2.3	La speranza di una variabile aleatoria discreta	43
1.2.4	Variabili aleatorie continue, densità di una legge, funzione di ripartizione	46
1.2.5	Il concetto generale di speranza	49
1.2.6	Vettori aleatori, leggi congiunte e leggi marginali	51
1.2.7	La legge condizionale	56
1.2.8	Varianza e covarianza di una variabile aleatoria	61
1.2.9	Leggi continue	64
1.2.10	La legge dei grandi numeri	69
	Esercizi	72
1.3	Il teorema limite centrale	76
1.3.1	Le leggi normali	76
1.3.2	La legge del chi–quadrato e la legge di Student	79
1.3.3	Uso della funzione di ripartizione, i quantili	81
1.3.4	Il teorema limite centrale	84
	Esercizi	88
2	Elementi di statistica inferenziale	92
2.1	La stima parametrica	92
2.1.1	Introduzione	92
2.1.2	Teoria della stima: la nozione di stimatore	96
2.1.3	La media empirica e la varianza empirica	98

2.1.4	Stimatori di massima verosimiglianza	101
2.1.5	Lo stimatore dei momenti	106
2.1.6	Gli intervalli di fiducia	108
2.1.7	Intervalli di fiducia per la media e per la varianza nei campioni gaussiani	111
	Esercizi	114
2.2	I test d'ipotesi statistiche	119
2.2.1	Introduzione	119
2.2.2	Il test di Student	122
2.2.3	Il test di Fisher–Snedecor	123
2.2.4	Il test del chi–quadro	125
	Esercizi	131
3	Cenni di statistica descrittiva	135
3.1	Rilevazione ed elaborazione dei dati	135
3.1.1	Introduzione	135
3.1.2	Ordinamento e frequenze	136
3.1.3	Rappresentazione grafica	139
3.1.4	Misure descrittive	142
3.1.5	Dati bidimensionali	146
3.1.6	Rette di regressione	149
	Esercizi	151
	Tavole numeriche	157
	La funzione di ripartizione della legge $\mathcal{N}(0, 1)$	158
	I quantili delle leggi $t(n)$ di Student	159
	I quantili delle leggi $\chi^2(n)$	160
	Indice analitico	160

Capitolo 1

Elementi di calcolo delle probabilità

1.1 Gli spazi probabilizzati

1.1.1 Esperimenti aleatori, ripetibilità

Un esperimento si dice *aleatorio*, per un certo individuo, in un certo istante, se l'individuo non è ancora in grado di indicarne con sicurezza il risultato (indipendentemente dal fatto che l'esperimento sia stato già eseguito o debba ancora essere eseguito). Se l'individuo che si trova in una tale situazione d'incertezza è interessato al risultato dell'esperimento (per esempio in vista di qualche scommessa), è naturale che egli si preoccupi innanzitutto di stabilire quali possano essere tutti i possibili risultati di questo esperimento (che, in termini probabilistici, prendono il nome di “eventualità”), indipendentemente dalla loro reale e concreta realizzabilità, ossia di “fissare un ventaglio completo di eventualità, a due a due incompatibili”. In termini più rigorosi, egli fisserà un insieme Ω , i cui elementi siano rappresentativi di tutti gli ipotetici risultati dell'esperimento, con la certezza che, comunque vadano le cose, il risultato effettivo dell'esperimento sicuramente “cadrà in Ω ”, nel senso che esso sarà rappresentato da uno ed un sol elemento di Ω . In altri termini, l'insieme Ω sarà costituito da tutti i “casi possibili” che si possono realizzare nell'ambito dell'esperimento aleatorio che l'individuo sta analizzando.

Esempio 1.1.1 (lancio di un dado) Si supponga che l'esperimento consista nel lanciare un dado. Se per “risultato” s'intende il numero della faccia che uscirà, si potrà prendere come Ω l'insieme degli interi compresi tra 1 e 6, cioè $\Omega = \{1, 2, 3, 4, 5, 6\}$.

Esempio 1.1.2 (estrazioni del lotto) Si supponga che l'esperimento sia costituito dalle estrazioni del lotto che verranno eseguite, oggi a mezzogiorno, sulla ruota di Firenze. Se per “risultato” s'intende l'insieme dei cinque numeri estratti (prescindendo dall'ordine d'estrazione), si potrà prendere come Ω l'insieme di tutte le cinquine, intendendo per *cinquina* un insieme di cinque distinti numeri interi compresi tra 1 e 90. Precisamente, si potrà prendere come Ω l'insieme formato da tutte le parti di $\{1, \dots, 90\}$ costituite da cinque elementi.

Esempio 1.1.3 (corse dei cavalli) Si supponga che l'esperimento consista in una corsa di 9 cavalli all'ippodromo di Livorno. Se per "risultato" s'intende l'ordine di arrivo dei cavalli partecipanti, ciascuno dei quali sia identificato da un numero compreso tra 1 e 9, si potrà prendere come Ω l'insieme di tutti i possibili ordinamenti degli interi $1, \dots, 9$, ovvero l'insieme formato da tutte le *permutazioni* dell'insieme $\{1, \dots, 9\}$. Così, ad esempio, la permutazione $\omega = (3, 7, 4, 9, 6, 2, 8, 1, 5)$ è quella in cui il cavallo numero 3 arriva primo, il cavallo numero 7 arriva secondo, il numero 4 arriva terzo e così via.

Esempio 1.1.4 (fila alla posta) Si supponga che l'esperimento aleatorio consista nello stabilire il numero di persone in coda ad un ufficio postale il venerdì alle ore 10:00. Sarà allora naturale prendere come Ω l'insieme di tutti i numeri interi positivi \mathbb{N} . Così, ad esempio, l'eventualità $\omega = 5$ significa che ci sono 5 persone in attesa. Come si vede, a differenza degli altri esempi, in questo caso l'insieme delle eventualità è infinito.

È bene osservare fin da subito che la scelta dell'insieme Ω è sempre in larga parte *arbitraria*. Non bisogna infatti dimenticare che gli elementi di Ω *rappresentano* ipotetici risultati dell'esperimento, secondo un opportuno *codice* che deve trascrivere nel linguaggio della matematica un evento concreto: è chiaro che la scelta di questo codice è, in larga misura, arbitraria. Ad esempio, se l'esperimento consiste nel lancio di una moneta, e se ci s'interessa solo alla faccia che apparirà (testa o croce), si potrà scegliere, come insieme delle eventualità, $\Omega = \{0, 1\}$, con la convenzione che 0 significhi croce, e 1 testa. Ma egualmente legittima sarebbe la convenzione inversa (0 = testa; 1 = croce), oppure la scelta, in luogo dell'insieme $\{0, 1\}$, di un qualsiasi altro insieme costituito da due elementi. Apparirà più chiaro in seguito, tuttavia, che limitare la scelta dell'insieme Ω ad insiemi numerici (dotati di strutture algebriche consistenti) sarà utile per schematizzare moltissimi problemi.

Osserviamo inoltre che la parola *esperimento*, così come viene utilizzata nelle scienze applicate ed in particolare in fisica, potrebbe indurre nella tentazione di credere che con la locuzione "esperimento aleatorio" si debba necessariamente intendere un esperimento *ripetibile quante volte si voglia*, e capace di produrre, in diverse esecuzioni, risultati diversi. A mettere in guardia contro una tale interpretazione (inutilmente riduttiva), dovrebbe bastare l'esempio seguente.

Esempio 1.1.5 (svuotamento di un'urna) Un individuo disponga di un'urna, contenente palline di due colori diversi: bianco e rosso. Egli conosca il numero totale n delle palline presenti nell'urna, ma non quello delle palline rosse (e dunque neppure quello delle palline bianche). Per conoscerlo egli effettui il banale esperimento che consiste nello svuotare l'urna e nel contare le palline rosse. Prima di compiere un tal esperimento, l'individuo non è in grado di predirne con certezza il risultato: egli è dunque di fronte ad un esperimento aleatorio, al quale potrà associare, come insieme Ω , l'insieme costituito da tutti gli interi compresi tra 0 e n (intendendo che un siffatto intero k rappresenti il risultato descritto dalle parole: "il numero delle palline rosse presenti nell'urna è k "). Per quel che riguarda la "ripetibilità", è del tutto evidente che, una volta compiuto l'esperimento, presa nota del suo risultato, e rimesse nell'urna le palline estratte, l'individuo potrebbe, volendo, ripetere l'esperimento: ma in ogni ripetizione

otterrebbe lo stesso risultato della prima volta, sicché non si troverebbe più dinanzi a un esperimento aleatorio (se non nel caso in cui egli giudicasse possibile aver commesso qualche errore di conteggio).

Tornando al caso generale di un arbitrario esperimento aleatorio, al quale sia stato associato un certo insieme Ω di eventualità, osserviamo subito che, molto spesso, non si è tanto interessati ad un singolo risultato possibile, ma ad un insieme di risultati che si giudicano in quel momento particolarmente “favorevoli” per lo studio che si vuole compiere. In altri termini si è spesso indotti a considerare una parte A di Ω che si può, in quest’ottica, interpretare come una rappresentazione di un certo *evento legato al risultato dell’esperimento*: l’evento, infatti, che si realizza se e soltanto se il risultato dell’esperimento “cade in A ”. In realtà, si può anzi *identificare* questo evento con l’insieme A stesso. Ovviamente, quando si sarà interessati ad uno specifico risultato ω_0 dell’esperimento aleatorio, sarà sufficiente limitarsi a considerare l’evento che si realizza se e soltanto se esce quel determinato risultato, ovvero l’evento $\{\omega_0\}$.

Esempio 1.1.6 (estrazione di una carta da un mazzo) Supponiamo che l’esperimento consista nell’estrazione di una carta da un mazzo contenente 40 carte ben mescolate. In questo caso, si potrà scegliere come Ω un insieme costituito da 40 elementi: per esempio, si potrà scegliere $\Omega = \{1, 2, \dots, 40\}$, con la convenzione che i primi 10 numeri rappresentino le carte a cuori, i secondi 10 le carte a quadri, i terzi le carte a picche e gli ultimi 10 le carte a fiori. Se ci s’interroga circa la possibilità che esca un asso, indipendentemente dal seme, sarà allora naturale rappresentare l’evento descritto dalle parole “esce un asso” come il sottoinsieme A di Ω costituito dai 4 numeri che rappresentano gli assi, per esempio $A = \{1, 11, 21, 31\}$.

Esempio 1.1.7 (uscita di una faccia pari) Riprendendo invece il caso del lancio di un dado (esempio 1.1.1), si può interpretare la parte $\{2, 4, 6\}$ dell’insieme $\Omega = \{1, 2, 3, 4, 5, 6\}$ come rappresentante dell’evento che si realizza se e soltanto se esce uno dei numeri 2, 4, 6, cioè l’evento descritto dalle parole: “uscita di una faccia pari”.

Esempio 1.1.8 (uscita del 18) Nel caso del lotto (esempio 1.1.2), l’evento “uscita del 18” è rappresentato dall’insieme di tutte le cinque ammettenti 18 come elemento.

Esempio 1.1.9 (vittoria di Cocabel) Nel caso delle corse dei cavalli descritto nell’esempio 1.1.3, l’evento “vince il cavallo Cocabel, contrassegnato dal n. 3” è rappresentato dall’insieme di tutti gli ordinamenti che cominciano con il numero 3, ovvero di tutte le permutazioni che scambiano 1 con 3.

1.1.2 La tribù degli eventi, misura del grado di fiducia

Nel caso generale di un arbitrario esperimento aleatorio, abbiamo detto che *ogni* parte A di Ω può essere interpretata come un evento. Tuttavia può darsi che certe parti di Ω corrispondano ad eventi non interessanti (ai fini di un determinato problema) oppure troppo complicati per essere studiati. In ciascun caso, dunque, e per ciascun problema da studiare, converrà scegliere una determinata classe \mathcal{A} (non vuota) di parti di Ω e

riservare il nome di *eventi* agli elementi di questa classe. Solo nei casi particolarmente semplici questa classe potrà coincidere con l'insieme $\mathcal{P}(\Omega)$ di *tutte* le parti di Ω , ma praticamente in tutti i problemi in cui vi è un'infinità più che numerabile di eventualità sarà necessario ridurre la classe degli eventi ad un sottoinsieme proprio di $\mathcal{P}(\Omega)$: questo sottoinsieme sarà comunque enormemente grande e non si correrà mai il rischio, nella pratica, di considerare sottoinsiemi di Ω che non vi appartengano. In ogni caso sarà però opportuno scegliere la famiglia \mathcal{A} in modo tale che essa possieda buone doti di stabilità (rispetto alle comuni operazioni insiemistiche), cioè sia dotata di una struttura algebrica che consenta di fare almeno le più comuni tra le operazioni. Precisamente, sarà conveniente esigere che il complementare (rispetto a Ω) di un qualsiasi elemento di \mathcal{A} sia ancora un elemento di \mathcal{A} ed inoltre che la classe \mathcal{A} sia stabile almeno per le operazioni di unione e di intersezione binaria, nel senso che se A e B sono due elementi di \mathcal{A} , tali siano anche $A \cup B$ e $A \cap B$. In realtà, per questioni di comodità matematica legate principalmente alla possibilità di svolgere operazioni di passaggio al limite, sarà conveniente pretendere che tali doti di stabilità si mantengano per la riunione e l'intersezione di una qualsiasi famiglia numerabile di elementi di \mathcal{A} . Diamo dunque la seguente definizione.

Definizione 1.1.10 Sia Ω un insieme. Una classe \mathcal{A} di parti di Ω si chiama una *tribù* su Ω se è dotata delle tre proprietà seguenti:

1. l'insieme Ω appartiene a \mathcal{A} ;
2. il complementare (rispetto a Ω) di ogni elemento di \mathcal{A} appartiene a \mathcal{A} ;
3. l'unione e l'intersezione di ogni famiglia numerabile di elementi di \mathcal{A} appartiene a \mathcal{A} .

Inoltre, se \mathcal{A} è una tribù su Ω , la coppia (Ω, \mathcal{A}) si chiama uno *spazio probabilizzabile*; nell'ambito di un fissato spazio probabilizzabile (Ω, \mathcal{A}) , l'insieme Ω prende il nome di *insieme delle eventualità*, mentre la tribù \mathcal{A} prende il nome di tribù degli eventi.

Usando il linguaggio appena introdotto, tutte le considerazioni fatte fino a questo momento si possono riassumere nel modo seguente:

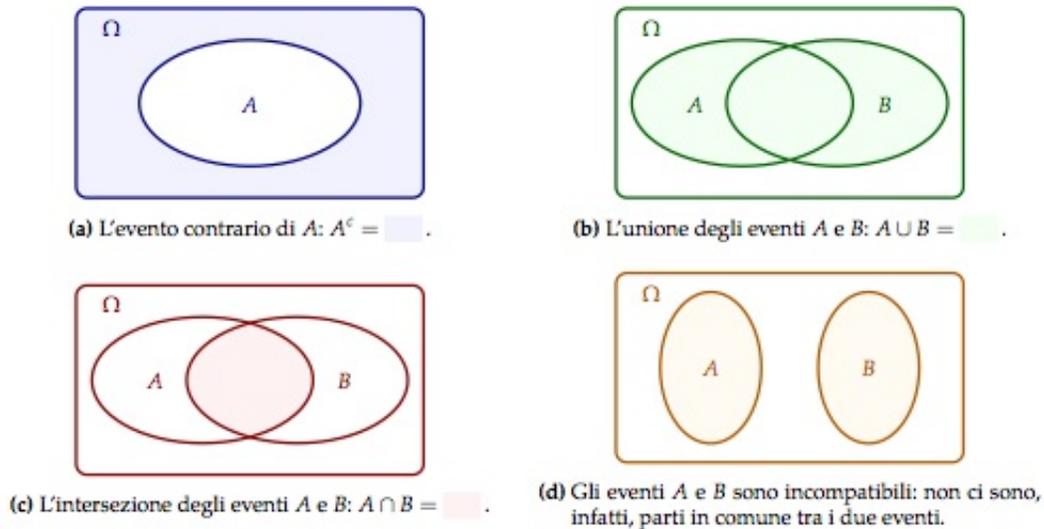
Quando si voglia studiare un esperimento aleatorio, il primo passo da compiere consiste nell'associargli uno spazio probabilizzabile (Ω, \mathcal{A}) .

Quando, per studiare un certo esperimento aleatorio, sia stato scelto un opportuno spazio probabilizzabile (Ω, \mathcal{A}) , si adopera abitualmente una terminologia particolarmente suggestiva che ha lo scopo di mantenere saldo il legame tra il modello matematico e l'esperimento aleatorio. Per questo, non solo si chiamano *eventualità* gli elementi di Ω , ed *eventi* gli elementi di \mathcal{A} , ma si usa anche dire che l'eventualità ω *realizza* l'evento A per dire che ω appartiene ad A . Inoltre:

- se A è un evento, il complementare di A , ossia l'evento A^c che è realizzato da tutte e sole le eventualità che non realizzano l'evento A , viene chiamato la *negazione* di A (o l'evento *contrario* di A);

- se A, B sono due eventi, la loro unione, ossia l'evento $A \cup B$ che è realizzato da tutte e sole le eventualità che realizzano *uno almeno* tra i due eventi A, B , viene chiamato l'evento “ A o B ”;
- se A, B sono due eventi, la loro intersezione, ossia l'evento $A \cap B$ che è realizzato da tutte e sole le eventualità che realizzano *entrambi* gli eventi A, B , viene chiamato l'evento “ A e B ”;
- due eventi A, B si dicono tra loro *incompatibili* se non esiste alcuna eventualità che li realizzi entrambi, cioè se i due insiemi A, B sono tra loro *disgiunti*, o, ciò ch'è lo stesso, se sono privi di elementi in comune. In altre parole, due eventi sono incompatibili quando il realizzarsi dell'uno esclude il realizzarsi dell'altro e viceversa.

Una rappresentazione grafica molto utile per illustrare le relazioni tra gli eventi è quella dei *diagrammi di Venn*. L'insieme delle eventualità Ω è rappresentato da un riquadro più grande all'interno del quale gli eventi sono designati per mezzo di ovali più o meno grandi. In questo modo, le relazioni tra gli eventi possono essere indicate evidenziando all'interno le regioni d'interesse. Per esempio, le principali operazioni tra eventi sono indicate nella figura sottostante.



Esempio 1.1.11 Si consideri l'esperimento consistente nello scegliere un punto a caso su un assegnato segmento. Quale spazio probabilizzabile converrà associare a un siffatto esperimento aleatorio? Usando un'opportuna unità di misura, si potrà rappresentare ciascun punto del segmento con un punto dell'intervallo $[0, 1]$. Come spazio delle eventualità si prenderà dunque l'intervallo $[0, 1]$. Se si ritiene interessante ogni evento rappresentato da un sottointervallo $[a, b]$ di $[0, 1]$ (ossia l'evento indicato dalle parole: “il punto scelto cade tra il punto di ascissa a e quello di ascissa b ”), la tribù degli eventi dovrà contenere la classe di tutti gli intervalli $[a, b]$, con $0 \leq a < b \leq 1$. La più piccola tra tutte le tribù che possiedono questa proprietà si chiama la tribù *boreliana* di $[0, 1]$ e si denota con $\mathcal{B}([0, 1])$.

In uno studio probabilistico di un complesso di eventi legati al risultato di un esperimento aleatorio, la scelta dello spazio probabilizzabile (Ω, \mathcal{A}) costituisce soltanto il primo passo. Un secondo passo consiste nella scelta di una “misura di probabilità”. Che cosa s’intende con ciò?

Per definizione stessa di esperimento aleatorio, l’individuo che considera un tal esperimento non è in grado (salvo casi banali) di stabilire con certezza, per ciascuno degli eventi legati al risultato dell’esperimento, se esso si realizzerà o no. Ciò tuttavia non gli impedisce di sentire, su un piano meramente psicologico, un diverso *grado di fiducia* nei confronti dei diversi eventi considerati. Sarà allora naturale, per quell’individuo, cercare di *misurare* questo grado di fiducia, associando a ciascun evento A della tribù \mathcal{A} un numero $P(A)$, ossia definendo una funzione P sugli elementi della tribù \mathcal{A} . Per convenzione, si può prendere questa funzione a valori in $[0, 1]$, e assumere che essa prenda il valore 1 sull’evento Ω ed il valore 0 sull’evento $\Omega^c = \emptyset$ (l’evento che non si realizza mai). Sarà anche naturale pretendere che essa sia (finitamente) *additiva*, nel senso che verifichi la relazione:

$$P(A \cup B) = P(A) + P(B)$$

ogni volta che A, B siano due eventi incompatibili. In realtà, per ragioni di comodità matematica, così come è stato richiesto che la tribù degli eventi sia stabile per l’operazione di unione (e quindi anche d’intersezione) numerabile, converrà esigere che essa verifichi una proprietà più forte, valida per le famiglie numerabili di eventi. Precisamente, sarà naturale richiedere che

$$P\left(\bigcup_{n=0}^{\infty} A_n\right) = \sum_{n=0}^{\infty} P(A_n)$$

per ogni famiglia numerabile $(A_n)_{n \in \mathbb{N}}$ di elementi di \mathcal{A} a due a due incompatibili. Quest’ultima condizione viene chiamata *additività numerabile* e, come subito si riconosce, essa si riduce all’additività finita quando la tribù \mathcal{A} sia finita e dunque, in particolare, quando Ω sia finito. Tutto ciò premesso, possiamo dare allora la seguente definizione di carattere generale.

Definizione 1.1.12 Fissato uno spazio probabilizzabile (Ω, \mathcal{A}) , una funzione P , definita sulla tribù \mathcal{A} a valori in $[0, 1]$, si chiama una *misura di probabilità* sullo spazio probabilizzabile (Ω, \mathcal{A}) (o, più semplicemente, sulla tribù \mathcal{A}) se

1. $P(\Omega) = 1$
2. P è numerabilmente additiva.

Inoltre, fissata una misura di probabilità P sullo spazio probabilizzabile (Ω, \mathcal{A}) , la terna (Ω, \mathcal{A}, P) è detta uno *spazio probabilizzato*, e, per ogni elemento A di \mathcal{A} , il numero $P(A)$ è detto la *probabilità* dell’evento A secondo P . Sempre per utilizzare un linguaggio più suggestivo e vicino alla probabilità, un evento A , con $P(A) = 0$, si dice anche *trascurabile*, mentre un evento A , con $P(A) = 1$, si dice *quasi certo*.

Per quanto ovvio, giova forse evidenziare che, secondo la precedente definizione, non ha senso parlare di “probabilità di un evento”, se non nell’ambito di un ben precisato spazio

probabilizzato. In particolare, se si è costruito soltanto lo spazio probabilizzabile (Ω, \mathcal{A}) , non ha ancora senso chiedersi quale sia la probabilità di un assegnato evento A (elemento della tribù \mathcal{A}): infatti esistono, in generale, molte misure di probabilità sulla stessa tribù \mathcal{A} , e ciascuna di queste può assegnare ad un evento A una diversa probabilità. Inoltre, ciascuna di queste probabilità è “corretta”, nel senso che ciascun individuo che si trovi a di fronte al medesimo esperimento aleatorio, potrà, secondo le proprie ragioni motivate nel modo che egli ritenga più giusto, assegnare misure di probabilità diverse a seconda delle proprie sensazioni, delle proprie credenze o addirittura delle proprie convinzioni più o meno razionali. Detto in altri termini: la misura di probabilità P , che un individuo dovrà mettere sulla tribù \mathcal{A} , è un oggetto matematico che ha il compito di “fotografare” nel miglior modo possibile la distribuzione della sua fiducia tra i diversi eventi legati al risultato dell’esperimento.

Ciò potrebbe indurre a pensare, a questo punto, che il calcolo delle probabilità sia così soggettivo da essere addirittura inutile, giacché ognuno di noi potrebbe, in linea di principio, stabilire per un dato evento un valore di probabilità diverso da quello scelto da chiunque altro soltanto per sua fede personale. D’altra parte, in molti problemi in cui si presentano esperimenti aleatori si manifestano delle condizioni più o meno “naturali” cui deve soddisfare la misura di probabilità: ebbene, alcuni risultati della teoria del calcolo delle probabilità forniscono, tra l’altro, dei criteri che garantiscono l’esistenza e l’unicità, su un assegnato spazio probabilizzabile (Ω, \mathcal{A}) , di una misura di probabilità che verifichi certe condizioni aggiuntive, come quelle che si presentano nei suddetti problemi aleatori. Dunque la scelta della misura di probabilità, molto spesso, è addirittura scontata se non obbligata dal problema stesso. Resta il fatto, tuttavia, che la scelta del “modello matematico” (Ω, \mathcal{A}, P) è, in ogni caso, un’operazione *pre-matematica*: chiedersi se una certa scelta sia “giusta o sbagliata” non ha dunque senso in nessun caso; o perlomeno, non ha lo stesso senso che chiedersi se siano giusti o sbagliati determinati calcoli eseguiti nell’ambito di un particolare modello scelto.

Non dilunghiamoci ulteriormente sulla pur interessante questione filosofica della scelta della misura di probabilità e ritorniamo piuttosto alla questione generale dell’analisi di un esperimento aleatorio. Usando il linguaggio sopra introdotto, possiamo così ulteriormente riassumere le considerazioni fatte fino a questo momento:

Compito preliminare, per un individuo che intenda studiare dal punto di vista probabilistico un esperimento aleatorio, è quello di associargli uno spazio probabilizzato (Ω, \mathcal{A}, P) .

Possiamo domandarci a questo punto: *che cos’è il calcolo delle probabilità?* Dal punto di vista del matematico puro, la risposta è semplice: è lo studio sistematico di quelle particolari misure che sono le misure di probabilità (dunque soltanto un capitolo della teoria della misura). Fare del calcolo delle probabilità significa perciò, in particolare, occuparsi dei due problemi seguenti:

Problema 1 *Studiare l’insieme di tutte le misure di probabilità che si possono definire su un fissato spazio probabilizzabile (Ω, \mathcal{A}) .*

Problema 2 *Per ogni fissata misura di probabilità P su (Ω, \mathcal{A}) , fornire delle tecniche che aiutino a calcolare esplicitamente la probabilità, secondo P , di certi eventi più o meno complicati.*

È chiaro tuttavia che, per l'utilizzatore che intenda soltanto applicare il calcolo delle probabilità, la situazione è drasticamente diversa. Costui parte di volta in volta da uno specifico problema concreto legato a un determinato esperimento aleatorio. Egli ha dunque davanti a sé i seguenti compiti preliminari:

1. *Fissare un adeguato insieme delle eventualità Ω .*
2. *Decidere quali sono, nell'ambito di questo insieme, gli eventi interessanti (ai fini del problema che si è posto), e scegliere, di conseguenza, la tribù \mathcal{A} degli eventi.*
3. *Scegliere una misura di probabilità P su (Ω, \mathcal{A}) .*

Solo dopo aver compiuto tutte queste operazioni preliminari, egli potrà valersi dei risultati del calcolo delle probabilità: ad esempio, per calcolare esplicitamente, nell'ambito dello spazio probabilizzato (Ω, \mathcal{A}, P) che ha costruito, le probabilità richieste dal problema, ossia le probabilità di certi specifici eventi, in generale piuttosto complicati. Nel seguito, noi introdurremo gli strumenti fondamentali del calcolo delle probabilità; ci concentreremo maggiormente sullo studio della possibile applicabilità delle tecniche proprie del calcolo delle probabilità per la risoluzione di problemi più o meno concreti e più o meno complessi, piuttosto che sullo studio teorico delle strutture algebriche che costituiscono l'impianto teorico del calcolo delle probabilità.

1.1.3 Proprietà generali degli spazi probabilizzati

In questa sezione vogliamo costruire un insieme di regole per il calcolo delle probabilità degli eventi, che siano sufficientemente generali da poter essere utilizzate per qualsiasi esperimento aleatorio. A questo scopo, sia (Ω, \mathcal{A}, P) lo spazio probabilizzato che un certo individuo ha deciso di associare ad un ben determinato esperimento aleatorio. Se A e B sono due eventi (elementi di \mathcal{A}), si può scrivere B come la riunione dei due eventi $A \cap B$ e $A^c \cap B$. Poiché questi sono evidentemente incompatibili, dall'additività della probabilità si trae:

$$P(B) = P(A \cap B) + P(A^c \cap B). \quad (1.1)$$

La relazione precedente è di per sé piuttosto importante: capita frequentemente, infatti, di non saper calcolare direttamente la probabilità di B , ma di saper “spezzare” l'evento B , tramite un evento ausiliario A , in due eventi tra loro incompatibili le cui probabilità sono più semplici da calcolare. Inoltre, questa relazione ha alcune conseguenze importanti. La prima di queste si ottiene ponendo $B = \Omega$. La (1.1) si può allora riscrivere nella forma $P(A) + P(A^c) = P(\Omega) = 1$, dalla quale si deduce:

$$P(A^c) = 1 - P(A). \quad (1.2)$$

D'altra parte, quando invece l'evento A è contenuto in B , la (1.1) si può scrivere nella forma:

$$P(B) = P(A) + P(B \setminus A) \quad (1.3)$$

e di qui, dal fatto che $P(B \setminus A)$ è certamente un numero non negativo, si deduce che

$$\text{se } A \subseteq B, \text{ allora } P(A) \leq P(B).$$

Questa ulteriore proprietà si chiama anche l'*isotonia* della probabilità. Da essa segue, come caso particolare, che ogni evento contenuto in un evento trascurabile è anch'esso trascurabile. Inoltre, se A è contenuto in B , da (1.3) segue immediatamente anche la relazione $P(B \setminus A) = P(B) - P(A)$.

Applichiamo adesso la relazione (1.1) sia ad A che a B , ottenendo così:

$$\begin{aligned} P(A) &= P(A \cap B) + P(A \cap B^c) \\ P(B) &= P(A \cap B) + P(A^c \cap B). \end{aligned}$$

Poiché i tre eventi $A \cap B$, $A^c \cap B$ e $A \cap B^c$ sono a due a due incompatibili e poiché la loro riunione dà l'evento $A \cup B$, sommando le due espressioni membro a membro ed usando l'additività della probabilità, si giunge a

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

Questa espressione, che prende il nome di *modularità*, si può anche riscrivere nella forma

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

e si può pensare come una generalizzazione della proprietà di additività della probabilità nel caso di due eventi tra loro non incompatibili.

Sia adesso $(A_n)_{n \in \mathbb{N}}$ una successione di eventi. Dalla ben nota relazione di De Morgan

$$\left(\bigcup_{n=0}^{\infty} A_n \right)^c = \bigcap_{n=0}^{\infty} A_n^c,$$

e dalla relazione (1.2), si trae:

$$P\left(\bigcup_{n=0}^{\infty} A_n\right) = 1 - P\left(\bigcap_{n=0}^{\infty} A_n^c\right). \quad (1.4)$$

Quest'ultima uguaglianza riesce spesso utile perché, in molte situazioni, la probabilità dell'intersezione di una successione di eventi è più semplice da calcolare, rispetto alla probabilità dell'unione di una successione di eventi. Ma anche quando gli eventi siano soltanto un numero finito (o addirittura soltanto due), questa relazione è spesso utile per il calcolo della probabilità, riducendosi alla forma $P(A \cup B) = 1 - P(A^c \cap B^c)$.

Tutte quante le proprietà viste fino a questo momento sono conseguenze della sola proprietà di additività della probabilità. I due teoremi che seguono si dimostrano invece a partire dall'additività numerabile ed hanno anche lo scopo incidentale di chiarire almeno in piccolissima parte l'importanza di questa proprietà, le cui conseguenze più profonde sono tutt'altro che scontate: essa permette, infatti, una certa "stabilità" per le operazioni di passaggio al limite, almeno per quanto riguarda le successioni monotone di eventi.

Teorema 1.1.13 *Su uno spazio probabilizzato (Ω, \mathcal{A}, P) , sia $(A_n)_{n \in \mathbb{N}}$ una successione crescente di eventi, nel senso che, per ciascun indice n , si ha $A_n \subseteq A_{n+1}$. Posto allora $A = \bigcup_{n=0}^{\infty} A_n$, si ha*

$$\lim_{n \rightarrow \infty} P(A_n) = P(A).$$

Teorema 1.1.14 *Su uno spazio probabilizzato (Ω, \mathcal{A}, P) , sia $(A_n)_{n \in \mathbb{N}}$ una successione decrescente di eventi, nel senso che, per ciascun indice n , si ha $A_{n+1} \subseteq A_n$. Posto allora $A = \bigcap_{n=0}^{\infty} A_n$, si ha,*

$$\lim_{n \rightarrow \infty} P(A_n) = P(A).$$

Come abbiamo detto, i teoremi 1.1.13 e 1.1.14 garantiscono la stabilità della misura di probabilità, nel passaggio al limite sulle successioni “monotone” di eventi. Si osservi infatti che la definizione di monotonia che scaturisce dagli enunciati dei due teoremi è perfettamente coerente con quella già nota per le successioni numeriche. In effetti, se la successione $(A_n)_{n \in \mathbb{N}}$ è crescente (rispettivamente decrescente), allora tale è (nel senso classico) la successione numerica $(P(A_n))_{n \in \mathbb{N}}$. Ora, il significato dei teoremi 1.1.13 e 1.1.14 è il seguente: se è già noto che una successione numerica monotona ha sempre limite, non è affatto detto a priori che questo limite debba coincidere, in generale, con il limite della successione d’insiemi (cioè con la riunione di tutti gli eventi, nel caso delle successioni crescenti o con la loro intersezione nel caso di quelle decrescenti). I teoremi 1.1.13 e 1.1.14 garantiscono invece che le due nozioni di limite (quello numerico delle probabilità e quello insiemistico degli eventi) sono completamente compatibili.

1.1.4 Misure di probabilità su uno spazio discreto

Concentriamo adesso la nostra attenzione soltanto sulle misure di probabilità su uno spazio probabilizzabile il cui insieme delle eventualità sia finito oppure numerabile, cioè relativo ad un esperimento aleatorio che ammetta un numero finito o, tutt’al più numerabile, di possibili risultati. Consideriamo dunque uno spazio probabilizzabile (Ω, \mathcal{A}) tale che l’insieme Ω sia numerabile, e supponiamo che la tribù \mathcal{A} sia costituita da tutte le parti di Ω . Un tale spazio probabilizzabile viene detto *discreto*. Per costruire su di esso una misura di probabilità si può così procedere. Si scelga una qualsiasi funzione positiva f , definita su Ω , a valori in $[0, 1]$ e verificante la relazione

$$\sum_{\omega \in \Omega} f(\omega) = 1.$$

Una tal funzione si chiama una *densità discreta di probabilità* (o, semplicemente, una *densità*) su Ω . Si consideri, poi, l’applicazione P , di \mathcal{A} in $[0, 1]$, che, ad ogni parte A di Ω , associa il numero

$$P(A) = \sum_{\omega \in A} f(\omega). \quad (1.5)$$

È immediato verificare che P è una misura di probabilità che si dice essere *definita dalla densità discreta f* . Viceversa, assegnata una qualsiasi misura di probabilità P sullo spazio probabilizzabile (Ω, \mathcal{A}) , esiste sempre un’unica densità discreta f su Ω ,

tale che la misura di probabilità da essa definita coincida con P ; precisamente, f è la funzione così definita su Ω :

$$f(\omega) = P(\{\omega\}) \quad \text{per ogni elemento } \omega \text{ di } \Omega. \quad (1.6)$$

Chiameremo tale funzione la *densità discreta di P* . Quanto appena stabilito si può così riassumere:

Assegnare una misura di probabilità P su uno spazio probabilizzabile discreto (Ω, \mathcal{A}) equivale ad assegnare su Ω una densità discreta f , essendo i due oggetti tra loro legati tramite le relazioni (1.5) e (1.6).

Supponiamo, in particolare, che l'insieme Ω sia *finito*, e precisamente che sia costituito da n elementi; supponiamo poi che, per la particolare simmetria del problema in esame, si giudichi sensato ritenere che *ogni eventualità sia egualmente probabile*, cioè che ciascun risultato dell'esperimento abbia la stessa probabilità di ogni altro di realizzarsi. In questo caso, sarà naturale scegliere su Ω una densità discreta costante perché, per definizione, essa corrisponde proprio alla probabilità che si realizzi ciascun risultato dell'esperimento. D'altro canto, se si denota con p il valore costante della densità discreta f scelta su Ω , dalla prima proprietà della misura di probabilità discende immediatamente che $P(\Omega) = \sum_{\omega \in \Omega} f(\omega) = np = 1$ e quindi che questo valore costante debba essere $p = 1/n$. La corrispondente misura di probabilità si chiama la *ripartizione uniforme* su Ω e il suo valore $P(A)$ sulla generica parte A di Ω coincide con il rapporto tra il numero di elementi di A (i *casi favorevoli* per il realizzarsi dell'evento A), che indichiamo con $\#(A)$, e il numero totale $\#(\Omega)$ di elementi di Ω (i *casi possibili*):

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

In modo equivalente, si può dire che:

la ripartizione uniforme sull'insieme finito Ω è l'unica misura di probabilità, sulla tribù di tutte le parti di Ω , che attribuisca la stessa probabilità a tutti i *singoletti*, cioè a tutti gli eventi costituiti da un sol elemento (ovvero a tutti i possibili risultati dell'esperimento).

Esempio 1.1.15 Riprendiamo l'esempio del lancio di un dado (esempio 1.1.1). La ripartizione uniforme sull'insieme $\Omega = \{1, 2, 3, 4, 5, 6\}$ appare come la più naturale, tra tutte le possibili misure di probabilità su $\mathcal{P}(\Omega)$, in quanto è l'unica che assegna la stessa probabilità a tutti i singoletti. Sarà questa, infatti, la scelta più naturale se si ritiene che il dado non sia truccato e che, dunque, l'uscita di ciascuna delle facce sia *equiprobabile*.

Per quanto ovvio, giova forse sottolineare ancora una volta che, affinché si possa utilizzare la ripartizione uniforme sull'insieme Ω che è stato scelto per studiare un determinato esperimento aleatorio, è necessario che siano soddisfatti due prerequisiti essenziali: innanzitutto che Ω sia finito, ovvero che vi siano un numero finito di possibili risultati dell'esperimento, e che inoltre ciascuno di questi risultati abbia la stessa probabilità degli altri di presentarsi.

1.1.5 Elementi di calcolo combinatorio

Ogni qual volta si scelga come misura di probabilità su uno spazio probabilizzabile finito la ripartizione uniforme, il calcolo delle probabilità degli eventi in esame si riduce al calcolo del numero di elementi di quegli eventi, ossia della loro cardinalità. Questo potrebbe sembrare, a prima vista, come una cosa piuttosto elementare (per esempio, è immediato riconoscere che l'insieme $\{1, 2, 3\}$ contiene soltanto 3 elementi), ma quando gli insiemi sono più complicati oppure molto numerosi, questo calcolo può diventare anche molto difficile. Giusto per dare un'idea della complessità di questo calcolo, si osservi l'insieme descritto nella figura qui sotto.



Come si vede, per quanto i punti siano disposti in maniera ordinata, non è così immediato dire quanti essi siano. Si può comunque agilmente contarli osservando che ve ne sono 12 in ogni riga e che vi sono 4 righe, per un totale di 48 punti. Se, tuttavia, i punti fossero disordinati, l'unica alternativa rimarrebbe quella di contarli uno per uno, con un procedimento inefficiente e piuttosto lungo, come si potrà subito verificare, provando a contare quanti sono i punti presenti nella figura sottostante.



Perdendo un po' di tempo a svolgere effettivamente il conto, si scopre che anche questi sono 48, cioè tanti quanti quelli della figura precedente, ma il fatto di essere disposti in maniera disordinata rende la verifica molto più lunga e soggetta ad errori di conto. Nel resto di questo paragrafo ci occuperemo dunque di stabilire alcuni metodi per calco-

lare la *cardinalità*, ossia il numero totale degli elementi, di insiemi più o meno complessi, utilizzando ragionamenti elementari che permettono di scomporre il problema in calcoli semplici e facilmente risolvibili. Questo ci permetterà, in particolare, di calcolare più agilmente le cardinalità degli eventi di cui ricerchiamo la probabilità.

In via del tutto generale, infatti, il *calcolo combinatorio* è quella disciplina che si occupa di studiare i modi in cui si combinano gli elementi di un insieme finito scegliendo tutti o alcuni degli elementi a disposizione, calcolando così il numero di modi possibili in cui essi si possono raggruppare. Se si pensa di raccogliere questi elementi all'interno di un insieme, dunque, si può parafrasare quanto appena detto, dicendo molto più semplicemente che il calcolo combinatorio si occupa del calcolo della cardinalità degli insiemi finiti.

A questo scopo, una volta per tutte, indichiamo con A un insieme finito di cardinalità k e con B un altro insieme finito di cardinalità n .

Raggruppamenti. Un *raggruppamento* è un insieme formato prendendo un elemento dai k di A ed un elemento dagli n di B . Esso corrisponde dunque al prodotto cartesiano $A \times B$, ossia dall'insieme costituito da tutte le coppie ordinate (a, b) con $a \in A$ e $b \in B$. Come subito si riconosce, la cardinalità di $A \times B$ è kn . Se fissiamo, infatti, un elemento a di A , il numero di possibili coppie distinte (a, b) è n , cioè una per ogni elemento b di B . Poiché, però, l'elemento a di A si può scegliere in k modi, il numero totale delle coppie ordinate è appunto kn .

Esempio 1.1.16 Se un individuo ha 3 camicie e 4 maglioni, in quanti modi egli si può vestire? Se indichiamo con $A = \{C_1, C_2, C_3\}$ l'insieme delle camicie e con $B = \{M_1, M_2, M_3, M_4\}$ l'insieme dei maglioni, si riconosce immediatamente che l'individuo in questione si può vestire combinando una camicia C_i tra le 3 disponibili, con un maglione M_j tra i 4 disponibili, ossia che i possibili modi di vestire sono tutti e soli racchiusi nell'insieme $A \times B$. Di conseguenza, egli avrà

$$3 \text{ camicie} \times 4 \text{ maglioni} = 12 \text{ modi di vestire.}$$

Disposizioni semplici. Supponiamo adesso che sia $k \leq n$. Si dice che si ha una *disposizione semplice* di n elementi e di classe k , quando si vogliono scegliere, *con ordine e senza ripetizione*, k elementi da un insieme di n . In altri termini, nel linguaggio della matematica, una disposizione semplice è il numero delle funzioni iniettive f dell'insieme A nell'insieme B , essendo l'insieme B , composto da n elementi, la totalità degli oggetti che si possono scegliere, l'immagine $f(A)$, composta da k elementi, l'insieme degli oggetti scelti, ed A l'insieme dei k posti vuoti nei quali disporre gli elementi di B che sono stati scelti.

Se indichiamo dunque con $D_{n,k}$ la cardinalità dell'insieme delle funzioni iniettive di A in B , per calcolare questo numero possiamo ragionare come segue. Per il primo elemento di A , si possono scegliere n elementi (tutti quelli di B); per il secondo elemento di A si possono scegliere $n - 1$ elementi di B , perché uno degli elementi di B è già stato preso; per il terzo elemento di A , si possono scegliere $n - 2$ elementi di B , perché stavolta ne sono stati già presi 2, e così via, fino al k -esimo elemento di A , per il quale sono disponibili $n - k + 1$ elementi di B . In totale, dunque, il numero di disposizioni semplici

di n elementi di classe k è

$$D_{n,k} = n \cdot (n-1) \cdot (n-2) \cdots (n-k+1). \quad (1.7)$$

Ricordiamo adesso che si chiama *fattoriale* di un numero intero n e si indica con $n!$ il prodotto dei numeri interi tra 1 e n , cioè il numero:

$$n! = n \cdot (n-1) \cdot (n-2) \cdots 3 \cdot 2 \cdot 1.$$

Ovviamente, affinché questa formula abbia senso, vi devono essere almeno due fattori, sicché questa definizione vale per $n \geq 2$. Tuttavia, per convenzione, si pone anche $0! = 1$ e $1! = 1$ cosicché si possa anche scrivere convenientemente la formula

$$n! = n \cdot (n-1)! \quad \text{per ogni } n \geq 1.$$

Usando la notazione fattoriale, la formula delle disposizioni semplici potrà anche essere scritta più facilmente nella forma

$$D_{n,k} = \frac{n!}{(n-k)!}.$$

Tuttavia questa formula, più comoda formalmente, è assai complessa da utilizzare nella pratica, comportando, in generale, calcoli che coinvolgono numeri molto grandi. Sarà dunque conveniente, nelle applicazioni, utilizzare sempre la formula (1.7).

Esempio 1.1.17 Ad una gara di salto ad ostacoli partecipano 6 cavalli. Verranno tuttavia premiati soltanto i primi 3, rispettivamente con medaglia d'oro, d'argento e di bronzo. Quante sono le possibili classifiche dei vincitori?

Per rispondere a questo problema, osserviamo che il primo classificato si potrà scegliere tra 6 cavalli e, per ciascuna scelta, si potrà scegliere il secondo classificato tra ciascuno dei 5 cavalli rimanenti e così, per il terzo posto, si potrà scegliere tra ciascuno dei 4 cavalli rimasti. In definitiva, si avranno

$$D_{6,3} = 6 \cdot 5 \cdot 4 = 120$$

possibilità per il podio finale.

Disposizioni con ripetizione. Si dice che si ha una *disposizione con ripetizione* di n elementi e di classe k , quando si vogliono scegliere, *con ordine* e *con ripetizione*, k elementi da un insieme di n elementi. Parafrasando anche questa affermazione nel linguaggio della matematica, una disposizione con ripetizione è il numero di tutte le funzioni f di A in B e, con un ragionamento analogo a quello fatto per le disposizioni semplici, si ottiene che il suo numero è

$$D_{n,k}^{(r)} = n^k,$$

giacché per ogni elemento di A si possono ogni volta associare n elementi di B .

Esempio 1.1.18 Calcoliamo il numero di parole (anche senza senso) che si possono comporre utilizzando soltanto tre vocali (ad esempio **aia**, **aeo**, **uoo**, **iie** ecc.). A questo scopo, denotiamo con B l'insieme delle cinque vocali, cioè $B = \{a, e, i, o, u\}$, ed osserviamo che ogni parola di tre lettere composta dalle sole vocali si può vedere come una funzione dall'insieme $A = \{1, 2, 3\}$ nell'insieme B e precisamente quella funzione che ad ogni numero i di A associa la lettera che si trova nel posto i -esimo. Dunque, il numero delle parole di tre lettere che si possono formare con le vocali è $5^3 = 125$.

Esempio 1.1.19 Calcolare il numero dei possibili PIN di un bancomat. Poiché il PIN è formato da 5 cifre fra 0 e 9, esso è una cinquina di elementi di $B = \{0, 1, \dots, 9\}$, ossia una funzione di $A = \{1, 2, 3, 4, 5\}$ in B . Il numero di possibili PIN è allora $10^5 = 100000$.

Permutazioni. Le *permutazioni semplici* sono le disposizioni semplici nelle quali k coincide con n , cioè in cui vengono presi tutti gli n elementi dell'insieme B . In altri termini, le permutazioni semplici si hanno quando le cardinalità di A e di B coincidono. Il loro numero si denota anche con P_n ed, evidentemente, si ha $P_n = D_{n,n} = n!$.

Esempio 1.1.20 Quanti sono gli anagrammi della parola **ciao**? Essendo questa parola costituita da 4 lettere che possono essere scambiate tra loro in tutti i modi possibili, si tratta di una permutazione semplice di classe 4 e quindi sono possibili

$$P_4 = 4! = 4 \cdot 3 \cdot 2 \cdot 1 = 24$$

anagrammi.

Spesso capita che, nell'effettuare una permutazione, alcuni degli elementi siano ripetuti, cosicché il loro scambio non comporta un vero cambiamento, lasciando inalterato, nel complesso, l'insieme degli elementi da scambiare. In questo caso si dice che si effettua una *permutazione con ripetizione*: in tal caso, per ciascun elemento ripetuto, occorrerà dividere per il fattoriale del numero di volte in cui esso compare, cioè per il numero delle possibili permutazioni che si possono fare con i soli elementi identici.

Esempio 1.1.21 Quanti sono gli anagrammi della parola **cavalla**? Osserviamo che le lettere **a** e **l** sono ripetute, la prima capitando 3 volte e la seconda 2 volte. Se dunque la parola **cavalla** è composta in totale da 7 lettere e $7!$ sarà il numero di permutazioni di queste lettere, va tuttavia osservato che, siccome la **a** e la **l** sono ripetute, ogni permutazione che scambi di posto fra loro le tre **a**, oppure le due **l**, lascerà invariato l'anagramma. Pertanto sarà necessario dividere per $3!$ (numero delle permutazioni della **a**) e per $2!$ (numero delle permutazioni della **l**), ottenendo così il corretto numero

$$\frac{7!}{3! \cdot 2!}$$

di permutazioni della parola **cavalla**.

Combinazioni semplici. Supponiamo adesso che sia $k \leq n$ e contiamo il numero dei sottoinsiemi U di B aventi esattamente k elementi. A questo scopo, indichiamo con m il numero che cerchiamo (di tutti i sottoinsiemi U di B con k elementi), che chiameremo

una *combinazione semplice* di n elementi presi a k a k . Detto A l'insieme $\{1, 2, \dots, k\}$, osserviamo che ogni funzione iniettiva $f : A \rightarrow B$ ha come immagine uno di tali insiemi U . Viceversa, ogni insieme U è immagine di qualche applicazione iniettiva $f : A \rightarrow B$. Fissato l'insieme U , contiamo quante sono le funzioni iniettive f con $U = f(A)$, dunque bigettive da A in U : esse sono esattamente $k!$, perché ogni permutazione di A individua una diversa funzione bigettiva da A ad U . Variando l'immagine U , il numero complessivo delle funzioni iniettive di A in B è dunque pari a $m \cdot k!$; ma tale numero è anche uguale, come sappiamo dalla (1.7), a $n(n-1) \cdots (n-k+1)$; da ciò si trae immediatamente

$$m = \frac{n(n-1) \cdots (n-k+1)}{k!} = \frac{n!}{k!(n-k)!},$$

quantità che si denoterà con $\binom{n}{k}$.

È comodo attribuire ai numeri $k!$ e $\binom{n}{k}$ un significato anche quando $k = 0$. Nel primo caso abbiamo già posto $0! = 1$ e così anche nel secondo caso poniamo $\binom{n}{0} = 1$. Quest'ultima eguaglianza si può spiegare semplicemente dicendo che ogni insieme ammette un unico sottoinsieme privo di elementi, l'insieme vuoto \emptyset .

Il numero $\binom{n}{k}$ si chiama *coefficiente binomiale* ed i sottoinsiemi di k elementi di un insieme costituito da n elementi si chiamano le *combinazioni semplici* di n oggetti presi a k a k . Si hanno le seguenti semplicissime eguaglianze:

$$\binom{n}{0} = \binom{n}{n} = 1, \quad \binom{n}{1} = \binom{n}{n-1} = n, \quad \binom{n}{2} = \binom{n}{n-2} = \frac{n(n-1)}{2}$$

e, più in generale,

$$\binom{n}{k} = \binom{n}{n-k},$$

perché, nella formula che definisce $\binom{n}{k}$, sostituire k con $n-k$ equivale soltanto a scambiare i due fattori che compaiono al denominatore. Vale inoltre la seguente formula

$$\binom{n}{k} = \binom{n-1}{k} + \binom{n-1}{k-1} \tag{1.8}$$

che si potrebbe pure dimostrare direttamente dalla formula che definisce il coefficiente binomiale, ma che è certamente più istruttiva se dedotta dallo studio della cardinalità dei sottoinsiemi di un insieme $B = \{b_1, \dots, b_n\}$ costituito da n elementi. Fissiamo un elemento b_1 in B ed osserviamo che i sottoinsiemi di B formati da k elementi si ripartiscono in due classi: la prima, costituita da quelli che non contengono b_1 ed il cui numero è lo stesso di quello dei sottoinsiemi di $\{b_2, \dots, b_n\}$ formati da k elementi, ossia $\binom{n-1}{k}$; la seconda, formata da quelli che contengono b_1 ed il cui numero è lo stesso di quello dei sottoinsiemi di $\{b_2, \dots, b_n\}$ costituiti da $k-1$ elementi, ossia $\binom{n-1}{k-1}$. In totale, avremo la desiderata formula (1.8).

Esempio 1.1.22 Da un gruppo di 9 studenti se ne vogliono scegliere 4. In quanti modi possibili si può fare questa scelta? Evidentemente, si tratta di contare il numero

di sottoinsiemi di 4 elementi formati da quelli di un insieme con 9 elementi. Come abbiamo detto questa è una combinazione semplice e dunque si possono realizzare

$$\binom{9}{4} = \frac{9 \cdot 8 \cdot 7 \cdot 6}{4!} = 126$$

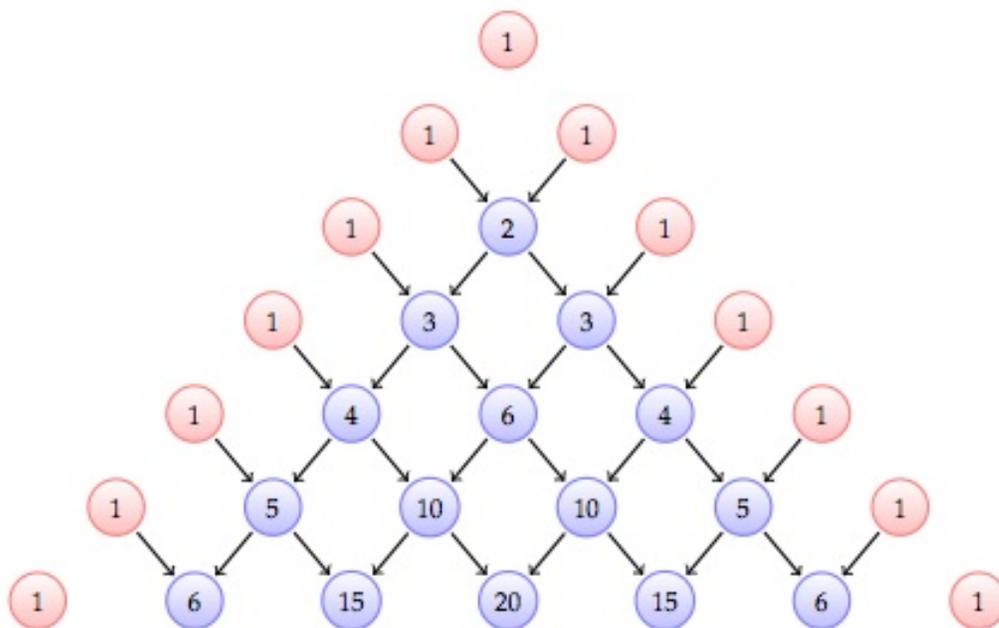
gruppi.

Esempio 1.1.23 Dato un insieme formato da 12 persone, nel quale 8 sono maschi e 4 sono femmine, in quanti modi si possono fare gruppi di 4 persone, prendendo dal gruppo 2 maschi e 2 femmine? Per rispondere a questo problema, osserviamo che gli uomini si possono prendere in $\binom{8}{2}$ modi, mentre le donne in $\binom{4}{2}$ modi. Quindi, raccogliendo questi due sottogruppi, si trova che si possono formare

$$\binom{8}{2} \cdot \binom{4}{2} = \frac{8 \cdot 7}{2} \cdot \frac{4 \cdot 3}{2} = 168$$

gruppi costituiti da 2 maschi e 2 femmine.

Tornando alla formula (1.8), osserviamo che essa permette di calcolare molto velocemente i coefficienti binomiali utilizzando il cosiddetto *triangolo di Tartaglia* (NICCOLÒ FONTANA detto TARTAGLIA, 1499–1557) o, secondo la tradizione francese, *triangolo di Pascal* (BLAISE PASCAL, 1623–1662) nel quale $\binom{n}{k}$ occupa la $(k + 1)$ -esima posizione della riga $(n + 1)$ -esima. Esso si ottiene semplicemente sommando i due numeri che lo sovrastano (nella figura sottostante compaiono le prime 7 righe).



Da ultimo, osserviamo che i coefficienti binomiali si possono utilizzare, in particolare, per sviluppare le potenze di un binomio e di qui traggono, in effetti, il loro nome. Precisamente, dati due numeri reali x, y vogliamo valutare $(x + y)^n$ al variare di $n \in \mathbb{N}$.

Il risultato di questo calcolo si chiama il *teorema del binomio* ed è dovuto a Newton. Innanzitutto osserviamo che sviluppando $(x+y)^n$ si ottengono tanti addendi il cui grado totale in x e y è sempre n . Per capire come si possano ottenere tutti questi addendi, scriviamo per esteso la potenza del binomio:

$$(x+y)^n = (x+y)(x+y)\cdots(x+y).$$

Si tratta quindi di scegliere da ognuno degli n fattori $x+y$ uno tra x e y , al variare di questa scelta si otterranno tutti i possibili addendi. Per prima cosa osserviamo che il risultato sarà simmetrico in x e y (questo segue in particolare dalla proprietà commutativa dell'addizione). Ora, per trovare quanti addendi della forma $x^k y^{n-k}$ ci sono, si tratta di scegliere k volte x e le restanti volte y . Per calcolare in quanti modi si possa scegliere k volte x , si tratta di valutare in quanti modi si possano scegliere k fattori su n , cioè quanti sottoinsiemi di k elementi ci sono in un insieme di n elementi. Ma questo numero, com'è noto, è $\binom{n}{k}$. Dunque:

$$(x+y)^n = \sum_{k=0}^n \binom{n}{k} x^k y^{n-k}. \quad (1.9)$$

Osserviamo che questo risultato dipende soltanto dal fatto che, nell'insieme numerico nel quale ci siamo messi, vale la proprietà commutativa. Dunque questo risultato si potrà estendere a qualsiasi altro insieme numerico soddisfacente alle medesime proprietà dell'insieme costituito da tutti i numeri reali.

Concludiamo questo paragrafo contando il numero di sottoinsiemi di un insieme finito X costituito da n elementi, cioè la cardinalità dell'insieme $\mathcal{P}(X)$ formato da tutti i sottoinsiemi di X . Per fare questo calcolo, iniziamo con l'introdurre l'insieme $Y = \{0, 1\}$ e consideriamo le funzioni di X in Y . Per ciascuna siffatta funzione $f: X \rightarrow Y$ l'immagine inversa $f^{-1}(1)$ è un sottoinsieme di X . Viceversa, ciascun sottoinsieme U di X si può vedere come l'immagine inversa di una ed una sola funzione $f: X \rightarrow Y$ e precisamente quella definita da $f(x) = 1$ per ogni $x \in U$ e $f(x) = 0$ per ogni $x \notin U$. Il numero di sottoinsiemi di X coincide dunque con il numero delle funzioni f di X in Y e dunque è pari a $\#(\mathcal{P}(X)) = 2^n$.

Un modo alternativo per contare il numero dei sottoinsiemi di X consiste nell'utilizzare la formula del binomio (1.9): infatti, il numero di sottoinsiemi sarà la somma di tutti quelli costituiti da k elementi, con k compreso tra 0 ed n , cioè dalla somma:

$$\sum_{k=0}^n \binom{n}{k}.$$

Prendendo $x = y = 1$ nella (1.9) si ottiene allora immediatamente la relazione:

$$\sum_{k=0}^n \binom{n}{k} = 2^n.$$

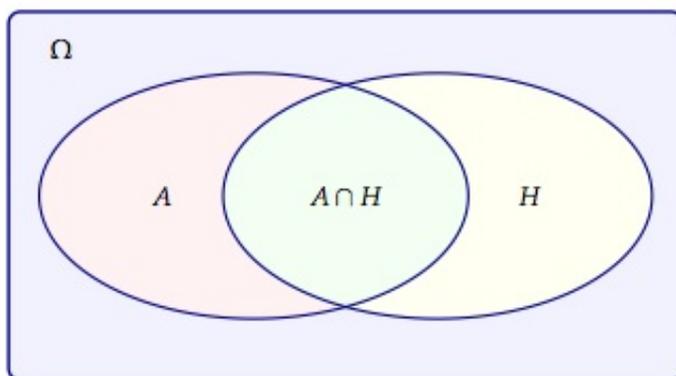
1.1.6 Probabilità condizionale, formula di Bayes

Molto spesso, nel corso di un esperimento aleatorio, può capitare che non si conosca l'esito dell'esperimento, ma che si sia portati a sapere (o a sperare) che l'esito dell'esperimento cada in un insieme noto di eventualità. In questo caso, si sarà portati a modificare la misura di probabilità scelta, per adattarla a questa nuova informazione. A questo scopo, sia (Ω, \mathcal{A}, P) uno spazio probabilizzato. Fissato un evento non trascurabile H (elemento di \mathcal{A}), si chiama *misura di probabilità dedotta da P sotto la condizione H* la misura di probabilità P_H così definita nella tribù degli eventi \mathcal{A} :

$$P_H(A) = \frac{P(A \cap H)}{P(H)}. \quad (1.10)$$

Per ogni evento A , la probabilità di A secondo P_H , ossia il numero $P_H(A)$ sopra definito, si chiama la *probabilità condizionale* di A , secondo P , *sotto la condizione H* , e si denota anche con $P(A|H)$. Si badi di non confondere P_H (che è una misura di probabilità sull'insieme delle parti di Ω) con $P_H(A)$ o con $P(H)$ (che sono due numeri).

Cerchiamo di capire perché questa nuova misura di probabilità è effettivamente quella che si desiderava costruire. Se (Ω, \mathcal{A}, P) è lo spazio probabilizzato che un certo individuo (in un determinato stato d'informazione) ha deciso di associare ad un esperimento aleatorio, allora, per ogni parte non trascurabile H di Ω , lo spazio probabilizzato $(\Omega, \mathcal{A}, P_H)$ è il nuovo spazio che l'individuo è naturalmente indotto a scegliere, in sostituzione del precedente, qualora egli riceva (e accetti per buona) la seguente informazione supplementare: “l'evento H si è realizzato” (ossia “il risultato dell'esperimento cade in H ”). Più precisamente: la scelta consistente nel sostituire (Ω, \mathcal{A}, P) con $(\Omega, \mathcal{A}, P_H)$ è la più naturale che l'individuo possa compiere qualora egli intenda aggiornare le proprie opinioni alla luce della nuova informazione, ma senza modificare l'insieme delle eventualità. Infatti P_H è l'unica misura di probabilità sull'insieme delle parti di Ω che prenda, sul generico evento A , un valore proporzionale alla probabilità, secondo la vecchia misura P , dell'insieme $A \cap H$ (che è, nel nuovo stato d'informazione, “la parte di A che conta”).



Utilizziamo subito questa nuova definizione per costruire una importantissima formula per il calcolo delle probabilità di un evento. A questo scopo, su un assegnato spazio probabilizzato (Ω, \mathcal{A}, P) , sia A un evento (cioè un elemento di \mathcal{A}), e sia \mathcal{H} un insieme *finito* (o *numerabile*) di eventi a due a due incompatibili, la riunione dei quali coincida con Ω . Si potrà pensare, intuitivamente, che \mathcal{H} sia un insieme di ipotetiche “cause”

a due a due incompatibili per il realizzarsi di un determinato “effetto” rappresentato dall’evento A . Si riconosce allora immediatamente che gli eventi della forma $A \cap H$, con $H \in \mathcal{H}$, sono a due a due incompatibili, e dunque, per l’additività della misura di probabilità, si ha:

$$P(A) = \sum_{H \in \mathcal{H}} P(A \cap H). \quad (1.11)$$

Se poi si suppone che ciascuno degli elementi H di \mathcal{H} non sia trascurabile, allora è possibile, al secondo membro della relazione precedente, moltiplicare e dividere ciascun termine della somma per $P(H)$. Si trova così:

$$P(A) = \sum_{H \in \mathcal{H}} P(H)P(A|H). \quad (1.12)$$

Questa formula è detta la *formula della disintegrazione*. Poiché la somma di tutte le probabilità $P(H)$, con $H \in \mathcal{H}$, è $P(\Omega) = 1$, essa esprime la probabilità di A secondo P come la *media ponderata* delle probabilità condizionali $P(A|H)$, con $H \in \mathcal{H}$: ciascuna di esse interviene nella media con il peso $P(H)$.

Nell’ambito di un fissato spazio probabilizzato (Ω, \mathcal{A}, P) , la probabilità condizionale e la formula della disintegrazione possono apparire inutili e artefatte, giacché la conoscenza della misura di probabilità P permette di calcolare la probabilità $P(A)$ di ogni evento A della tribù \mathcal{A} senza bisogno di aggiungere altro. In realtà, però, le cose sono un po’ più complicate di come possono apparire sul piano teorico: nei problemi pratici di calcolo delle probabilità, infatti, può capitare che la funzione P non si conosca mai completamente (questo, a dire il vero, capita quasi di continuo) e, anche nel caso in cui vi sia una formula esplicita per questa funzione, talvolta, l’evento A è troppo complicato per fare un calcolo diretto. D’altra parte, nelle applicazioni pratiche, molto spesso, è noto un insieme \mathcal{H} di possibili *cause* ed è nota la probabilità $P(H)$ di ciascuna di queste cause H in \mathcal{H} . Questo basta, in virtù dei teoremi generali del calcolo delle probabilità, a garantire l’esistenza e l’unicità di una misura di probabilità P pur senza conoscerla completamente. Inoltre, dato un evento A , è spesso nota (o comunque più facile da calcolare) la probabilità $P(A|H)$, che rappresenta la probabilità che A si realizzi sapendo che H si è realizzato, cioè sapendo che H lo ha causato.

Tuttavia, quando ci si trova in una situazione come quella che stiamo descrivendo, si è interessati non soltanto a stabilire qual è la probabilità che l’evento A si realizzi (cosa a cui può agilmente rispondere la formula della disintegrazione), ma anche e soprattutto a stabilire quale delle ipotesi H in \mathcal{H} possa averlo causato. A questo scopo, supponiamo che si sia realizzato un particolare evento A e di voler stabilire quale possa essere la causa H che lo abbia scatenato. La prima osservazione che si potrebbe fare è che, se vi è una causa K più probabile di ogni altra (cioè con $P(K) \geq P(H)$ per ogni $H \in \mathcal{H}$), si sarebbe portati a pensare che essa sia probabilmente la sua vera causa. D’altro canto, si supponga anche che la probabilità $P(A|K)$ sia molto piccola, se non addirittura nulla. È chiaro allora che, per quanto si sia portati a fidarsi che K sia più probabile delle altre cause, difficilmente saremo portati a credere che essa si sia realizzata proprio in questo caso, perché non ci fidiamo affatto che K possa aver davvero causato A .

Occorrerà dunque vedere quanto valgono le probabilità $P(H|A)$, che in questo contesto vengono chiamate le *probabilità a posteriori*, degli elementi H di \mathcal{H} . Esse, infatti,

rappresentano la probabilità che la causa H si sia realizzata, sapendo che l'effetto A si è realizzato. Anche in questo caso, però, nella pratica saranno più facilmente note le probabilità $P(H)$ e $P(A|H)$ e dunque sarà comodo legare a queste ultime l'espressione di ciascuna delle probabilità a posteriori $P(H|A)$. Per far questo, dalla formula (1.10), si ricava immediatamente, per ogni evento non trascurabile A e ogni elemento di \mathcal{H} , la seguente espressione che prende il nome di *formula di Bayes*:

$$P(K|A) = \frac{P(K)P(A|K)}{P(A)} = \frac{P(K)P(A|K)}{\sum_{H \in \mathcal{H}} P(H)P(A|H)}.$$

Si potrà allora dire che l'ipotesi K ha causato A , almeno con probabilità $P(K|A)$.

Esempio 1.1.24 Una popolazione è composta al 40% da fumatori e al 60% da non fumatori. È noto che il 25% dei fumatori ed il 7% dei non fumatori sono affetti da una forma di malattia respiratoria cronica. Qual è la probabilità che, scelto a caso un individuo dalla popolazione, egli sia affetto dalla malattia?

Supponiamo di aver costruito uno spazio probabilizzato (Ω, \mathcal{A}, P) in grado di modellizzare questo problema. Un siffatto spazio probabilizzato dovrà certamente contenere tre eventi: H, K, A , il primo dei quali rappresenta l'evento che si realizza se e soltanto se si è scelto un individuo fumatore, il secondo dei quali rappresenta l'evento che si realizza se e soltanto se si è scelto un individuo non fumatore e il terzo dei quali che rappresenta l'evento che si realizza se e soltanto se si è scelto un individuo affetto dalla malattia. Dovrà poi essere:

$$\begin{aligned} P(H) &= 0.4, & P(K) &= 0.6, \\ P(A|H) &= 0.25, & P(A|K) &= 0.07. \end{aligned}$$

Inoltre i due eventi H, K sono tra loro incompatibili e la loro riunione coincide con Ω . È quindi possibile calcolare la probabilità di A , utilizzando la formula della disintegrazione di A rispetto alla famiglia finita $\{H, K\}$. Si ha così:

$$P(A) = P(H)P(A|H) + P(K)P(A|K) = 0.142.$$

Calcoliamo adesso la probabilità che una persona affetta dalla malattia sia un fumatore. Basterà per questo calcolare la probabilità $P(H|A)$. Utilizziamo a questo scopo la formula di Bayes:

$$P(H|A) = \frac{P(H)P(A|H)}{P(A)} = 0.704.$$

In altre parole, possiamo affermare che se un individuo estratto da questa popolazione è affetto dalla patologia respiratoria, allora al 70% di probabilità egli sarà un fumatore.

Esempio 1.1.25 Tre mobili tra loro indistinguibili contengono ciascuno due cassetti. Il primo contiene una moneta d'oro in ciascuno dei due cassetti, il secondo una moneta d'oro nel primo cassetto ed una moneta d'argento nel secondo, il terzo una moneta d'argento in ciascuno dei due. Si apre un cassetto a caso e si trova una moneta d'oro. Qual è la probabilità che anche l'altro cassetto dello stesso mobile contenga una moneta d'oro?

Per risolvere questo problema, consideriamo uno spazio probabilizzato (Ω, \mathcal{A}, P) il quale contenga quattro eventi A_1, A_2, A_3, B , che si potranno interpretare nel modo seguente: l'evento A_i ($i = 1, 2, 3$) è l'evento che si realizza se e soltanto se è stato aperto un cassetto del mobile i -esimo; l'evento B è invece quello che si realizza se e soltanto se la moneta estratta dal cassetto prescelto è d'oro. Si ha allora, tenuto conto che i mobili sono stati scelti completamente a caso:

$$\begin{aligned} P(B | A_1) &= 1, & P(B | A_2) &= 1/2, & P(B | A_3) &= 0, \\ P(A_1) &= P(A_2) = P(A_3) &= 1/3. \end{aligned}$$

La formula della disintegrazione fornisce immediatamente la probabilità dell'evento B :

$$P(B) = P(A_1)P(B | A_1) + P(A_2)P(B | A_2) + P(A_3)P(B | A_3) = 1/2.$$

Per rispondere alla domanda, è sufficiente calcolare la probabilità $P(A_1 | B)$. Per questo, basta utilizzare la formula di Bayes:

$$P(A_1 | B) = \frac{P(A_1)P(B | A_1)}{P(B)} = \frac{2}{3};$$

risultato, questo, che è probabilmente diverso da quello che ci viene suggerito dall'intuizione, che ci porterebbe erroneamente a pensare che la probabilità sia $1/2$, ritenendo che, avendo escluso che si tratti del terzo mobile, tale probabilità sia la medesima che quella di scegliere a caso tra il primo e il secondo mobile. In realtà, poiché il primo mobile ha a disposizione due monete d'oro, contro una sola disponibile per il secondo, è più probabile che sia stato scelto il primo mobile che non il secondo.

Esempio 1.1.26 Ad un certo stadio delle indagini su un crimine, l'ispettore Clouseau è convinto al 60% della colpevolezza di un certo sospettato. Supponiamo che si scopra un nuovo indizio che mostra che il colpevole deve possedere una determinata caratteristica distintiva (per esempio, l'aver i capelli di un certo colore, oppure essere mancino ecc.) e supponiamo che il sospettato dell'ispettore Clouseau lo possieda. Se tale particolarità interessa il 20% della popolazione, quanto sicuro dev'essere l'ispettore Clouseau della colpevolezza del sospettato?

Per risolvere il problema, supponiamo assegnato uno spazio probabilizzato (Ω, \mathcal{A}, P) e, su di esso, gli eventi H e A descritti rispettivamente dalle parole "il sospettato è colpevole" e "il sospettato possiede il tratto distintivo del colpevole". Si ha dunque:

$$P(H) = 0.6, \quad P(A | H) = 1, \quad P(A | H^c) = 0.2.$$

La formula della disintegrazione ci dà subito

$$P(A) = P(H)P(A | H) + P(H^c)P(A | H^c) = 0.68.$$

Per la formula di Bayes, ora, si ha

$$P(H | A) = \frac{P(H)P(A | H)}{P(A)} \approx 0.882.$$

Dunque la fiducia dell'ispettore Clouseau circa la colpevolezza del sospettato dovrebbe passare dal 60% all'88%.

1.1.7 Indipendenza

Gli esempi discussi nel paragrafo precedente mostrano bene il fatto che la probabilità condizionale di A sapendo H , cioè il numero $P(A|H)$, è generalmente diversa dalla probabilità di A , cioè $P(A)$. In altri termini, sapere che si è verificato l'evento H modifica, in generale, la probabilità che si verifichi l'evento A . Nel caso particolare in cui questo non avvenga, tuttavia, cioè quando la probabilità di A sapendo H coincide con la probabilità di A , significa che il sapere che l'evento H si sia verificato o meno, non influenza la probabilità del verificarsi dell'evento A . Quanto appena osservato, come vedremo, giustifica la definizione seguente.

Definizione 1.1.27 Sia (Ω, \mathcal{A}, P) uno spazio probabilizzato. Dati due eventi A, H , si dice che essi sono tra loro *indipendenti* (o che l'uno è indipendente dall'altro) se risulta

$$P(A \cap H) = P(A)P(H). \quad (1.13)$$

Questa relazione è automaticamente verificata (riducendosi alla forma $0 = 0$) se H ha probabilità nulla. Supposto invece che H non sia trascurabile, se si dividono per $P(H)$ i due membri della precedente relazione, questa assume la forma equivalente

$$P(A|H) = P(A). \quad (1.14)$$

Come abbiamo evidenziato all'inizio del paragrafo, il contenuto intuitivo di quest'ultima eguaglianza è evidente; ribadiamolo in altri termini: per un individuo che abbia deciso di distribuire la propria fiducia tra i vari eventi secondo la misura di probabilità P , il fatto che A risulti indipendente da H significa semplicemente che l'ipotetica informazione supplementare “ H si è realizzato”, anche se può convincere l'individuo a cambiare la distribuzione globale della fiducia tra i vari eventi considerati (inducendolo a sostituire P con P_H), non è però capace di alterare la fiducia dell'individuo nell'evento A (nel senso che questo evento continua a ricevere, secondo P_H , la stessa probabilità che riceveva secondo P).

Dimostriamo immediatamente un utile risultato di estrema naturalezza sugli eventi indipendenti, il quale afferma che se A e H sono indipendenti, allora la probabilità di A non è influenzata né dal realizzarsi di H né dal non realizzarsi di H stesso, cioè dal realizzarsi di H^c . In altri termini, dimostriamo che vale il seguente

Teorema 1.1.28 *Se A, H sono due eventi indipendenti, tali sono anche A, H^c .*

Dimostrazione Basterà verificare che è $P(A \cap H^c) = P(A)P(H^c)$. A questo scopo, osserviamo che, siccome l'evento A si può scrivere come l'unione disgiunta degli eventi $A \cap H$ e $A \cap H^c$, dalle proprietà della misura di probabilità P si deduce immediatamente che

$$\begin{aligned} P(A \cap H^c) &= P(A) - P(A \cap H) \\ &= P(A) - P(A)P(H) \\ &= P(A)(1 - P(H)) = P(A)P(H^c), \end{aligned}$$

dove la seconda eguaglianza segue dal fatto che A e H sono tra loro indipendenti. Tanto basta per concludere che A e H^c sono eventi indipendenti. \square

Esempio 1.1.29 (Lancio di due monete) Si supponga che l'esperimento consista nel lancio di una moneta per due volte consecutive. Cerchiamo il naturale spazio probabilizzato da associare a questo esperimento aleatorio.

1. Come *insieme delle eventualità* si potrà prendere l'insieme Ω formato da tutte le possibili coppie composte dagli interi 0 e 1, con la convenzione che 0 significhi croce e 1 testa. Naturalmente, è da intendere che la generica di queste coppie $\omega = (\omega_1, \omega_2)$ rappresenti il risultato che si ottiene quando il primo lancio dà come risultato il numero ω_1 e il secondo lancio dà come risultato il numero ω_2 . Come *tribù degli eventi* si potrà tranquillamente scegliere la tribù di tutte le parti di Ω .
2. Come *misura di probabilità* P da mettere sulla tribù degli eventi (cioè sull'insieme delle parti di Ω) sarà naturale scegliere la ripartizione uniforme. In effetti, non c'è nessun motivo razionale per credere che, per una moneta qualsiasi, un risultato sia più o meno probabile di un altro.

Nell'ambito dello spazio probabilizzato appena costruito, calcoliamo la probabilità che esca testa nel corso del primo lancio. Le eventualità che compongono questo evento sono, evidentemente, $(1, 0)$ e $(1, 1)$. (In effetti, la prima di questa significa "è uscita testa nel corso del primo lancio e croce nel corso del secondo", mentre la seconda significa "è uscita testa in entrambi i lanci".) Detto allora A l'evento in questione, la probabilità richiesta è:

$$P(A) = \frac{\#(A)}{\#(\Omega)} = \frac{1}{2}.$$

Nello stesso modo si ottiene che la probabilità che esca testa nel corso del secondo lancio è $1/2$. Essa è infatti la probabilità dell'evento $B = \{(0, 1), (1, 1)\}$. Si riconosce subito che, come ci si aspetta dall'intuizione, gli eventi A e B sono tra loro indipendenti. Basta per questo osservare che è $A \cap B = \{(1, 1)\}$ e dunque

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

La definizione di indipendenza 1.1.27 si può generalizzare ad un arbitrario numero di eventi nel modo che segue.

Definizione 1.1.30 Nell'ambito di un fissato spazio probabilizzato (Ω, \mathcal{A}, P) , si prendano n eventi A_1, \dots, A_n . Si dice che essi sono tra loro *indipendenti* se per ogni loro sottogruppo A_{n_1}, \dots, A_{n_r} , con $1 \leq n_1 < \dots < n_r \leq n$, si ha

$$P\left(\bigcap_{i=1}^r A_{n_i}\right) = \prod_{i=1}^r P(A_{n_i}).$$

Questa definizione può apparire strana, coinvolgendo ogni sottogruppo dell'insieme degli n eventi. In realtà, tuttavia, essa è un'ovvia generalizzazione della Definizione 1.1.27, ed anzi la amplia e la completa. Per fare un esempio, si osservi innanzitutto che se la condizione che compare nella Definizione 1.1.30 è soddisfatta, in particolare, questa dovrà essere vera per ogni sottogruppo di due elementi, sicché, in particolare, gli eventi A_i

sono a due a due indipendenti secondo la definizione iniziale. Inoltre, se prendiamo ad esempio quattro eventi A, B, C, D tra loro indipendenti, oltre alle ovvie relazioni

$$\begin{aligned} P(A \cap B) &= P(A)P(B), & P(A \cap C) &= P(A)P(C), & P(A \cap D) &= P(A)P(D), \\ P(B \cap C) &= P(B)P(C), & P(B \cap D) &= P(B)P(D), & P(C \cap D) &= P(C)P(D), \end{aligned}$$

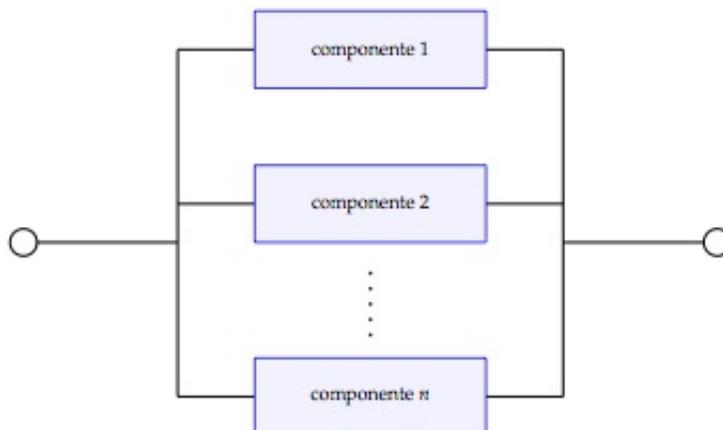
devono essere soddisfatte anche le seguenti:

$$\begin{aligned} P(A \cap B \cap C) &= P(A)P(B)P(C), & P(A \cap B \cap D) &= P(A)P(B)P(D), \\ P(A \cap C \cap D) &= P(A)P(C)P(D), & P(B \cap C \cap D) &= P(B)P(C)P(D), \\ P(A \cap B \cap C \cap D) &= P(A)P(B)P(C)P(D). \end{aligned}$$

Inoltre, osserviamo che, se gli eventi A, B, C sono indipendenti, allora ciascuno di essi è indipendente pure da qualsiasi evento si possa costruire con gli altri due per mezzo delle usuali operazioni insiemistiche. Ad esempio, A è indipendente da $B \cup C$. Infatti:

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

Esempio 1.1.31 Un sistema composto di n componenti distinti si dice *in parallelo* se funziona fino a che almeno uno dei componenti funziona, cioè smette di funzionare soltanto quando tutti i componenti hanno smesso di funzionare (vedi figura sottostante). Sia dato un sistema di questo tipo, per il quale, per $i = 1, 2, \dots, n$ il componente i -esimo funziona (indipendentemente da tutti gli altri) con probabilità p_i . Qual è la probabilità che l'intero sistema funzioni?



All'interno di un opportuno spazio probabilizzato (Ω, \mathcal{A}, P) , per ciascun indice i , denotiamo con A_i l'evento che si realizza se il componente i funziona. Chiaramente, si avrà allora $P(A_i) = p_i$ e gli eventi A_i sono tra loro indipendenti. Inoltre, l'evento descritto dalle parole "il sistema funziona" coincide con l'evento $A_1 \cup \dots \cup A_n$ e di questo

evento è richiesto il calcolo della probabilità. Applicando la relazione (1.4), si ottiene immediatamente

$$P\left(\bigcup_{i=1}^n A_i\right) = 1 - P\left(\bigcap_{i=1}^n A_i^c\right) = 1 - \prod_{i=1}^n P(A_i^c) = 1 - \prod_{i=1}^n (1 - p_i).$$

Esercizi del §1.1

1. Un'urna contiene tre palline indistinguibili: una rossa, una verde e una blu. Si descriva l'insieme delle eventualità dell'esperimento aleatorio consistente nell'estrarre in sequenza due palline, rimettendo la prima pallina nell'urna dopo l'estrazione.
2. Si lanci tre volte una moneta equilibrata. Si descriva l'insieme delle eventualità di questo esperimento aleatorio e si esprima l'evento descritto dalle parole "si ottengono più teste che croci".
3. Sull'insieme $\Omega = \{a, b, c, d, e, f\}$, si considerino gli eventi $A = \{a, c, e\}$, $B = \{d, e, f\}$ e $C = \{a, d\}$. Si descrivano gli elementi dei seguenti eventi:

$$\begin{array}{ll} (a) & A \cap B \\ (b) & A \cap C^c \\ (c) & A^c \cap (B \cup C) \\ (d) & A \cup (B \cap C) \end{array}$$

4. Un sistema è composto da 4 componenti, ciascuno dei quali funziona oppure è guasto. Si osserva lo stato dei componenti, ottenendo un vettore (x_1, x_2, x_3, x_4) , dove la componente x_i vale 1 se il componente i -esimo è funzionante e 0 altrimenti.
 - Si dica da quanti elementi è composto l'insieme di tutte le eventualità.
 - Si sa che il sistema funziona finché entrambi i componenti 1 e 2 oppure 3 e 4 funzionano. Si scriva esplicitamente l'evento costituito dalle parole "il sistema funziona".
 - Sia A l'evento descritto dalle parole "i componenti 1 e 3 sono guasti". Quanti esiti contiene?
5. Si collocano a caso tre palline in tre scatole. Si stabilisca un opportuno spazio probabilizzato (Ω, \mathcal{A}, P) adeguato a rappresentare il problema; quindi si calcoli, nell'ambito del modello scelto, la probabilità che *almeno* una scatola sia vuota.
6. Fissato uno spazio probabilizzato (Ω, \mathcal{A}, P) , si dimostri la *proprietà subadditiva* della misura di probabilità: se A e B sono due eventi qualsiasi, allora $P(A \cup B) \leq P(A) + P(B)$.
7. Su uno spazio probabilizzato (Ω, \mathcal{A}, P) , si dimostrino le due proprietà seguenti:
 - $P(A \setminus B) = P(A) - P(A \cap B)$.
 - $P(A^c \cap B^c) = 1 - P(A) - P(B) + P(A \cap B)$.
8. Su uno spazio probabilizzato (Ω, \mathcal{A}, P) , siano A e B due eventi, con $P(A) = 0.9$ e $P(B) = 0.9$. Si dimostri che $P(A \cap B) \geq 0.8$. Si dimostri poi che, in generale, vale la disuguaglianza

$$P(A \cap B) \geq P(A) + P(B) - 1.$$

9. Su uno spazio probabilizzato (Ω, \mathcal{A}, P) , siano A e B due eventi qualsiasi. Si dimostri che la probabilità che si realizzi uno ed uno solo tra gli eventi A e B è pari a $P(A) + P(B) - 2P(A \cap B)$.
10. Si calcoli ciascuno dei seguenti coefficienti binomiali: $\binom{9}{3}$, $\binom{9}{7}$, $\binom{7}{2}$, $\binom{7}{4}$ e $\binom{10}{7}$.
11. Si scrivono a caso tre numeri interi distinti su tre bigliettini, i quali vengono poi inseriti in tre buste etichettate con A , B e C . Qual è la probabilità che il minore tra i numeri delle buste A e B sia anche minore del numero della busta C ?
12. Da un mazzo ben mescolato composto da 40 carte se ne estraggono 8 in blocco. Dopo aver costruito un opportuno spazio probabilizzato (Ω, \mathcal{A}, P) adeguato a descrivere questa situazione, si calcoli la probabilità che escano:
- tre assi e cinque figure;
 - tutte carte rosse.
13. Da un'urna, contenente 50 palline, numerate da 1 a 50, se ne estraggono in sequenza quattro, rimettendo ogni volta la pallina nell'urna. Dopo aver costruito un opportuno spazio probabilizzato (Ω, \mathcal{A}, P) adeguato a descrivere questa situazione, si calcoli la probabilità che escano:
- esattamente tre palline con un numero maggiore di 45;
 - al più tre palline con un numero inferiore a 21.
14. Da un'urna, contenente quattro palline bianche e tre nere, si eseguono due estrazioni, rimettendo ogni volta la pallina nell'urna. Calcolare la probabilità che:
- le due palline estratte siano dello stesso colore;
 - almeno una delle due palline estratte sia nera.
15. Una moneta viene lanciata 8 volte. Determinare la probabilità che:
- al quinto lancio esca testa;
 - al terzo e al quinto lancio esca testa;
 - ci siano tante teste quante croci;
 - ci siano teste nei tiri pari;
 - la prima testa esca al quinto lancio;
 - siano apparse almeno due teste.
16. Vengono lanciati 5 dadi in sequenza. Qual è la probabilità che:
- il terzo dado mostri la faccia con il numero 4;
 - la somma dei risultati del terzo e del quarto faccia 9.
17. Si lanciano contemporaneamente 4 monete equilibrate. Si calcoli la probabilità che:
- escano almeno tre teste;
 - escano al più tre teste.

18. Da un mazzo composto da 40 carte, se ne estraggono in sequenza 10, rimettendo ogni volta la carta nel mazzo. Si calcoli la probabilità che:
- esattamente tre delle carte uscite siano figure;
 - escano sempre carte di picche;
 - escano cinque carte rosse e cinque carte nere.
19. Un'urna contiene una pallina rossa e due palline bianche. Se ne estraggono cinque, rimettendo ogni volta la pallina nell'urna. Calcolare la probabilità che esca sempre la pallina rossa.
20. Si sceglie a caso un numero di esattamente 4 cifre. Qual è la probabilità che le cifre siano tutte differenti? Qual è la probabilità che ogni cifra sia strettamente maggiore della cifra alla sua destra?
21. Un test di matematica è composto da dieci domande alle quali si può rispondere soltanto *sì* oppure *no*. Calcolare la probabilità che, rispondendo a caso al test, si possa rispondere bene ad esattamente sei domande.
22. Quattro individui si danno appuntamento al Grand Hotel di Parigi. Ci sono però cinque alberghi che portano questo nome. Se le quattro persone si recano, a caso e indipendentemente l'una dall'altra, in uno dei cinque alberghi, qual è la probabilità che tutti e quattro si trovino nello stesso albergo?
23. Nella città di Paperopoli i numeri di telefono degli abitanti sono formati da 5 cifre scelte tra 0 e 9. Determinare la probabilità che il numero di telefono di un abitante di Paperopoli:
- contenga due 1 un 7 e due 3;
 - abbia esattamente tre 1;
 - abbia due e solo due cifre distinte;
 - abbia le cifre tutte distinte;
 - abbia come prodotto delle cifre un numero pari.
24. Una compagnia aerea dispone di due tipi di aereo: uno da 20 posti e un altro da 10 posti. Poiché si sa che i passeggeri che prenotano, poi non si presentano con una probabilità del 10%, vengono sempre accettate 22 prenotazioni sui voli da 20 posti e 11 su quelli da 10 posti. In quale dei due tipi di aereo è maggiore il rischio di lasciare a terra almeno un passeggero che ha regolarmente prenotato, per un volo in cui si è accettato il massimo delle prenotazioni?
25. Un'urna contiene due monete: una di esse ha entrambe le facce nere, mentre l'altra ha una faccia nera ed una faccia bianca. Viene estratta dall'urna una moneta e se ne guarda il colore di una faccia: è nera. Calcolare la probabilità che anche l'altra faccia sia nera.
26. Un'urna contiene due palline rosse e tre palline bianche. Si lancia una moneta equilibrata; indi, se è uscita testa, si estraggono in sequenza due palline dall'urna, rimettendo ogni volta la pallina nell'urna, mentre altrimenti, cioè se è uscita croce, si estrae una sola pallina.

- Calcolare la probabilità che sia uscita esattamente una pallina rossa.
 - Sapendo che alla prima estrazione è uscita una pallina rossa, qual è la probabilità che non ci sia una seconda estrazione?
27. Un'urna contiene r palline rosse e b palline bianche. Si estrae una pallina che viene messa da parte senza guardarla. Dopodiché si estrae una seconda pallina. Calcolare la probabilità che la seconda pallina estratta sia bianca.
28. Un giocatore gioca al lotto i numeri 1, 3, 7, 37 e 42. Determinare la probabilità che egli ottenga:
- una cinquina;
 - una quaterna;
 - un terno;
 - un ambo;
 - nessun numero estratto.
29. (*Il paradosso dei compleanni*). Consideriamo una classe di n persone e a ognuna di esse associamo il numero fra 1 e 365 (per semplicità non consideriamo gli anni bisestili) che corrisponde al numero di giorni tra il primo di gennaio ed il giorno del rispettivo compleanno. Qual è la probabilità che almeno due persone abbiano il compleanno lo stesso giorno? Tracciare un grafico della probabilità trovata al variare del numero n di persone e determinare per quale valore questa probabilità è maggiore di $1/2$.
30. Si dimostri che, su un fissato spazio probabilizzato (Ω, \mathcal{A}, P) , se A, H, K sono tre eventi, vale la formula

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

Si supponga che, prima di ottenere una nuova informazione, l'ipotesi H fosse tre volte più probabile di K . Se l'informazione aggiuntiva A è due volte più probabile quando è vera K rispetto a quando è vera H , qual è l'ipotesi più credibile, tenendo conto della nuova informazione?

31. Un individuo chiede al vicino di innaffiare una piantina delicata mentre egli si trova in viaggio per lavoro. Egli pensa che, senza acqua, la piantina muoia con probabilità 0.80, mentre, se innaffiata, questa abbia una probabilità di morire di 0.15. La sua fiducia che il vicino si ricordi di innaffiarla è del 90%. (a) Si calcoli la probabilità che la piantina sia ancora viva al ritorno dell'individuo. (b) Se la piantina fosse morta, quale sarebbe la probabilità che il vicino si sia dimenticato di innaffiarla?
32. Un test per diagnosticare un certo tipo di malattia ha un'affidabilità del 95% sia per le persone malate che per quelle sane. Se lo 0.4% della popolazione soffre di questa malattia, si calcoli qual è la probabilità che un soggetto che sia risultato positivo al test sia realmente malato.
33. Su 1000 membri di un'associazione di pensionati americani, 600 si dichiarano repubblicani, mentre gli altri si dicono democratici. In occasioni di un'elezione interna in cui hanno votato tutti, 60 repubblicani hanno dato la loro preferenza al candidato democratico e 50 democratici hanno votato a favore del repubblicano. Se un membro

dell'associazione scelto a caso ha votato repubblicano, con che probabilità si tratta di un democratico?

34. (*Urna di Pólya*). Un'urna contiene r palline rosse e b palline bianche. Si estrae una pallina, se ne controlla il colore e si rimette la pallina nell'urna, aggiungendovi una pallina dello stesso colore. Si riestraie di seguito una pallina. Calcolare la probabilità che la prima pallina estratta sia rossa, sapendo che tale è la seconda.
35. Un giocatore lancia due dadi. Se il primo dado mostra la faccia con il numero 3, qual è la probabilità che la somma dei due risultati sia almeno 6?
36. Due palline vengono colorate di rosso oppure di blu, indipendentemente l'una dall'altra, con probabilità $1/2$.
 - Sapendo che la vernice rossa è stata certamente usata, qual è la probabilità che siano entrambe rosse?
 - Si estrae una pallina rossa. Qual è la probabilità che entrambe le palline siano rosse?
37. Due scatole contengono palline colorate. La prima scatola contiene 5 palline rosse e 7 blu, la seconda 8 palline rosse e 3 blu. Si sceglie a caso una scatola e da essa si estraggono due palline, che risultano essere entrambe rosse. Qual è la probabilità che le palline siano state estratte dalla prima scatola?
38. Un gioco consiste nel lancio di una moneta e successivamente di un dado. Se nel lancio della moneta è uscita croce, il concorrente vincerà tante monete quanti il numero che compare sulla faccia del dado. Altrimenti, se esce testa, egli vincerà il doppio. Un giocatore partecipa al gioco: sapendo che egli ha vinto quattro monete, calcolare la probabilità che, durante il lancio della moneta, sia uscita testa.
39. Un paesino, disperso tra le campagne toscane tra Firenze e Pisa, riceve il segnale televisivo al 40% dall'antenna del monte Morello, e per il restante 60% dall'antenna del monte Serra. Il segnale può essere di due tipi: "lungo" o "breve". È noto che l'antenna del monte Morello trasmette un segnale "lungo" il 52% delle volte, mentre l'antenna del monte Serra trasmette il medesimo segnale soltanto il 37% delle volte. Se un abitante del paesino riceve, in un certo istante, un segnale "breve", qual è la probabilità che esso provenga dall'antenna sul monte Serra?
40. Una compagnia di assicurazioni classifica i suoi clienti in tre fasce: *basso rischio*, *medio rischio* e *alto rischio*. Le sue statistiche indicano che le probabilità che il cliente delle tre fasce abbia un incidente entro un periodo di un anno sono rispettivamente 0.05, 0.15 e 0.30. Se i clienti sono per il 20% a basso rischio, per il 50% a medio rischio e per il 30% ad alto rischio, che percentuale di clienti avrà mediamente incidenti entro un anno?
41. Il tumore alla prostata è il più comune tipo di tumore nella popolazione maschile. Un marker solitamente usato dai medici come indicatore della presenza di questo tumore è il livello della proteina PSA (*prostate specific antigen*), che viene prodotto esclusivamente nella ghiandola prostatica. Nonostante livelli più alti di PSA siano associati alla presenza di tumore, questo tipo di test è notoriamente non affidabile. In particolare, la probabilità che un individuo sano abbia valori di PSA al di sopra della soglia di stabilità

è circa 0.135, mentre quella per un individuo malato sale appena a 0.268.

Un medico, basandosi su altri elementi, stima al 70% la probabilità che un certo individuo abbia il tumore alla prostata. Determinare la probabilità che esso sia veramente malato, sapendo che:

- il livello di PSA è superiore alla soglia;
- il livello di PSA è inferiore alla soglia.

42. (*Il dilemma dei tre prigionieri*). Tre prigionieri A , B e C sono in attesa dell'esecuzione di uno dei tre, il cui nome a loro è ignoto, ma noto alla guardia che li sorveglia. Il prigioniero A chiede alla guardia di rivelargli chi tra B e C si salverà o, se saranno entrambi salvi, di riferirgli uno dei due nomi a caso. La guardia si rifiuta, sostenendo che A sa già di avere il 33% di possibilità di essere condannato. Ma, se la guardia rivelasse chi tra B e C si salverà, a quel punto A avrebbe il 50% di possibilità di essere messo a morte. La guardia dice il vero?
43. (*Il gioco delle tre carte*). Il gioco delle tre carte consiste nell'indovinare dov'è l'asso fra tre carte, due delle quali sono re. Un giocatore propone una variante del gioco: le carte vengono disposte dal giocatore (in una maniera che il giocatore stesso conosce) ed il malcapitato di turno deve indovinare, indicando a caso, quale delle tre carte coperte è l'asso. Una volta indicata la carta, il giocatore scopre una delle due carte non scelte rivelando un re (se entrambe le carte non indicate sono re, la carta da scoprire è scelta a caso) e proponendo al malcapitato la possibilità di cambiare la sua scelta, spostandola sulla carta rimanente. Cosa conviene fare? Perché?
44. Un difetto di produzione ha una incidenza dell'1%. Un test per identificare il difetto risulta positivo il 95% delle volte per i prodotti effettivamente difettosi, e il 5% delle volte per i prodotti integri. Si esamina un prodotto e il test realizzato su di esso risulta positivo. Qual è la probabilità che sia effettivamente un prodotto difettoso?
45. In un mucchietto di 65 monete c'è una moneta falsa (una moneta avente testa su entrambe le facce). Una moneta viene estratta a caso dal mucchietto e lanciata 6 volte. Tutte e sei le volte esce testa. Qual è la probabilità che la moneta scelta sia proprio quella falsa?
46. Un giornalista vuole fare una stima di quanti tra gli imprenditori italiani investono denaro all'estero. Poiché la risposta potrebbe essere imbarazzante per qualche imprenditore, egli decide di far tirare a ciascun imprenditore un dado, in modo tale che egli scelga di dire:
- *sì*, se esce il numero 1 oppure il numero 2;
 - *no*, se esce il numero 3 oppure il numero 4;
 - *la verità*, se esce il numero 5 oppure il numero 6.

Si scopre così che il 60% degli imprenditori dichiara di investire denaro all'estero. Calcolare la probabilità che un imprenditore investa realmente denaro all'estero.

47. Un'urna contiene 112 dadi di cui 56 (cioè la metà) sono equilibrati, mentre gli altri sono stati manipolati in maniera tale che, per ciascuno di essi, la probabilità di ottenere 1 sia $1/2$, mentre ogni altro risultato si verifica con probabilità $1/10$. Un dado viene

estratto a caso e lanciato. Calcolare la probabilità che esca la faccia corrispondente al numero 1.

48. Tre amici lasciano le proprie giacche al guardaroba di un locale. Al momento di ritirarle si accorgono che i tagliandi si sono mescolati. Sia A_i l'evento per cui la persona i riceve la propria giacca. Dire se gli eventi A_1 , A_2 e A_3 siano o meno indipendenti. Dire se essi siano o meno a due a due indipendenti.
49. Due dadi vengono tirati. Consideriamo i tre seguenti eventi:

$$\begin{aligned} A &= \{\text{il primo dado dà un numero dispari}\}, \\ B &= \{\text{il secondo dado dà un numero pari}\}, \\ C &= \{\text{la somma dei due risultati è pari}\}. \end{aligned}$$

Dire se i tre eventi A, B, C sono indipendenti. Dire se sono a due a due indipendenti.

1.2 Le variabili aleatorie

1.2.1 Definizione di variabile aleatoria, legge, indipendenza

Un individuo che compia un esperimento aleatorio non è spesso interessato a studiare il risultato dell'esperimento in quanto tale, ma piuttosto egli sarà interessato a studiare certe quantità numeriche che sono "funzioni" del risultato dell'esperimento stesso. A dire il vero, si può tranquillamente affermare che, in moltissime situazioni probabilistiche, è proprio la funzione del risultato che è interessante, più dell'esperimento in sé e per sé. Si capisce dunque perché queste funzioni hanno assunto un ruolo centrale nel calcolo delle probabilità dove intervengono da indiscusse protagoniste sotto il nome di "variabili aleatorie".

Definizione 1.2.1 Su uno spazio probabilizzato (Ω, \mathcal{A}, P) , si chiama una *variabile aleatoria* (reale) ogni funzione $X : \Omega \rightarrow \mathbb{R}$ che rispetti la condizione seguente: per ogni intervallo $A \subseteq \mathbb{R}$, l'insieme

$$\{\omega \in \Omega : X(\omega) \in A\} \tag{1.15}$$

è un evento, cioè appartiene alla tribù \mathcal{A} . Ora, se chiamiamo tribù *boreliana* di \mathbb{R} , indicandola con il simbolo $\mathcal{B}(\mathbb{R})$, la più piccola tribù che contenga gli intervalli, e se chiamiamo *boreliani* gli elementi di questa tribù, si può dimostrare che, se \mathcal{A} è sufficientemente ricca da contenere gli insiemi della forma (1.15), con A intervallo, essa contiene anche tutti gli insiemi della forma (1.15), con A boreliano di \mathbb{R} . Inoltre, l'insieme costituito da tutti gli eventi della forma (1.15) con A boreliano di \mathbb{R} forma una sottotribù di \mathcal{A} (cioè un sottoinsieme di \mathcal{A} che è esso stesso una tribù su Ω), che prende il nome di *tribù generata* da X e che si denota usualmente con $\mathcal{T}(X)$.

Osservazione 1.2.2 Talvolta è utile estendere la definizione appena data anche nel caso di funzioni definite sull'insieme Ω a valori nella *retta reale estesa* $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, +\infty\}$, ossia è utile permettere alla variabile aleatoria di assumere anche i valori $-\infty$ e $+\infty$. In questo caso, la definizione appena data non cambia, pur di considerare come intervalli

possibili anche quelli della forma $[a, +\infty]$ e $[-\infty, a]$; una siffatta variabile aleatoria si dice di solito una variabile aleatoria *numerica*.

Se X è una variabile aleatoria, allora, per ogni insieme boreliano A di \mathbb{R} , l'evento (1.15) (che è realizzato da tutte e sole le eventualità ω di Ω tali che $X(\omega)$ appartenga all'intervallo A) si denota brevemente con il simbolo $\{X \in A\}$ (da leggere: “ X cade in A ”). Inoltre, la probabilità di tale evento, anziché con $P(\{X \in A\})$, si denoterà più semplicemente con $P(X \in A)$ e si chiama anche la “probabilità che la variabile aleatoria X cada in A ”.

Definizione 1.2.3 *Su uno spazio probabilizzato, supponiamo assegnata una variabile aleatoria X . Si chiama la legge (o distribuzione) di X (secondo P) l'applicazione*

$$A \mapsto P(X \in A)$$

che ad ogni insieme boreliano A di \mathbb{R} associa la probabilità che X cada in A . Non è difficile riconoscere che la legge di X secondo P altro non è che una misura di probabilità sulla tribù boreliana di \mathbb{R} .

Nelle applicazioni, è importante ricordare anche il seguente *criterio fondamentale per la coincidenza di due misure di probabilità*:

se le leggi di due variabili aleatorie coincidono su ogni intervallo di \mathbb{R} , esse sono identiche.

Da questo fatto segue immediatamente che, per avere delle informazioni sulla legge di una variabile aleatoria, sarà sufficiente conoscere come essa si comporta su tutti gli intervalli di \mathbb{R} . Addirittura, come si vedrà in seguito, sarà sufficiente conoscere il comportamento della legge di una variabile aleatoria soltanto su certi intervalli particolarmente semplici, per ottenere informazioni su tutta la legge.

La legge di una variabile aleatoria X (definita su un opportuno spazio probabilizzato) può essere pensata come una “fotografia” delle varie probabilità assegnate a tutti gli eventi della forma $\{X \in A\}$, con A insieme boreliano di \mathbb{R} .

Esempio 1.2.4 Supponiamo di scommettere sul lancio di due dadi equilibrati e di essere interessati, per l'esito della scommessa, alla somma dei valori usciti dal lancio dai due dadi. Denotiamo dunque con X la somma delle facce uscite nei due lanci. Evidentemente, X è una variabile aleatoria definita sullo spazio probabilizzato (Ω, \mathcal{A}, P) che abbiamo scelto per rappresentare il lancio dei due dadi (esempio 1.1.29): Ω è l'insieme formato da tutte le possibili coppie d'interi compresi tra 1 e 6, ove la generica di queste coppie (ω_1, ω_2) rappresenta il risultato che si ottiene quando il primo dado dà come risultato il numero ω_1 e il secondo dado dà come risultato il numero ω_2 ; \mathcal{A} è la tribù di tutte le parti di Ω , e la misura di probabilità P è la ripartizione uniforme.

Evidentemente la variabile aleatoria X può assumere soltanto un numero finito di valori e, precisamente, tutti gli interi compresi tra un minimo di 2 e un massimo di 12. La sua

legge sarà dunque univocamente determinata una volta noto il suo valore su ciascuno di questi valori. Precisamente:

$$\begin{aligned}
 P(X = 2) &= P(\{(1, 1)\}) = \frac{1}{36} \\
 P(X = 3) &= P(\{(1, 2), (2, 1)\}) = \frac{2}{36} \\
 P(X = 4) &= P(\{(1, 3), (2, 2), (3, 1)\}) = \frac{3}{36} \\
 P(X = 5) &= P(\{(1, 4), (2, 3), (3, 2), (4, 1)\}) = \frac{4}{36} \\
 P(X = 6) &= P(\{(1, 5), (2, 4), (3, 3), (4, 2), (5, 1)\}) = \frac{5}{36} \\
 P(X = 7) &= P(\{(1, 6), (2, 5), (3, 4), (4, 3), (5, 2), (6, 1)\}) = \frac{6}{36} \\
 P(X = 8) &= P(\{(2, 6), (3, 5), (4, 4), (5, 3), (6, 2)\}) = \frac{5}{36} \\
 P(X = 9) &= P(\{(3, 6), (4, 5), (5, 4), (6, 3)\}) = \frac{4}{36} \\
 P(X = 10) &= P(\{(4, 6), (5, 5), (6, 4)\}) = \frac{3}{36} \\
 P(X = 11) &= P(\{(5, 6), (6, 5)\}) = \frac{2}{36} \\
 P(X = 12) &= P(\{(6, 6)\}) = \frac{1}{36}
 \end{aligned}$$

Questa legge si può anche sintetizzare nel modo seguente:

X	2	3	4	5	6	7	8	9	10	11	12
P	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36

Se ci domandiamo, così, qual è la probabilità che la somma dei valori usciti sia maggiore di 9, essa si potrà facilmente calcolare nel modo seguente:

$$\begin{aligned}
 P(X > 9) &= P(X = 10) + P(X = 11) + P(X = 12) \\
 &= \frac{3}{36} + \frac{2}{36} + \frac{1}{36} = \frac{6}{36} = \frac{1}{6}.
 \end{aligned}$$

A ben guardare, almeno nel caso di una variabile aleatoria X che prende i suoi valori in un insieme E finito o al più numerabile, la legge L_X di X è univocamente determinata da una funzione p_X , definita sull'insieme E a valori in $[0, 1]$, tale che

$$\sum_{k \in E} p_X(k) = 1,$$

perché, come subito si riconosce,

$$\sum_{k \in E} p_X(k) = \sum_{k \in E} P(X = k) = P\left(\bigcup_{k \in E} \{X = k\}\right) = P(\Omega) = 1.$$

Dunque la legge non dipende dallo spazio probabilizzato su cui la variabile aleatoria X è definita, ma solo dai valori che assume X . In questo senso, non ci importa più dello spazio probabilizzato sottostante: ad esempio, nel caso in cui gli esiti possibili sono solo due, come nel lancio di una moneta, non ci importa se si tratti di testa o croce, pari o dispari, malato o sano, maschio o femmina: parleremo solo di successo e insuccesso, e ci interessano solo le relative probabilità, siano esse $\frac{1}{2}$ e $\frac{1}{2}$ o magari p e $q = 1 - p$, con $p \in [0, 1]$.

In analogia con quanto abbiamo definito nel paragrafo 1.1.7, due variabili aleatorie saranno considerabili come indipendenti l'una dall'altra quando tutti gli eventi che si possono formare con l'una saranno indipendenti da tutti quelli che si possono formare con l'altra. Ora, poiché, come abbiamo peraltro già osservato, gli eventi della forma $\{X \in A\}$, dove A è un intervallo di \mathbb{R} , sono rappresentativi di tutti gli eventi di questa forma con A boreliano di \mathbb{R} , ciò giustifica totalmente la definizione che segue e che diamo per semplicità per due variabili aleatorie (ricalcando la Definizione 1.1.27) ma che si estende naturalmente anche al caso di una sequenza di variabili aleatorie (ricalcando stavolta la Definizione 1.1.30).

Definizione 1.2.5 Su uno spazio probabilizzato (Ω, \mathcal{A}, P) siano X, Y due variabili aleatorie. Esse si dicono tra loro *indipendenti* se accade che, per ogni coppia I, J d'intervalli di \mathbb{R} , gli eventi $\{X \in I\}$ e $\{Y \in J\}$ sono tra loro indipendenti. Precisamente, le variabili aleatorie X, Y sono tra loro indipendenti se e soltanto se risulta

$$P(X \in I, Y \in J) = P(X \in I)P(Y \in J)$$

per ogni coppia I, J d'intervalli di \mathbb{R} .

Esempio 1.2.6 Riprendiamo l'esempio 1.1.29 e denotiamo con X e con Y le applicazioni che ad ogni coppia (ω_1, ω_2) associano rispettivamente i numeri ω_1 e ω_2 . Non è difficile riconoscere che si tratta di due variabili aleatorie, e che esse sono per giunta indipendenti. In effetti, poiché entrambe prendono soltanto i valori 0 oppure 1, la condizione d'indipendenza si riduce a richiedere

$$P(X = k, Y = h) = P(X = k)P(Y = h) \quad \forall h, k \in \{0, 1\},$$

condizione che è completamente determinata dai quattro eventi $\{X = 0\}$, $\{X = 1\}$, $\{Y = 0\}$, $\{Y = 1\}$.

Abbiamo motivato la nozione di variabile aleatoria con l'opportunità di considerare delle funzioni di un esperimento aleatorio. In realtà, la loro importanza va molto più in là: d'ora in avanti il modello fondamentale dello studio di un esperimento aleatorio sarà costituito da un opportuno spazio probabilizzato (Ω, \mathcal{A}, P) , di cui spesso ignoreremo addirittura l'esatta natura, sul quale sono definite delle variabili aleatorie con certe leggi assegnate. La teoria del calcolo delle probabilità garantirà, sotto ipotesi più o meno ragionevoli, che tali modelli di volta in volta siano realmente esistenti, ma la nostra attenzione sarà sempre indirizzata verso le leggi che stabiliremo, di volta in volta, essere le più adatte alla risoluzione dei vari problemi probabilistici.

Capiterà molto spesso di aver necessità di costruire uno spazio probabilizzato (Ω, \mathcal{A}, P) sul quale sono definite un certo numero (oppure anche una successione) di variabili aleatorie tutte indipendenti tra di loro e dotate di leggi prestabilite, che spesso e volentieri (ma non sempre) sono una medesima legge fissata. In linea di principio, per far questo, occorrerebbe costruire un opportuno insieme delle eventualità Ω , munirlo di un'opportuna tribù \mathcal{A} e quindi costruire una misura di probabilità P che renda indipendenti le variabili aleatorie richieste e, pure, che assegni loro le desiderate leggi. Fortunatamente, nel calcolo delle probabilità, esiste un teorema che risolve la questione una volta per tutte senza dover ricorrere ogni volta alla costruzione concreta dello spazio probabilizzato (Ω, \mathcal{A}, P) . Precisamente:

Teorema 1.2.7 (Schema delle prove indipendenti) *Sia data una famiglia $(\mathcal{L}_t)_{t \in T}$ di leggi su \mathbb{R} , con $\mathcal{L}_t : E_t \subseteq \mathbb{R} \rightarrow [0, 1]$. Allora esistono uno spazio probabilizzato (Ω, \mathcal{A}, P) ed una famiglia $(X_t)_{t \in T}$ di variabili aleatorie indipendenti, con $X_t : \Omega \rightarrow E_t$, tali che, per ogni indice t in T , la legge di X_t secondo P è \mathcal{L}_t .*

1.2.2 Variabili aleatorie discrete, leggi discrete

Nello studio delle variabili aleatorie distingueremo due casi, a seconda che le variabili aleatorie in questione possano assumere un insieme *continuo* di valori, oppure un insieme *discreto*. Considereremo dapprima quest'ultimo caso, che è più semplice; in particolare, vedremo alcune situazioni tipiche e le leggi delle variabili aleatorie che in esse compaiono.

Consideriamo dunque una variabile aleatoria X , definita su uno spazio probabilizzato (Ω, \mathcal{A}, P) , che possa assumere soltanto un insieme discreto E di valori: nella maggior parte delle applicazioni si tratterà dei numeri $0, 1, \dots, n$ oppure di tutti gl'interi naturali. Chiameremo *discreta* una siffatta variabile aleatoria. In questo caso, la legge di X è individuata, come abbiamo visto nel paragrafo precedente, non appena sia determinata la probabilità che X assuma il valore k (con $k \in E$), ovvero non appena si conoscano i numeri

$$p_X(k) = P(X = k).$$

La funzione p_X , di E in $[0, 1]$ è una *densità discreta* di probabilità su E e si chiama la *densità discreta* della legge di X . Essa assume soltanto un numero finito o tutt'al più numerabile di valori in corrispondenza degli elementi di E e deve soddisfare alla relazione

$$\sum_{x \in E} p_X(x) = 1.$$

Dedichiamo il resto di questo paragrafo alla descrizione di alcune delle leggi più importanti delle variabili aleatorie discrete, cercando di descriverne le proprietà che le caratterizzano e quali siano i relativi campi d'applicazione. A questo scopo, supponiamo fissato una volta per tutte uno spazio probabilizzato (Ω, \mathcal{A}, P) .

La legge uniforme

Sia N un intero positivo. Si chiama la *legge uniforme* sull'insieme $\{1, 2, \dots, N\}$ la legge di una variabile aleatoria X , definita su (Ω, \mathcal{A}, P) , che prenda tutti i valori k fra 1 e N

con la stessa probabilità: la misura di probabilità sarà quindi la ripartizione uniforme di parametro $p = \frac{1}{N}$. Questa legge si denoterà con il simbolo $\mathcal{U}(N)$. La densità di questa variabile aleatoria al variare di k è costantemente uguale a $\frac{1}{N}$. Ad esempio, il risultato del lancio di un dado equilibrato si rappresenta con una variabile aleatoria di legge $\mathcal{U}(6)$.

La legge di Bernoulli

Sia p un qualsiasi numero reale compreso tra 0 e 1. Si chiama la *legge di Bernoulli* di parametro p la legge di una variabile aleatoria X , definita su (Ω, \mathcal{A}, P) , che prenda due soli valori: il valore 1 con probabilità p e il valore 0 con probabilità $1 - p$. Questa legge si denoterà con il simbolo $\mathcal{B}(p)$ e una variabile aleatoria dotata di una tal legge si chiamerà anche una variabile aleatoria *bernoulliana* di parametro p .

Solitamente, siamo in presenza di una variabile aleatoria bernoulliana quando si studia un esperimento aleatorio nel quale l'esito dell'esperimento può essere soltanto un "successo" (che si rappresenta tipicamente con il valore 1) o un "insuccesso" (che si rappresenta tipicamente con il valore 0). In questo senso, una variabile aleatoria bernoulliana X si potrà descrivere come

$$X = \begin{cases} 1 & \text{in caso di successo} \\ 0 & \text{in caso di insuccesso} \end{cases}$$

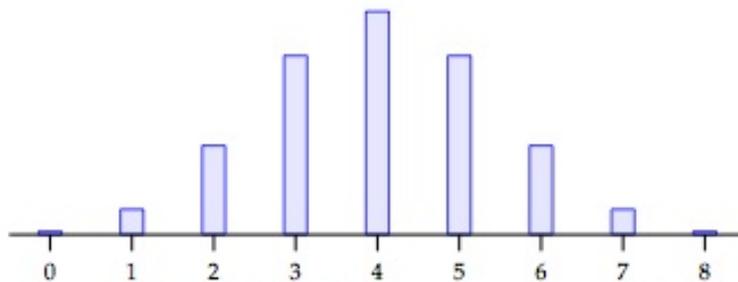
e il numero p rappresenta la *probabilità di successo*.

La legge binomiale

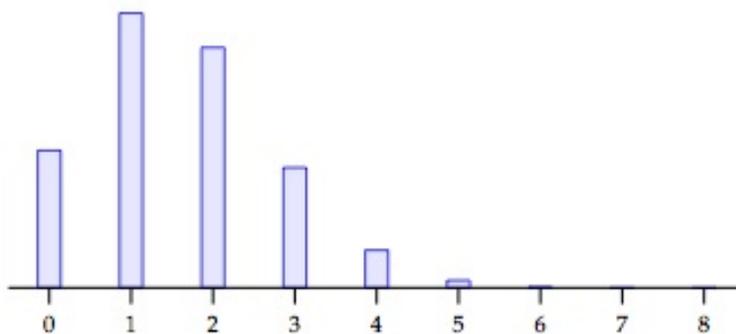
Siano ora p un numero reale compreso tra 0 e 1, e n un intero naturale. Si chiama la *legge binomiale* di parametri n, p la legge di una variabile aleatoria X , definita su (Ω, \mathcal{A}, P) , che prenda i valori $0, 1, 2, \dots, n$ con le probabilità

$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k} \quad k = 0, 1, 2, \dots, n. \quad (1.16)$$

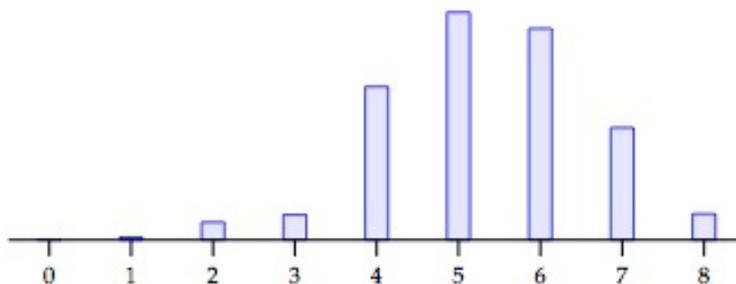
Le leggi binomiali si indicano con il simbolo $\mathcal{B}(n, p)$. Osserviamo che, per $n = 1$, si ottiene, come caso particolare, la legge di Bernoulli, cioè la legge $\mathcal{B}(1, p)$ coincide con $\mathcal{B}(p)$. I grafici nella figura sottostante mostrano l'andamento di alcune leggi binomiali. Come subito si riconosce, al crescere di k la densità cresce fino ad un valore massimo (che si trova non lontano dal valore np) per poi decrescere nuovamente. Notiamo anche che le densità sono tanto più asimmetriche quanto più p è vicino ai valori estremi 0 e 1.



(a) Andamento della densità $\mathcal{B}(8, 0.5)$. C'è una simmetria intorno al valore centrale $k = 4$.



(b) Andamento della densità $\mathcal{B}(8, 0.2)$: i valori 6, 7 e 8 vengono assunti con probabilità prossima a 0.



(c) Andamento della densità $\mathcal{B}(8, 0.65)$. Si può vedere che con l'aumentare di p aumenta la probabilità di osservare valori grandi e diminuisce quella di osservare i valori piccoli.

Per capire il significato intuitivo della legge binomiale, consideriamo un esperimento aleatorio costituito da n prove ripetute e indipendenti (come ad esempio n lanci successivi di una stessa moneta) in ciascuna delle quali sono possibili due risultati, che indicheremo convenzionalmente con 0 e con 1 (un tale modello esiste per il teorema 1.2.7). Supponiamo inoltre che, in ogni singola prova, il risultato 1 si verifichi con probabilità p (con p numero reale compreso tra 0 e 1). Allora la probabilità che il numero 1 appaia k volte è appunto data dalla (1.16). In effetti, la variabile aleatoria X che rappresenta il numero di volte in cui 1 compare nel corso di n prove ha una legge binomiale $\mathcal{B}(n, p)$. Per dimostrarlo, iniziamo col considerare, sullo spazio probabilizzato (Ω, \mathcal{A}, P) , due variabili aleatorie X, Y indipendenti, la prima delle quali abbia legge $\mathcal{B}(n-1, p)$ e la seconda abbia legge $\mathcal{B}(1, p)$; calcoliamo la legge della variabile aleatoria $Z = X + Y$. Poiché Y può assumere soltanto i valori 0 e 1, se $Z = k$ vi sono due possibilità: $X = k$

e $Y = 0$, oppure $X = k - 1$ e $Y = 1$. Dunque:

$$\begin{aligned} P(Z = k) &= P(X = k, Y = 0) + P(X = k - 1, Y = 1) \\ &= P(X = k)P(Y = 0) + P(X = k - 1)P(Y = 1) \\ &= \binom{n-1}{k} p^k (1-p)^{n-1-k} (1-p) + \binom{n-1}{k-1} p^{k+1} (1-p)^{n-k} p \\ &= \left[\binom{n-1}{k} + \binom{n-1}{k-1} \right] p^k (1-p)^{n-k}. \end{aligned}$$

D'altra parte, sappiamo che è $\binom{n-1}{k} + \binom{n-1}{k-1} = \binom{n}{k}$ (si veda la (1.8)), e dunque

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

cioè Z ha legge $\mathcal{B}(n, p)$. Tanto premesso, torniamo allo schema delle prove ripetute e indipendenti e consideriamo, sullo spazio probabilizzato (Ω, \mathcal{A}, P) , le variabili aleatorie X_1, \dots, X_n così definite: per ciascun indice i compreso tra 1 e n , sia

$$X_i = \begin{cases} 1 & \text{se l}'i\text{-esima prova ha dato risultato 1,} \\ 0 & \text{altrimenti.} \end{cases}$$

Le variabili aleatorie X_1, \dots, X_n hanno legge di Bernoulli di parametro p e sono indipendenti. Inoltre, il numero totale delle volte in cui compare il numero 1 è dato dalla variabile aleatoria $X = X_1 + \dots + X_n$ che, per il conto che abbiamo fatto, ha legge binomiale $\mathcal{B}(n, p)$.

Esempio 1.2.8 Un problema frequente, che si riconduce alle leggi binomiali, è il seguente. Si consideri una popolazione composta da due tipi di individui: quelli di tipo A e quelli di tipo B (per esempio, maschi e femmine, sani e malati, fumatori e non fumatori, ecc.). Si supponga inoltre che la percentuale di individui di tipo A all'interno della popolazione sia p e, di conseguenza, la percentuale di individui di tipo B sia $q = 1 - p$. Da una siffatta popolazione, si scelgano a caso n individui e si voglia stabilire quanti di essi siano di tipo A . Poniamo a questo scopo $X_k = 1$ se il k -esimo individuo nel campione è di tipo A e $X_k = 0$ altrimenti.

Se la scelta degli individui è fatta in modo casuale, si può supporre che le variabili aleatorie X_1, \dots, X_n siano indipendenti, assumano il valore 1 con probabilità p e 0 con probabilità $q = 1 - p$, e dunque siano variabili aleatorie bernoulliane di parametro p . Il numero totale d'individui di tipo A nel campione è dunque dato dalla variabile aleatoria $X = X_1 + \dots + X_n$ che ha legge binomiale $\mathcal{B}(n, p)$.

Nella pratica, dunque, quando scegliamo un campione e contiamo il numero d'individui di tipo A nel campione, il numero ottenuto è una variabile aleatoria che segue una legge $\mathcal{B}(n, p)$, dove n è il numero d'individui nel campione e p la proporzione d'individui di tipo A nella popolazione. Chiaramente, il numero di individui di tipo B si potrà poi contare tramite la variabile aleatoria $Y = n - X$ che, per ovvie ragioni di simmetria, ha legge binomiale $\mathcal{B}(n, q)$, giacché è la somma delle n variabili aleatorie $Y_i = 1 - X_i$ (con $i = 1, \dots, n$) indipendenti e bernoulliane di parametro $q = 1 - p$.

La legge geometrica

Sullo spazio probabilizzato (Ω, \mathcal{A}, P) sia adesso X_1, \dots, X_n, \dots una sequenza infinita di variabili aleatorie indipendenti e bernoulliane di parametro p . Come abbiamo detto, ciascuna di queste variabili aleatorie descrive un esperimento aleatorio nel quale sono possibili soltanto due risultati: successo oppure insuccesso. L'intera sequenza si potrà dunque pensare come un insieme (potenzialmente infinito) di prove che si ripetono con indipendenza e l'insieme degli indici si potrà pensare come la sequenza discreta degli istanti di tempo in cui la singola prova ha luogo. Vogliamo considerare l'*istante di primo successo* dell'intera sequenza. Evidentemente, esso potrà essere descritto dalla funzione:

$$T(\omega) = \inf\{n \in \mathbb{N}^+ : X_n(\omega) = 1\},$$

con la convenzione che, se l'insieme $\{n \in \mathbb{N}^+ : X_n(\omega) = 1\}$ è vuoto, cioè se per ciascun indice i si ha $X_i(\omega) = 0$, allora si porrà $T(\omega) = \inf \emptyset = +\infty$. Si riconosce subito che T è una variabile aleatoria discreta, a valori in $\mathbb{N}^+ \cup \{+\infty\}$: infatti, per ogni intero n si ha

$$\{T > n\} = \{X_1 = 0, X_2 = 0, \dots, X_n = 0\} = \bigcap_{i=1}^n \{X_i = 0\}$$

ed il secondo membro di questa eguaglianza è un evento (cioè appartiene alla tribù \mathcal{A}) grazie alle proprietà delle tribù. Calcoliamo la probabilità di questo evento: tenuto conto che X_1, \dots, X_n sono indipendenti e tutte dotate della medesima legge bernoulliana di parametro p , e ponendo al solito $q = 1 - p$, si ha

$$P(T > n) = \prod_{i=1}^n P(X_i = 0) = q^n.$$

Osservando dunque che l'evento $\{T = +\infty\}$ è contenuto in ciascuno degli eventi $\{T > n\}$, per la proprietà di isotonia e passando al limite per $n \rightarrow +\infty$, si deduce immediatamente che $P(T = +\infty) \leq q^n \rightarrow 0$ e quindi $P(T = +\infty) = 0$, cioè $\{T = +\infty\}$ è un evento trascurabile. Inoltre, osservando che $\{T = n\} = \{T > n - 1\} \setminus \{T > n\}$ si ricava immediatamente la legge di T :

$$P(T = n) = P(T > n - 1) - P(T > n) = q^{n-1} - q^n = pq^{n-1}.$$

Questa legge si chiama la *legge geometrica* di parametro p , si denota con $\mathcal{G}(p)$ e rappresenta, come abbiamo detto, la legge dell'istante di primo successo di una sequenza (potenzialmente infinita) di prove indipendenti e ripetute.

Una delle proprietà più importanti della legge geometrica è la cosiddetta *assenza di memoria*. Per descrivere questa proprietà, sia $(X_i)_{i \in \mathbb{N}}$ una sequenza infinita di variabili aleatorie indipendenti e bernoulliane di parametro p e si denoti con T l'istante di primo successo. Fissato un intero positivo n , si consideri l'evento $H = \{T > n\} = \{X_1 = 0, \dots, X_n = 0\}$. Imitando il linguaggio dei giocatori del lotto, questo evento si potrebbe descrivere con le parole "ritardo di n estrazioni nell'apparizione del primo successo". Su di esso, la variabile aleatoria $T - n$ rappresenta il *tempo residuo* da attendere per veder apparire il primo successo dopo i primi n istanti trascorsi invano. Osserviamo adesso

che la legge di $T - n$ secondo la probabilità condizionale P_H coincide con la legge di T secondo P , cioè con la legge geometrica di parametro p : per riconoscerlo basta infatti osservare che

$$P(T - n = k | H) = \frac{P(T - n = k, T > n)}{P(T > n)} = \frac{P(T = n + k)}{P(T > n)} = \frac{q^{n+k-1}p}{q^n} = q^{k-1}p$$

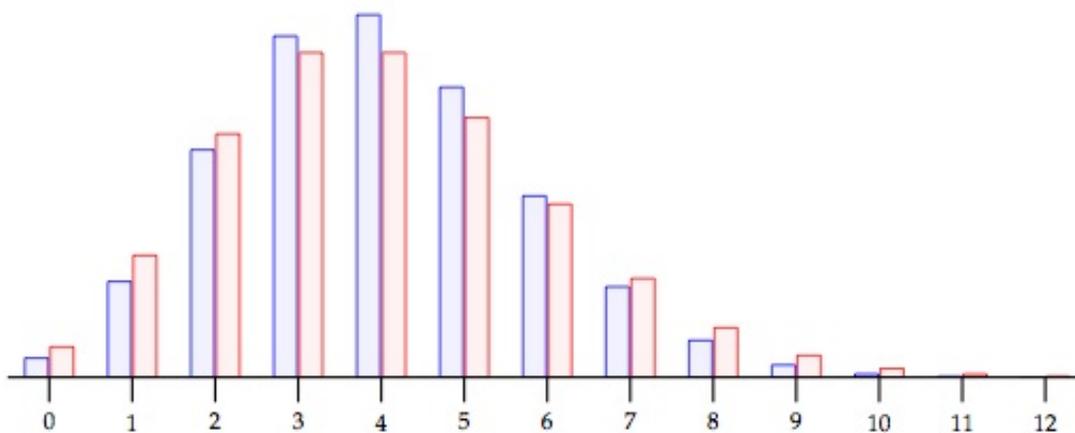
dove la seconda eguaglianza segue dal fatto che gli eventi $\{T = n + k\}$ e $\{T - n = k, T > n\}$ coincidono, essendo uguali all'evento $\{X_1 = 0, \dots, X_{n+k-1} = 0, X_{n+k} = 1\}$. Intuitivamente, questo risultato significa che, sotto la condizione di non aver mai avuto successo fino all'istante n incluso, la legge del tempo residuo da attendere, a partire da quell'istante, per vedere apparire il primo successo è sempre la stessa, qualunque sia n (cioè comunque grande sia il "ritardo"). Questa legge, infatti, è sempre uguale alla legge geometrica $\mathcal{G}(p)$.

La legge di Poisson

Un'altra legge naturale in molte situazioni è la cosiddetta "legge di Poisson". Si chiama *legge di Poisson* di parametro λ la legge di una variabile aleatoria X , definita su uno spazio probabilizzato (Ω, \mathcal{A}, P) , a valori nell'insieme di tutti gl'interi naturali, con

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

Indicheremo questa legge con il simbolo $\mathcal{P}(\lambda)$. L'importanza della legge di Poisson deriva dal fatto che, se n è grande e p è piccolo, una legge binomiale $\mathcal{B}(n, p)$ si può approssimare con una legge di Poisson $\mathcal{P}(np)$: si veda la figura sottostante, dove si confrontano la legge binomiale $\mathcal{B}(20, 0.2)$, in azzurro, e la legge di Poisson $\mathcal{P}(4)$, in rosa.



In altre parole

$$\binom{n}{k} p^k (1-p)^{n-k} \sim \frac{(np)^k}{k!} e^{-np}.$$

Per riconoscerlo, osserviamo che, se X è una variabile aleatoria, su un opportuno spazio probabilizzato (Ω, \mathcal{A}, P) , con legge binomiale $\mathcal{B}(n, \lambda/n)$, si ha, al tendere del

parametro n all'infinito:

$$\begin{aligned} P(X = k) &= \binom{n}{k} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} = \frac{n!}{k!(n-k)!} \frac{\lambda^k}{n^k} \left(1 - \frac{\lambda}{n}\right)^{n-k} \\ &= \frac{\lambda^k}{k!} \frac{n(n-1)\cdots(n-k+1)}{n^k} \left(1 - \frac{\lambda}{n}\right)^n \left(1 - \frac{\lambda}{n}\right)^{-k} \rightarrow \frac{\lambda^k}{k!} e^{-\lambda}, \end{aligned}$$

dove abbiamo usato i ben noti limiti, per $n \rightarrow \infty$,

$$\begin{aligned} \left(1 - \frac{\lambda}{n}\right)^{-k} &\rightarrow 1, \\ \left(1 - \frac{\lambda}{n}\right)^n &\rightarrow e^{-\lambda}, \\ \frac{n(n-1)\cdots(n-k+1)}{n^k} &\rightarrow 1. \end{aligned}$$

Esempio 1.2.9 Riprendiamo l'esempio 1.1.4: si vuole studiare il numero di persone che si trovano in fila all'ufficio postale il venerdì alle 10:00. A questo scopo, se indichiamo con N il numero dei possibili abitanti della città, per ogni indice i compreso tra 1 e N , indichiamo con X_i la variabile aleatoria bernoulliana così definita:

$$X_i = \begin{cases} 1 & \text{se l'abitante } i\text{-esimo si trova in fila alla posta,} \\ 0 & \text{altrimenti.} \end{cases}$$

È evidente che le variabili aleatorie X_i sono tutte indipendenti e si può assumere che esse siano tutte bernoulliane del medesimo parametro p , che rappresenta la probabilità che un generico cittadino si trovi alla posta in quel momento. Il numero di persone in fila all'ufficio postale si potrà allora rappresentare mediante la variabile aleatoria $X = X_1 + \cdots + X_N$ che ha legge binomiale $\mathcal{B}(N, p)$. D'altra parte, poiché il numero N è molto grande e la probabilità p molto piccola (si deve immaginare, infatti, che l'evento descritto dalle parole "il cittadino i si trova all'ufficio postale il venerdì alle 10:00" sia abbastanza raro), la legge di X si potrà approssimare con la legge di Poisson $\mathcal{P}(Np)$ e quindi, posto $\lambda = Np$, si avrà

$$P(X = k) \approx \frac{\lambda^k}{k!} e^{-\lambda} \quad \text{per ogni } k \in \mathbb{N}.$$

Esempio 1.2.10 Si vogliono contare il numero di particelle alfa emesse da un grammo di un certo materiale radioattivo \mathbf{X} in un determinato istante di tempo. Sappiamo dalla teoria del decadimento alfa che ciascun atomo del materiale \mathbf{X} ha una bassissima probabilità p di emettere una particella alfa e, per contro, un grammo di materiale è composto da un enorme numero N di atomi (dell'ordine di 10^{24} per ogni mole). La variabile aleatoria X che descrive il numero di decadimenti in un secondo ha dunque, approssimativamente, legge di Poisson di parametro $\lambda = Np$. Se, ad esempio, per un certo materiale si scopre sperimentalmente che è $\lambda = 3.2$, allora la probabilità che esso emetta 2 particelle alfa è

$$P(X = 2) = e^{-3.2} \frac{(3.2)^2}{2} \approx 0.21.$$

Questi esempi sono interessanti perché mostrano che in varie situazioni concrete bastano alcune semplici ipotesi per riuscire a stabilire quale sia la natura della legge delle variabili osservate.

1.2.3 La speranza di una variabile aleatoria discreta

Data, su uno spazio probabilizzato (Ω, \mathcal{A}, P) , una variabile aleatoria discreta X , e indicato con E l'insieme (finito o numerabile) dei valori assunti da X , si dice che essa è *integrabile* se il numero

$$\sum_{k \in E} |k| P(X = k)$$

è finito. In tal caso, si chiama *speranza* (o *media*) di X il numero

$$\mathbb{E}[X] = \sum_{k \in E} k P(X = k). \quad (1.17)$$

La notazione $\mathbb{E}[X]$ (in cui la lettera \mathbb{E} ricorda le parole *espérance*, *Erwartung*, *expectation* usate in francese, tedesco e inglese per indicare la speranza) ha il grave difetto di non far apparire la misura di probabilità secondo la quale si considera la legge di X . Questo non è un grande svantaggio quando si abbia a che fare con un'unica misura di probabilità, ma specialmente nei casi in cui si vuole condizionare rispetto a qualche evento (o, come vedremo, rispetto a qualche variabile aleatoria) questo potrebbe causare qualche disagio. Tuttavia, nei casi in cui ciò potrebbe creare delle ambiguità, si ricorrerà alle notazioni $\mathbb{E}_P[X]$, $\mathbb{E}^P[X]$ o ad altre consimili, al fine di limitare il disagio.

Osserviamo immediatamente che la speranza non è altro che la somma dei valori che una variabile aleatoria discreta può prendere, moltiplicati per la probabilità con cui questi valori vengono assunti. Poiché la somma di queste probabilità è 1, essa è dunque la *media ponderata* dei valori k assunti: il generico valore k interviene nella media col peso $P(X = k)$.

Esiste un modo molto naturale di passare dalla probabilità di un evento alla speranza di una variabile aleatoria: basta utilizzare una particolare funzione che introduciamo brevemente. Assegnato a questo scopo un evento A (elemento della tribù \mathcal{A}), la funzione, definita su Ω , che assume il valore 1 in tutti i punti di A e il valore 0 in tutti i punti di A^c si chiama la *funzione indicatrice* (o, semplicemente, l'*indicatrice*) di A , e si denota con il simbolo I_A . Si ha cioè, per definizione:

$$I_A(\omega) = \begin{cases} 1 & \text{se } \omega \in A, \\ 0 & \text{se } \omega \notin A. \end{cases}$$

È immediato riconoscere che si tratta di una variabile aleatoria discreta integrabile. Inoltre, dalla definizione di speranza si ha

$$\mathbb{E}[c I_A] = c P(A),$$

per ogni evento A e ogni numero reale c . Inoltre, se X è una variabile aleatoria discreta, e se si denota con E l'insieme (finito o numerabile) dei valori assunti da X , allora essa

si può scrivere come combinazione lineare di indicatori nel modo seguente:

$$X = \sum_{k \in E} k I_{\{X=k\}}.$$

Questa formula mostra che la definizione di speranza è una sorta di generalizzazione della misura di probabilità: si passa dagli eventi, cioè dalle funzioni indicatori, a tutte le variabili aleatorie discrete.

Ritornando al caso generale di una variabile aleatoria discreta, osserviamo adesso che, evidentemente, ciascuna variabile aleatoria discreta integrabile *quasi certamente positiva*, cioè tale che l'evento $\{X \geq 0\}$ abbia misura 1 secondo P , ha speranza positiva. Inoltre, due variabili aleatorie discrete integrabili X, Y , che siano equivalenti secondo P , cioè tali che l'insieme $\{X = Y\}$ contenga un evento quasi certo, hanno la stessa legge, dunque hanno la stessa speranza.

Sussiste inoltre, per la speranza, la seguente proprietà di *linearità*, che ci contenteremo di enunciare senza dimostrazione. Se X_1, \dots, X_n sono variabili aleatorie discrete integrabili, definite sullo stesso spazio probabilizzato, e se a_1, \dots, a_n sono numeri reali, allora la funzione $Y = a_1 X_1 + \dots + a_n X_n$ è ancora una variabile aleatoria discreta integrabile, e la sua speranza è data dalla formula

$$\mathbb{E}[Y] = a_1 \mathbb{E}[X_1] + \dots + a_n \mathbb{E}[X_n].$$

Inoltre, per ogni coppia X, Y di variabili aleatorie discrete integrabili, definite sullo stesso spazio probabilizzato, con $X \leq Y$, si ha $\mathbb{E}[X] \leq \mathbb{E}[Y]$. In effetti, poiché la differenza $Y - X$ è una variabile aleatoria discreta, integrabile e positiva, si ha, grazie alla linearità della speranza

$$\mathbb{E}[Y] - \mathbb{E}[X] = \mathbb{E}[Y - X] \geq 0.$$

Questa proprietà si chiama l'*isotonia* della speranza.

Sempre su un fissato spazio probabilizzato (Ω, \mathcal{A}, P) , sia X una variabile aleatoria discreta, a valori nell'insieme finito E , e sia ϕ una funzione di E in \mathbb{R} . Allora, affinché la variabile aleatoria $\phi(X)$ sia integrabile occorre e basta che il numero $\sum_{k \in E} |\phi(k)| P(X = k)$ sia finito; inoltre, se questa condizione è soddisfatta, vale la seguente legge, che qualcuno chiama **lotus**, acronimo di *Law of the Unconscious Statistician*:

$$\mathbb{E}[\phi(X)] = \sum_{k \in E} \phi(k) P(X = k). \quad (1.18)$$

Supponiamo ora che X sia una variabile aleatoria discreta, definita su uno spazio probabilizzato (Ω, \mathcal{A}, P) , a valori nell'insieme \mathbb{N} degli interi naturali. Si ha allora la seguente utile espressione per la speranza (la cui dimostrazione è una conseguenza non troppo semplice delle proprietà "generalì" della speranza, che vedremo nel paragrafo 1.2.5):

$$\mathbb{E}[X] = \sum_{n \in \mathbb{N}} P(X > n). \quad (1.19)$$

Per dimostrarla, basterà infatti osservare che si ha:

$$X = \sum_{n \in \mathbb{N}} I_{\{X > n\}},$$

ossia, per ogni elemento ω di Ω ,

$$X(\omega) = \sum_{n \in \mathbb{N}} I_{\{X > n\}}(\omega).$$

Infatti, il secondo membro di quest'ultima eguaglianza è la somma di tanti termini, tutti eguali a 1, quanti sono gl'indici n verificanti la condizione $0 \leq n < X(\omega)$: e il numero di questi indici è appunto $X(\omega)$. A questo punto, per ottenere la proprietà (1.19) basta osservare che

$$\mathbb{E}[X] = \sum_{n \in \mathbb{N}} \mathbb{E}[I_{\{X > n\}}],$$

e che

$$\mathbb{E}[I_{\{X > n\}}] = \sum_{k \in \mathbb{N}} k P(I_{\{X > n\}} = k) = P(I_{\{X > n\}} = 1) = P(X > n).$$

Esempio 1.2.11 Calcoliamo la speranza delle variabili aleatorie dotate delle leggi discrete illustrate nel paragrafo 1.2.2.

Legge uniforme: per una variabile aleatoria dotata di legge $\mathcal{U}(N)$ si ha $P(X = k) = \frac{1}{N}$ per ogni k compreso fra 1 e N : quindi

$$\mathbb{E}[X] = \sum_{k=1}^N k P(X = k) = \frac{1}{N} \sum_{k=1}^N k = \frac{1}{N} \frac{N(N+1)}{2} = \frac{N+1}{2}.$$

Legge di Bernoulli: sappiamo che per una variabile aleatoria X dotata di legge $\mathcal{B}(p)$ si ha $P(X = 1) = p$ e $P(X = 0) = 1 - p$; ne segue

$$\mathbb{E}[X] = \sum_{k \in \mathbb{N}} k P(X = k) = P(X = 1) = p.$$

Legge binomiale: per una variabile aleatoria X dotata di legge $\mathcal{B}(n, p)$ si ha da (1.16)

$$\mathbb{E}[X] = \sum_{k \in \mathbb{N}} k P(X = k) = \sum_{k=1}^n k \binom{n}{k} p^k (1-p)^{n-k},$$

essendo $k \binom{n}{k} = n \binom{n-1}{k-1}$, si deduce, ponendo $h = k - 1$:

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

Legge geometrica: sia T una variabile aleatoria dotata di legge $\mathcal{G}(p)$: risulta dunque $P(T = n) = p(1 - p)^{n-1}$. Si ha allora:

$$\begin{aligned}\mathbb{E}[T] &= \sum_{n=1}^{\infty} np(1-p)^{n-1} = p \sum_{n=1}^{\infty} n(1-p)^{n-1} = -p \sum_{n=1}^{\infty} \frac{d}{dp} (1-p)^n = \\ &= -p \frac{d}{dp} \sum_{n=0}^{\infty} (1-p)^n = -p \frac{d}{dp} \frac{1}{p} = \frac{1}{p}.\end{aligned}$$

Legge di Poisson: per una variabile aleatoria dotata di legge $\mathcal{P}(\lambda)$ vale la relazione $P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}$, e dunque

$$\mathbb{E}[X] = \sum_{k=1}^{\infty} \frac{\lambda^k}{(k-1)!} e^{-\lambda} = \lambda \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} e^{-\lambda} = \lambda \sum_{h=0}^{\infty} \frac{\lambda^h}{h!} e^{-\lambda} = \lambda.$$

1.2.4 Variabili aleatorie continue, densità di una legge, funzione di ripartizione

Come abbiamo già detto, in moltissime situazioni una variabile aleatoria discreta non permette di rappresentare il risultato di un esperimento aleatorio: questo accade, ad esempio, quando si voglia scegliere un punto su un segmento (esempio 1.1.11). Sarà allora opportuno, in certi casi, pretendere che una variabile aleatoria possa assumere un insieme continuo di valori (per esempio, tutti i valori reali, oppure tutti i valori di un determinato intervallo della retta reale). Una variabile aleatoria che non sia discreta, cioè che non assuma solo un numero finito o numerabile di valori, si dice una variabile aleatoria *reale*. Lo studio generale delle variabili aleatorie reali può essere molto complicato; tuttavia vi è un caso particolare, ma anche particolarmente importante e comunque molto comune, in cui questo studio è parecchio semplificato: il caso delle *variabili aleatorie continue*.

A questo scopo, supponiamo assegnata, su uno spazio probabilizzato (Ω, \mathcal{A}, P) , una variabile aleatoria reale X . Se esiste una funzione positiva f_X , tale che, per ogni intervallo I di \mathbb{R} , abbia senso l'integrale di f_X su I e risulti

$$P(X \in I) = \int_I f_X(x) dx, \quad (1.20)$$

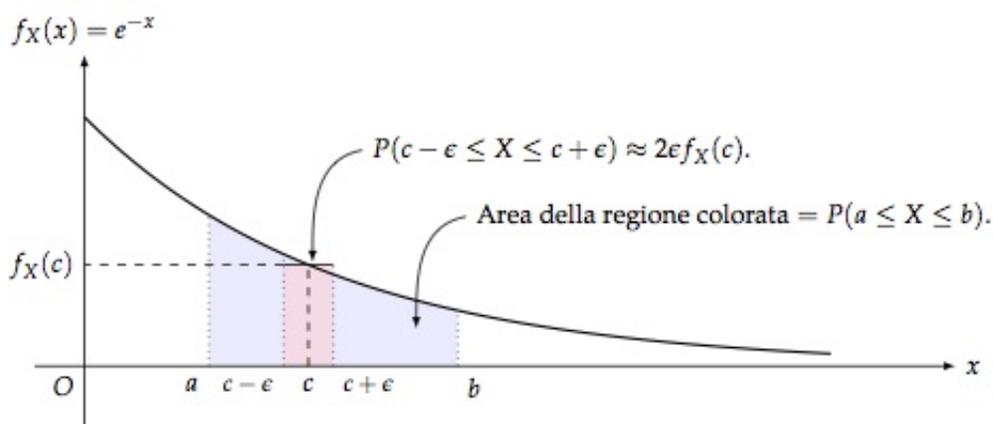
si dice che X è una variabile aleatoria *dotata di densità* o, più semplicemente, *continua*, e la funzione f_X si chiama una *densità* di X (secondo P). Dalla relazione (1.20) discende dunque che, per una variabile aleatoria X dotata di densità f_X , il calcolo delle probabilità di eventi del tipo $\{a \leq X \leq b\}$ si riconduce al calcolo di un integrale ovvero di un'area. In particolare, la funzione f_X deve soddisfare alla relazione $\int_{-\infty}^{+\infty} f_X(x) dx = 1$, perché questo integrale coincide con la probabilità $P(\Omega)$. Inoltre, le regioni in cui f_X assume valori grandi sono le regioni nelle quali X prende valori con probabilità più elevata. Per riconoscerlo, osserviamo che, se ϵ è un qualsiasi numero reale positivo, si può approssimare la probabilità $P(c - \epsilon \leq X \leq c + \epsilon)$ attraverso il teorema della media

integrale nel modo seguente:

$$P(c - \epsilon \leq X \leq c + \epsilon) = \int_{c-\epsilon}^{c+\epsilon} f_X(x) dx \approx 2\epsilon f_X(c),$$

dalla quale si deduce che la probabilità che X cada in un intorno di c di ampiezza 2ϵ è approssimativamente uguale a $2\epsilon f_X(c)$, dunque direttamente proporzionale a $f_X(c)$; dunque quest'ultimo numero si può ragionevolmente interpretare come un'indicazione di quanto X possa cadere vicino al numero c .

Inoltre, osserviamo che una variabile aleatoria dotata di densità è sempre *diffusa*, nel senso che ogni evento della forma $\{X = c\}$ è trascurabile (secondo P): infatti la probabilità di un tal evento si può sempre scrivere nella forma $P(X = c) = \int_c^c f_X(x) dx = 0$.



Il grafico rappresenta la densità $f_X(x) = e^{-x}$. L'area colorata è pari alla probabilità che la variabile aleatoria X prenda valori nell'intervallo $I = [a, b]$. L'area della regione in rosso, per $\epsilon \rightarrow 0^+$ approssima dunque la probabilità che la variabile aleatoria X assuma valori vicini a c .

Esempio 1.2.12 Per ogni numero reale positivo λ , si consideri la funzione f definita su \mathbb{R} da

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

Come subito si riconosce, la funzione f è sempre positiva ed inoltre

$$\int_{-\infty}^{+\infty} f(x) dx = \int_0^{+\infty} \lambda e^{-\lambda x} dx = 1.$$

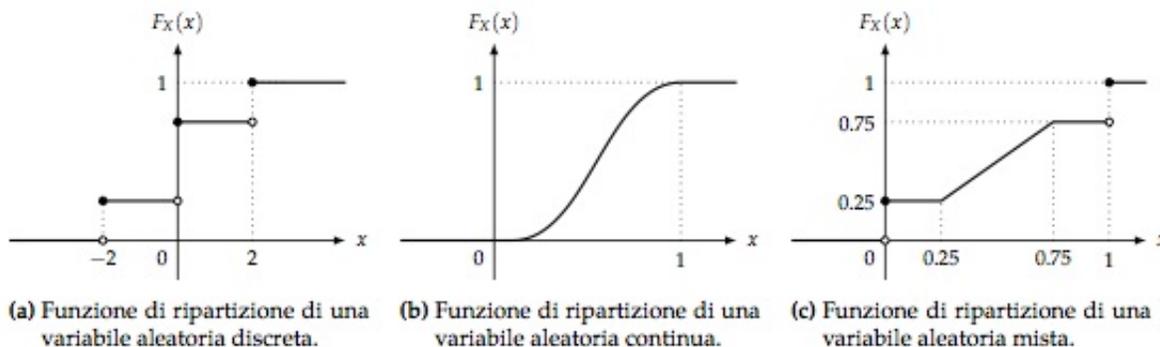
Essa rappresenta dunque la densità di una variabile aleatoria reale quasi certamente positiva, nel senso che può assumere con probabilità non nulla soltanto valori positivi. Una variabile aleatoria dotata di questa densità si dice *esponenziale* di parametro λ e la sua legge si denota con il simbolo $\mathcal{E}(\lambda)$.

Introduciamo un'utile definizione del tutto generale che, tuttavia, sarà molto utile, specialmente nella pratica, nel caso di variabili aleatorie reali dotate di densità.

Definizione 1.2.13 Sia X una variabile aleatoria definita su uno spazio probabilizzato (Ω, \mathcal{A}, P) . Si chiama *funzione di ripartizione* di X la funzione F_X , di \mathbb{R} in $[0, 1]$, così definita:

$$F_X(x) = P(X \leq x) \quad \text{per ogni } x \in \mathbb{R}.$$

Si riconosce immediatamente che la funzione di ripartizione di una variabile aleatoria X è crescente, che $\lim_{x \rightarrow -\infty} F_X(x) = 0$, $\lim_{x \rightarrow +\infty} F_X(x) = 1$ e, un po' meno facilmente, che essa è continua a destra (nel senso che, per ogni numero reale c , si ha $\lim_{x \rightarrow c^+} F_X(x) = F_X(c)$). In generale, tuttavia, la funzione F_X non è invertibile. L'invertibilità è però assicurata, ad esempio, quando la variabile aleatoria X sia dotata di densità.



Osserviamo subito che tutti i problemi che si pongono per una variabile aleatoria reale si possono risolvere utilizzando la relativa funzione di ripartizione. Ad esempio, volendo calcolare la probabilità $P(a < X \leq b)$, basta osservare che l'evento $\{X \leq b\}$ è l'unione dei due eventi incompatibili $\{X \leq a\}$ e $\{a < X \leq b\}$. Pertanto, dall'additività della misura P , risulta immediatamente,

$$P(X \leq b) = P(X \leq a) + P(a < X \leq b),$$

da cui

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

Inoltre, se la legge della variabile aleatoria X è *diffusa* (cioè se ogni evento della forma $\{X = c\}$ è trascurabile secondo P), allora quest'eguaglianza vale anche per $P(a \leq X \leq b)$, giacché essa si può sempre scrivere (per la proprietà di additività di P) nella forma

$$P(a \leq X \leq b) = P(X = a) + P(a < X \leq b).$$

Se, adesso, X è una variabile aleatoria reale dotata di densità f_X , dalla definizione di funzione di ripartizione e dalla (1.20) segue immediatamente la seguente importante relazione che lega la funzione di ripartizione alla densità:

$$F_X(x) = \int_{-\infty}^x f_X(t) dt \quad \text{per ogni } x \in \mathbb{R}.$$

Se la densità f_X è una funzione continua, tale relazione si può invertire derivando ambo i membri, ottenendo così la relazione che permette, nota la funzione di ripartizione, di trovare la densità della variabile aleatoria X :

$$\frac{dF_X}{dx} = f_X(x).$$

Quest'ultima relazione esprime, in particolare, un fatto ancora più importante e profondo: se una variabile aleatoria X ammette una funzione di ripartizione ovunque derivabile, allora essa è dotata di densità e tale densità è proprio la derivata prima della funzione di ripartizione.

Esempio 1.2.14 Su uno spazio probabilizzato (Ω, \mathcal{A}, P) , sia X una variabile aleatoria la cui funzione di ripartizione è

$$F_X(x) = \begin{cases} \frac{1}{2}e^x & \text{per } x < 0, \\ 1 - \frac{1}{2}e^{-x} & \text{per } x \geq 0. \end{cases}$$

Si tratta di una funzione continua e derivabile in ogni punto eccetto $x = 0$. Dunque la variabile aleatoria X ammette come densità la funzione

$$f_X(x) = \frac{dF_X}{dx} = \begin{cases} \frac{1}{2}e^x & \text{per } x < 0, \\ \frac{1}{2}e^{-x} & \text{per } x \geq 0, \end{cases}$$

ovvero

$$f_X(x) = \frac{1}{2}e^{-|x|}.$$

1.2.5 Il concetto generale di speranza

La nozione di speranza vista nel paragrafo 1.2.3 non permette, ovviamente, di calcolare la speranza per le variabili aleatorie reali in generale, perché, limitandoci anche soltanto al caso delle variabili aleatorie reali dotate di densità, la somma (1.17) si ridurrebbe ad una somma di una quantità più che numerabile di addendi nulli, giacché la legge di una siffatta variabile aleatoria è diffusa e dunque la probabilità $P(X = x)$ risulterebbe trascurabile. D'altra parte, sarà comodo poter definire una nozione di speranza più generale, tanto da poter essere applicata in tutte le situazioni in cui le variabili aleatorie in questione non siano necessariamente discrete. Ovviamente, perché questa estensione si possa chiamare a sua volta “speranza”, essa si dovrà ridurre a quella già precedentemente definita nel paragrafo 1.2.3 quando le variabili aleatorie considerate siano discrete. Faremo questo “assiomaticamente”, cioè semplicemente dichiarando l'esistenza della speranza fornendone soltanto le proprietà che la caratterizzano, senza dilungarci nel tecnicismo di doverla costruire concretamente; lasceremo dunque all'analisi matematica il compito di dimostrare, nella maniera più astratta e generale, la sua esistenza e la sua unicità.

Fissato uno spazio probabilizzato (Ω, \mathcal{A}, P) , è possibile costruire, nella classe di tutte le variabili aleatorie *positive*, un'applicazione $X \mapsto \mathbb{E}[X]$ (detta *operatore di speranza*) che ad ogni variabile aleatoria reale positiva X associa un numero $\mathbb{E}[X]$ (detto, appunto, la *speranza* di X), non necessariamente finito, in modo tale che valgano le seguenti proprietà:

1. Se $X = I_A$ è l'indicatrice di un evento A appartenente ad \mathcal{A} , si ha $\mathbb{E}[X] = P(A)$.
2. Per ogni variabile aleatoria reale positiva X e per ogni numero reale $c \geq 0$ si ha $\mathbb{E}[cX] = c\mathbb{E}[X]$.
3. Se X, Y sono due variabili aleatorie reali positive, si ha $\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$.

4. Se $(X_n)_{n \in \mathbb{N}}$ è una successione crescente di variabili aleatorie positive, convergente puntualmente verso una variabile aleatoria X (necessariamente positiva), cioè se, per ciascuna eventualità ω in Ω , risulta $X_n(\omega) \uparrow X(\omega)$ per $n \rightarrow +\infty$, allora si ha $\mathbb{E}[X_n] \uparrow \mathbb{E}[X]$ per $n \rightarrow +\infty$.

Data ora una variabile aleatoria reale X , si chiama la *parte positiva* di X , e si denota con X^+ , quella variabile aleatoria positiva che coincide con X sull'evento $\{X \geq 0\}$ e con 0 altrove; si chiama invece la *parte negativa* di X , e si denota con X^- , quella variabile aleatoria positiva che coincide con $-X$ sull'evento $\{X \leq 0\}$ e con 0 altrove. Diremo allora che una variabile aleatoria X è *integrabile* se sono finiti entrambi i numeri $\mathbb{E}[X^+]$ e $\mathbb{E}[X^-]$, ed in tal caso si pone:

$$\mathbb{E}[X] = \mathbb{E}[X^+] - \mathbb{E}[X^-].$$

Il numero $\mathbb{E}[X]$ si chiama allora la *speranza* (o la *media*) di X (secondo P). Si riconosce immediatamente, ricorrendo alla definizione, che l'operatore di speranza, nella classe formata da tutte le variabili aleatorie integrabili, gode delle proprietà di "linearità" ed "isotonia" (proprio come accadeva nel caso discreto). Questa definizione estende l'omonima definizione data nel paragrafo 1.2.3. Per convincersi di ciò, grazie alla decomposizione $X = X^+ - X^-$, è sufficiente considerare una variabile aleatoria discreta X che sia positiva, ossia a valori in una parte E (finita o numerabile) di \mathbb{R}_+ . Una siffatta variabile aleatoria si può scrivere nella forma

$$X = \sum_{k \in E} k I_{\{X=k\}},$$

e di qui, utilizzando le proprietà appena elencate per la speranza, se ne deduce facilmente la formula (1.17). D'altra parte, se adesso la variabile aleatoria reale X ammette f_X come densità, allora si può dimostrare che X è integrabile se e soltanto se l'integrale $\int_{\mathbb{R}} |x| f_X(x) dx$ è finito, e in tal caso la speranza di X si può esprimere per mezzo della densità f tramite la formula:

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

Più in generale, si può estendere anche la legge (1.18) (che abbiamo chiamato **lotus**) anche nel caso delle variabili aleatorie continue. Precisamente, se X ammette f come densità, allora, per ogni variabile aleatoria della forma $\phi(X)$, con ϕ funzione continua di \mathbb{R} in \mathbb{R} , affinché la variabile aleatoria $\phi(X)$ sia integrabile occorre e basta che l'integrale $\int_{\mathbb{R}} |\phi(x)| f_X(x) dx$ sia finito, e se questa condizione è soddisfatta, vale la relazione, che continueremo a chiamare con l'acronimo di **lotus**:

$$\mathbb{E}[\phi(X)] = \int_{\mathbb{R}} \phi(x) f_X(x) dx. \quad (1.21)$$

Esempio 1.2.15 (La ripartizione uniforme) Sia $A = [a, b]$ un intervallo, e poniamo $c = b - a$. Una variabile aleatoria X , definita su di un opportuno spazio probabilizzato, avente come densità la funzione $c^{-1}I_A$, si dirà una variabile aleatoria *uniformemente*

ripartita su A , e la sua legge si chiamerà la *ripartizione uniforme* su A . La speranza di una siffatta variabile aleatoria X è $\mathbb{E}[X] = c^{-1} \int_a^b x dx = (a+b)/2$, cioè il punto medio dell'intervallo $[a, b]$.

Come abbiamo detto, un sinonimo del termine “speranza” è “media”. Talvolta si utilizzano anche locuzioni come “valore atteso” oppure “valore di aspettazione”, traduzioni più o meno grossolane dalla lingua inglese dei termini *mean value* e *expected value* utilizzati per descrivere la speranza. Un'altra locuzione utilizzata per la nozione di speranza è *momento del prim'ordine*, con riferimento alla definizione che segue.

Definizione 1.2.16 Su uno spazio probabilizzato (Ω, \mathcal{A}, P) , sia X una variabile aleatoria reale e sia r un numero intero positivo. La quantità $\mathbb{E}[X^r]$, quando esiste finita, si chiama il *momento di ordine r* di X e si denoterà con $m_r(X)$.

Nelle ipotesi della definizione precedente, se X è una variabile aleatoria discreta, dotata di densità discreta p_X , la formula (1.18) permette di calcolare il momento di ordine r riducendosi alla forma:

$$m_r(X) = \sum_{k \in E} k^r p_X(k).$$

Se, invece, X è una variabile aleatoria reale dotata di densità f_X , allora la formula per il calcolo del momento di ordine r diventerà, in base alla relazione (1.21),

$$m_r(X) = \int_{\mathbb{R}} x^r f_X(x) dx.$$

1.2.6 Vettori aleatori, leggi congiunte e leggi marginali

Nello studio di un esperimento aleatorio ci sono molte situazioni in cui può capitare che il risultato dell'esperimento produca più di una variabile interessante, oppure che si vogliano studiare le relazioni tra due o più grandezze legate all'esito dell'esperimento. Per esempio, supponiamo che un individuo abbia un'urna contenente sei palline rosse (che, convenzionalmente, indicheremo con 1) e quattro palline bianche (che, convenzionalmente, indicheremo con 0). Egli ne estragga una, la metta da parte, e ne estragga una seconda. Indichiamo dunque con X e Y le variabili aleatorie descritte dalle parole “il colore della prima pallina” e “il colore della seconda pallina”. In questo caso, una domanda interessante potrebbe essere: qual è la probabilità che escano due palline rosse? In simboli si tratterà di calcolare la “probabilità congiunta” $P(X=1, Y=1)$. Ora, se le due variabili fossero indipendenti, la risposta sarebbe ovvia. Per calcolare questa probabilità anche nel caso di variabili aleatorie dipendenti (come in questo esempio), occorre introdurre una generalizzazione della nozione di variabile aleatoria: i cosiddetti “vettori aleatori”.

Definizione 1.2.17 In tutta generalità, sia dato uno spazio probabilizzato (Ω, \mathcal{A}, P) e, su di esso, siano definite n variabili aleatorie reali X_1, \dots, X_n . Si chiamerà allora *vettore aleatorio* la funzione $\mathbf{X} = (X_1, \dots, X_n)$, definita su Ω , a valori in \mathbb{R}^n , da

$$\mathbf{X}(\omega) = (X_1(\omega), \dots, X_n(\omega)) \quad \text{per ogni } \omega \in \Omega.$$

Dato adesso un vettore aleatorio $\mathbf{X} = (X_1, \dots, X_n)$, le variabili aleatorie X_1, \dots, X_n si chiameranno le sue *componenti*. Inoltre, come di consueto nell'ambito geometrico, quando sia $n = 2$ oppure $n = 3$, in luogo di (X_1, X_2) e (X_1, X_2, X_3) , le componenti di un generico vettore aleatorio si indicheranno di preferenza con (X, Y) ovvero (X, Y, Z) .

Generalizzando quanto abbiamo detto a proposito delle variabili aleatorie reali, se si chiama *tribù boreliana* di \mathbb{R}^n , e si denota con $\mathcal{B}(\mathbb{R}^n)$, la più piccola tra tutte le tribù su \mathbb{R}^n che contengano gli insiemi della forma $I_1 \times \dots \times I_n$ con I_1, \dots, I_n intervalli di \mathbb{R} (oppure, equivalentemente, boreliani di \mathbb{R}) e se si conviene di chiamare *boreliano* di \mathbb{R}^n ogni elemento di questa tribù, si può dimostrare che, se $\mathbf{X} = (X_1, \dots, X_n)$ è un vettore aleatorio e se A è un qualsiasi boreliano di \mathbb{R}^n , allora l'insieme

$$\{\mathbf{X} \in A\} = \{\omega \in \Omega : (X_1(\omega), \dots, X_n(\omega)) \in A\}$$

è un evento, ossia appartiene alla tribù \mathcal{A} . Pertanto, sempre in analogia con il caso delle variabili aleatorie reali, la funzione che ad ogni boreliano di \mathbb{R}^n associa il numero $P(\mathbf{X} \in A)$ si chiamerà la *legge* del vettore aleatorio \mathbf{X} secondo P e, come subito si riconosce, si tratta di una misura di probabilità sulla tribù boreliana di \mathbb{R}^n , proprio come la legge di una variabile aleatoria reale è una misura di probabilità sulla tribù boreliana di \mathbb{R} .

Una volta definiti i vettori aleatori in tutta generalità, sempre in analogia a quanto abbiamo fatto per le variabili aleatorie reali, analizziamo dapprima il caso di un vettore aleatorio discreto. Supponiamo cioè che $\mathbf{X} = (X_1, \dots, X_n)$, definito sullo spazio probabilizzato (Ω, \mathcal{A}, P) , ammetta una quantità finita o numerabile di valori possibili e indichiamo con E l'insieme di questi possibili valori. Equivalentemente, ciò significa che le variabili aleatorie X_1, \dots, X_n sono discrete e che E è un sottoinsieme del prodotto $E_1 \times \dots \times E_n$ dove, per ciascun indice i , E_i denota l'insieme dei possibili valori assunti dalla variabile aleatoria discreta X_i . In tal caso, la funzione $p_{\mathbf{X}}$ di E in $[0, 1]$ definita per ogni elemento $\mathbf{x} = (x_1, \dots, x_n)$ di E , da

$$p_{\mathbf{X}}(\mathbf{x}) = P(\mathbf{X} = \mathbf{x}) = P(X_1 = x_1, \dots, X_n = x_n)$$

prende il nome di *densità discreta* del vettore aleatorio \mathbf{X} oppure di *densità discreta congiunta* delle variabili aleatorie X_1, \dots, X_n . Evidentemente, se le variabili aleatorie X_1, \dots, X_n fossero indipendenti, questa sarebbe banalmente il prodotto delle loro densità discrete, giacché il terzo membro della precedente uguaglianza si potrebbe scrivere, equivalentemente, nella forma $P(X_1 = x_1) \dots P(X_n = x_n)$. Tuttavia, quando invece le variabili aleatorie non sono indipendenti tra loro, la densità discreta congiunta non si potrà scrivere direttamente per mezzo delle densità discrete delle componenti, che prenderanno il nome di densità discrete *marginali*. Si può tuttavia vedere immediatamente che, se si conosce la densità congiunta delle variabili aleatorie X_1, \dots, X_n , da essa si possono ricavare le marginali. Mostriamolo, per semplicità di scrittura, soltanto nel caso di due variabili aleatorie discrete X e Y a valori negli insiemi discreti E e F rispettivamente. Il passaggio ad un arbitrario numero finito di variabili aleatorie si farà poi con una generalizzazione elementarissima. In questo caso, se si denotano con $p_{X,Y}$ la densità congiunta di X e Y e con p_X e p_Y le due densità marginali, basterà osservare

che l'evento $\{X = x\}$ si potrà senz'altro scrivere nella forma $\bigcup_{y \in F} \{X = x, Y = y\}$ e di qui, utilizzando l'additività della misura P , si trae immediatamente

$$\begin{aligned} p_X(x) &= P(X = x) = P\left(\bigcup_{y \in F} \{X = x, Y = y\}\right) \\ &= \sum_{y \in F} P(X = x, Y = y) = \sum_{y \in F} p_{X,Y}(x, y). \end{aligned}$$

Analogamente, per Y , si avrà:

$$p_Y(y) = \sum_{x \in E} p_{X,Y}(x, y).$$

Esempio 1.2.18 Da un'urna contenente 6 palline numerate da 1 a 6, se ne estraggono due in sequenza: indichiamo con X e con Y i risultati della prima e della seconda estrazione. Per descrivere questo esperimento aleatorio, si potrà prendere lo spazio probabilizzato così composto:

1. Come *insieme delle eventualità* si prenderà l'insieme Ω formato da tutte le possibili coppie d'interi compresi tra 1 e 6 con coordinate diverse tra loro. Naturalmente, è da intendere che la generica di queste coppie $\omega = (\omega_1, \omega_2)$ rappresenti il risultato che si ottiene quando il risultato della prima estrazione è ω_1 e il risultato della seconda estrazione è ω_2 .
2. Come *tribù degli eventi* si potrà tranquillamente scegliere la tribù di tutte le parti di Ω .
3. Come *misura di probabilità* P da mettere sulla tribù degli eventi (cioè sull'insieme delle parti di Ω) sarà naturale scegliere la ripartizione uniforme.
4. Le variabili aleatorie discrete X e Y , a valori in $\{1, \dots, 6\}$ saranno dunque le proiezioni sulla prima e sulla seconda coordinata rispettivamente, cioè le funzioni $X(\omega) = \omega_1$ e $Y(\omega) = \omega_2$.

La densità discreta congiunta di queste due variabili aleatorie sarà, allora, tenuto conto che l'insieme Ω è costituito da 30 elementi,

$$p_{X,Y}(x, y) = P(X = x, Y = y) = \frac{1}{30},$$

mentre le densità marginali di X e Y saranno, evidentemente,

$$p_X(x) = \sum_{y \neq x} P(X = x, Y = y) = 5 \cdot \frac{1}{30} = \frac{1}{6},$$

cioè la ripartizione uniforme sull'insieme $\{1, \dots, 6\}$, e così pure per la densità discreta di Y .

Passiamo adesso ad analizzare il caso di un vettore aleatorio continuo. In questo caso, la definizione di densità congiunta sarà un po' più complicata perché si dovrà tenere conto dei problemi d'integrabilità delle funzioni di più variabili. Ancora, la definizione procede in completa analogia con la sua analoga nel caso di una variabile aleatoria continua.

Definizione 1.2.19 Su un dato spazio probabilizzato (Ω, \mathcal{A}, P) , sia $\mathbf{X} = (X_1, \dots, X_n)$ un vettore aleatorio. Diremo che esso ammette una *densità*, ovvero che le variabili aleatorie X_1, \dots, X_n ammettono una *densità congiunta*, se esiste una funzione reale $f_{\mathbf{X}}$ di n variabili reali avente la proprietà che, per ogni boreliano A di \mathbb{R}^n , sia

$$P(\mathbf{X} \in A) = \int_A f_{\mathbf{X}}(\mathbf{x}) \, d\mathbf{x} = \int_A f_{\mathbf{X}}(x_1, \dots, x_n) \, dx_1 \cdots dx_n.$$

Esattamente come nel caso delle variabili aleatorie discrete, inoltre, se X_1, \dots, X_n ammettono la densità congiunta $f_{\mathbf{X}}$, allora esse sono dotate di densità, dette *marginali*, e tali densità si possono ricavare dalla densità congiunta. Precisamente, limitandoci ancora, per semplicità di scrittura, al caso di due variabili aleatorie X e Y , e riferendoci per esempio alla variabile aleatoria X , basterà osservare che, se A è un boreliano di \mathbb{R} , l'evento $\{X \in A\}$ si può scrivere nella forma $\{X \in A, Y \in \mathbb{R}\}$ e di qui, per la definizione di densità congiunta, si trova immediatamente

$$P(X \in A) = \iint_{A \times \mathbb{R}} f_{X,Y}(x, y) \, dx dy = \int_A \left[\int_{\mathbb{R}} f_{X,Y}(x, y) \, dy \right] dx$$

da cui segue immediatamente che

$$f_X(x) = \int_{\mathbb{R}} f_{X,Y}(x, y) \, dy.$$

Analogamente, per la variabile aleatoria Y si avrà

$$f_Y(y) = \int_{\mathbb{R}} f_{X,Y}(x, y) \, dx.$$

Osserviamo adesso che, se le due variabili aleatorie X e Y sono indipendenti, allora la probabilità $P(X \in A, Y \in B)$ si può scrivere nella forma equivalente $P(X \in A)P(Y \in B)$ e di qui, dette f_X e f_Y le marginali delle due variabili aleatorie X e Y , si traggono le due uguaglianze

$$\begin{aligned} P(X \in A, Y \in B) &= \iint_{A \times B} f_{X,Y}(x, y) \, dx dy \\ P(X \in A)P(Y \in B) &= \int_A f_X(x) \, dx \int_B f_Y(y) \, dy = \iint_{A \times B} f_X(x)f_Y(y) \, dx dy, \end{aligned}$$

da cui segue immediatamente, per l'arbitrarietà di A e B , la relazione

$$f_{X,Y}(x, y) = f_X(x)f_Y(y).$$

In altre parole, quando due (o più) variabili aleatorie continue sono indipendenti la loro densità congiunta è il prodotto delle loro densità marginali.

Esempio 1.2.20 Su uno spazio probabilizzato (Ω, \mathcal{A}, P) , siano X e Y due variabili aleatorie reali la cui densità congiunta sia data da:

$$f_{X,Y}(x, y) = \begin{cases} 2e^{-x}e^{-2y} & \text{per } x > 0 \text{ e } y > 0, \\ 0 & \text{altrimenti.} \end{cases}$$

Calcoliamo la probabilità $P(X > 1, Y < 1)$. A questo scopo, sarà sufficiente calcolare l'integrale della densità congiunta $f_{X,Y}$ nella regione definita da $x > 1$ e $y < 1$; ma la seconda diseuguaglianza si riduce a $0 < y < 1$ giacché la funzione $f_{X,Y}$ è nulla per $y < 0$. Si ha dunque:

$$\begin{aligned} P(X > 1, Y < 1) &= \int_0^1 \left[\int_1^{+\infty} 2e^{-x}e^{-2y} dx \right] dy \\ &= \left(\int_1^{+\infty} e^{-x} dx \right) \cdot \left(\int_0^1 2e^{-2y} dy \right) \\ &= [-e^{-x}]_1^{+\infty} \cdot [-e^{-2y}]_0^1 = e^{-1}(1 - e^{-2}). \end{aligned}$$

Calcoliamo adesso la probabilità $P(X < Y)$. In questo caso, l'integrale dovrà essere fatto nella regione $x < y$. Gli estremi d'integrazione che corrispondono a questo dominio possono essere scelti in due modi: o si integra internamente rispetto alla variabile x tra gli estremi 0 e y (infatti $x > 0$ altrimenti la funzione $f_{X,Y}$ è nulla, mentre $x < y$ è la definizione della regione che stiamo considerando) e quindi si integra quanto trovato rispetto a y tra 0 e $+\infty$ (infatti, basta porre la condizione $x < y$ sull'integrale interno), oppure si integra internamente rispetto a y tra x e ∞ (per rispettare $x < y$) e, successivamente, in x tra 0 e ∞ . Scegliamo la prima strada.

$$\begin{aligned} P(X < Y) &= \iint_{\{0 < x < y\}} f_{X,Y}(x, y) dx dy = \int_0^{+\infty} \left[\int_0^y 2e^{-x}e^{-2y} dx \right] dy \\ &= \int_0^{+\infty} 2e^{-2y} \cdot [-e^{-x}]_0^y dy = \int_0^{+\infty} 2e^{-2y}(1 - e^{-y}) dy \\ &= \int_0^{+\infty} 2e^{-2y} dy - \int_0^{+\infty} 2e^{-3y} dy = 1 - \frac{2}{3} = \frac{1}{3}. \end{aligned}$$

Osserviamo, infine, che le due variabili aleatorie X e Y sono tra loro indipendenti, giacché la loro densità congiunta coincide con il prodotto delle due densità marginali. Le leggi di queste due densità, come vedremo nel paragrafo 1.2.9, sono le leggi esponenziali $\mathcal{E}(1)$ e $\mathcal{E}(2)$ di parametro 1 e 2 rispettivamente.

Chiaramente, anche per i vettori aleatori, vale la legge della speranza di una funzione composta (ciò che abbiamo chiamato il **lotus**). Precisamente, se su uno spazio probabilizzato (Ω, \mathcal{A}, P) è dato un vettore aleatorio \mathbf{X} a valori in \mathbb{R}^n dotato di densità $f_{\mathbf{X}}$ e se ϕ è una funzione di \mathbb{R}^n in \mathbb{R} , allora la variabile aleatoria $\phi(\mathbf{X})$ è integrabile se esiste finito l'integrale $\int |\phi(\mathbf{x})| f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x}$ e in tal caso, la speranza di $\phi(\mathbf{X})$ è data dalla formula:

$$\mathbb{E}[\phi(\mathbf{X})] = \int_{\mathbb{R}^n} \phi(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x}. \quad (1.22)$$

concludiamo questo paragrafo utilizzando il **lotus** per dimostrare un'utile proprietà di carattere generale, almeno nel caso delle variabili aleatorie continue. Dimostriamo cioè che, per due variabili aleatorie indipendenti, la speranza del loro prodotto coincide con il prodotto delle loro speranze, cosa questa assolutamente particolare e caratteristica dell'indipendenza. Precisamente:

Teorema 1.2.21 *Se, su uno spazio probabilizzato (Ω, \mathcal{A}, P) , X e Y sono due variabili aleatorie integrabili e indipendenti, il loro prodotto è ancora una variabile aleatoria integrabile; inoltre la sua speranza è data dalla relazione $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$.*

Dimostrazione Come abbiamo detto, faremo la dimostrazione limitandoci al caso di due variabili aleatorie continue. In questo caso, come sappiamo, la legge congiunta del vettore aleatorio (X, Y) coincide con il prodotto delle leggi marginali, cioè $f_{X,Y}(x, y) = f_X(x)f_Y(y)$. D'altra parte, se si considera la funzione $\phi(x, y) = xy$, si ottiene subito, per la (1.22),

$$\begin{aligned} \mathbb{E}[XY] &= \iint_{\mathbb{R}^2} xyf_{X,Y}(x, y) \, dx dy = \iint_{\mathbb{R}^2} xyf_X(x)f_Y(y) \, dx dy \\ &= \int_{-\infty}^{+\infty} \left[\int_{-\infty}^{+\infty} xyf_X(x)f_Y(y) \, dx \right] dy \\ &= \int_{-\infty}^{+\infty} yf_Y(y) \cdot \left[\int_{-\infty}^{+\infty} xf_X(x) \, dx \right] dy \\ &= \left[\int_{-\infty}^{+\infty} xf_X(x) \, dx \right] \cdot \left[\int_{-\infty}^{+\infty} yf_Y(y) \, dy \right] = \mathbb{E}[X]\mathbb{E}[Y]. \end{aligned}$$

Tanto basta per concludere, almeno nel caso delle variabili aleatorie continue. \square

1.2.7 La legge condizionale

Negli esempi che abbiamo fatto fino a questo momento, ci siamo occupati quasi esclusivamente di variabili aleatorie tra loro indipendenti ed abbiamo visto solo incidentalmente che, quando due (o più) variabili aleatorie sono invece dipendenti, il loro studio richiede la conoscenza della legge congiunta. D'altra parte, in moltissime circostanze, può capitare che due variabili aleatorie (o due vettori aleatori) siano legate tra loro e che, addirittura, la legge di una di queste possa essere in qualche modo definita soltanto tramite l'altra.

Esempio 1.2.22 Supponiamo che un esperimento aleatorio consista nel lancio di un dado e, successivamente, registrato il numero che compare sulla faccia uscita nel lancio, si effettui quello stesso numero di lanci di una moneta. Se ci s'interessa al numero di teste che si possono realizzare, si potrebbero costruire, su un opportuno spazio probabilizzato (Ω, \mathcal{A}, P) , due variabili aleatorie discrete X e Y , la prima delle quali rappresenti il risultato del lancio del dado, e la seconda che rappresenti il numero di teste uscite nel lancio delle monete.

Se la legge di X è ovviamente la ripartizione uniforme sull'insieme $E = \{1, 2, \dots, 6\}$,

la legge di Y dovrà essere la legge binomiale $\mathcal{B}(n, 1/2)$ sotto la condizione che l'evento $\{X = n\}$ si sia realizzato, cioè sotto la condizione che il numero uscito nel lancio del dado sia n . In altre parole, la variabile aleatoria Y avrà legge $\mathcal{B}(n, 1/2)$ non secondo la probabilità P ma secondo la probabilità condizionale $P(\cdot | X = n)$. Con un innocente abuso di notazione, in questa circostanza, scriveremo anche, talvolta, che la legge di Y sarà $\mathcal{B}(X, 1/2)$.

Tornando al caso generale e fissato uno spazio probabilizzato (Ω, \mathcal{A}, P) , supponiamo assegnata una variabile aleatoria X a valori nell'insieme E . Lo scopo di tutto il presente paragrafo è quello di studiare l'insieme di tutte le misure di probabilità della forma

$$Q(x, C) = P(C | X = x) \quad \text{per ogni } C \in \mathcal{A} \text{ e ogni } x \in E.$$

Come di consueto, analizziamo dapprima il caso di una variabile aleatoria discreta e, successivamente, analizzeremo il caso di una variabile aleatoria continua. Ora, mentre nel primo caso non vi sono difficoltà a definire la misura di probabilità $Q(x, \cdot)$ in base alla definizione di probabilità condizionale (si veda la (1.10)), nel caso in cui X sia una variabile aleatoria continua ciò comporterà alcune difficoltà perché, in quella circostanza, gli eventi rispetto ai quali si vorrebbe condizionare, cioè gli eventi della forma $\{X = x\}$, sono tutti trascurabili e quindi non si potrà effettuare il condizionamento secondo la definizione usuale. D'altra parte, sebbene questi eventi siano trascurabili, quando la variabile aleatoria X assume un determinato valore x come risultato di un esperimento aleatorio, per quanto a priori questo evento capiti con probabilità nulla, sicuramente l'informazione $X = x$ dovrebbe in generale alterare le probabilità che si vorrà adesso assegnare a tutti gli altri eventi. In effetti, si vedrà che sarà pur sempre possibile definire la probabilità in un senso che si potrebbe definire *improprio*, ma che sarà una naturale generalizzazione di quello già noto.

Definizione 1.2.23 Supponiamo che, su (Ω, \mathcal{A}, P) , sia assegnata una variabile aleatoria discreta X a valori nell'insieme (finito o numerabile) E . Supponiamo inoltre che tutti gli eventi della forma $\{X = n\}$ non siano trascurabili (abbiano cioè probabilità non nulla). Si ponga:

$$Q(n, C) = P(C | X = n) = \frac{P(X = n, C)}{P(X = n)} \quad \text{per ogni } n \in E \text{ e ogni } C \in \mathcal{A}.$$

La famiglia di misure di probabilità $(Q(n, \cdot))_{n \in E}$ si chiama allora una *legge condizionale* della tribù \mathcal{A} rispetto alla variabile aleatoria X .

Il motivo per cui la locuzione “legge condizionale” si riferisce alla tribù è che tutte le misure di probabilità $Q(n, \cdot)$ sono definite su \mathcal{A} . Addirittura, solitamente, si generalizza questa definizione considerando una sottotribù \mathcal{F} di \mathcal{A} e si definisce la legge condizionale di \mathcal{F} rispetto a X come una famiglia di misure di probabilità su \mathcal{F} . Il significato di quest'ultima definizione è il seguente: se si suppone che si sia realizzato l'evento $\{X = n\}$, si riterrà anche di voler stimare la probabilità di ciascuno degli eventi facenti parte della tribù \mathcal{F} non più con la misura P ma con la misura $Q(n, \cdot)$, noncuranti di alterare la misura di probabilità (che fuori di \mathcal{F} rimane pur sempre P) quando invece gli eventi

non appartengano alla tribù \mathcal{F} .

Tornando alla definizione data nel caso generale, osserviamo che questa permette di calcolare facilmente la probabilità di ciascuno degli elementi di \mathcal{A} . Per questo scopo, generalizzando il ragionamento fatto nel paragrafo 1.1.6 che portava alla (1.12), anzi adattandolo alla nuova notazione introdotta, fissiamo un sottoinsieme A di E ed iniziamo con l'osservare che, al variare di n in A , gli eventi della forma $\{X = n\}$ sono a due a due incompatibili e la loro riunione coincide con l'insieme $\{X \in A\}$. In altri termini, ciò significa che vale la relazione $\{X \in A\} = \bigcup_{n \in A} \{X = n\}$. Ora, dalla definizione di legge condizionale, moltiplicando ambo i membri per $P(X = n)$ e sommando su tutti i possibili valori di n , al variare di n in A , si ottiene immediatamente la seguente *formula della disintegrazione*:

$$P(X \in A, C) = \sum_{n \in A} Q(n, C)P(X = n) \quad \text{per ogni } A \subset E \text{ e ogni } C \in \mathcal{A}.$$

Nel caso particolare in cui si ponga $A = E$ nell'espressione appena trovata, si ottiene una formula che permette di calcolare la probabilità di qualsiasi evento, conoscendo la legge condizionale rispetto alla variabile aleatoria X , come la media ponderata delle probabilità condizionali $Q(n, C)$ ciascuna assunta con il peso $p_X(n) = P(X = n)$. In altri termini, sarà

$$P(C) = \sum_{n \in E} Q(n, C)p_X(n) \quad \text{per ogni } C \in \mathcal{A}.$$

Per generalizzare questa definizione nel caso di una variabile aleatoria continua, cioè dotata di densità, osserviamo che, per ogni fissato elemento x di E , la funzione $C \mapsto Q(x, C)$ è una misura di probabilità sulla tribù \mathcal{A} . Inoltre, la formula della disintegrazione caratterizza integralmente la legge condizionale. Questa semplice osservazione permette così di dare la seguente definizione, sempre di carattere generale.

Definizione 1.2.24 Su un fissato spazio probabilizzato (Ω, \mathcal{A}, P) , sia X una variabile aleatoria continua e si denoti con f_X la sua densità. Supponiamo che esista un insieme boreliano D di \mathbb{R} in cui f_X sia strettamente positiva. Sia inoltre \mathcal{F} una sottotribù di \mathcal{A} , ossia un sottoinsieme di \mathcal{A} che sia essa stessa una tribù. Si dice che la sottotribù \mathcal{F} *ammette una legge condizionale* rispetto alla variabile aleatoria X se esiste una famiglia $Q = (Q(x, \cdot))_{x \in D}$ di misure di probabilità su \mathcal{F} che soddisfi la relazione

$$P(X \in A, C) = \int_A Q(x, C)f_X(x) dx \quad \text{per ogni } C \in \mathcal{F} \text{ e ogni } A \in \mathcal{B}(\mathbb{R}).$$

In tal caso, la famiglia Q si dice una *legge condizionale* di \mathcal{F} rispetto a X .

Intuitivamente, nelle ipotesi della definizione precedente, la misura di probabilità $Q(x, \cdot)$ si può interpretare come la misura di probabilità, sulla tribù \mathcal{F} , con cui si vuole misurare il grado di fiducia degli eventi appartenenti alla tribù \mathcal{F} , quando si sappia che la variabile aleatoria ha assunto il valore $X = x$, per quanto questo evento sia estremamente improbabile (tant'è vero che $\{X = x\}$ è un evento trascurabile).

Ora, sempre sullo spazio probabilizzato (Ω, \mathcal{A}, P) , supponiamo che sia data un'altra variabile aleatoria reale Y e supponiamo che la tribù $\mathcal{T}(Y)$ da essa generata (ovvero la tribù contenente tutti gli eventi della forma $\{Y \in B\}$, con B boreliano di \mathbb{R}) ammetta una legge condizionale rispetto a X . Per semplificare la notazione, inoltre, supponiamo che f_X sia ovunque strettamente positiva. In questo caso, per definizione di legge condizionale, si ha

$$P(X \in A, Y \in B) = \int_A Q(x, \{Y \in B\}) f_X(x) dx \quad \text{per ogni } A, B \in \mathcal{B}(\mathbb{R}).$$

Inoltre, la funzione $B \mapsto Q(x, \{Y \in B\})$, definita sulla tribù boreliana di \mathbb{R} , si può pensare come la legge di Y secondo la misura di probabilità $Q(x, \cdot)$, ovvero assumendo che X abbia preso il valore x . Se ammettiamo che, per ogni numero reale x , questa legge sia dotata di una densità, che indicheremo con $f_{Y|X}(y|x)$ e che chiameremo una *densità condizionale* di Y rispetto a X , allora la formula precedente si potrà riscrivere nella forma

$$P(X \in A, Y \in B) = \int_A \left[\int_B f_{Y|X}(y|x) dy \right] dx = \iint_{A \times B} f_{Y|X}(y|x) f_X(x) dx dy.$$

D'altra parte, se si ammette anche che X e Y abbiano densità congiunta $f_{X,Y}$, la precedente probabilità si può anche esprimere nella forma

$$P(X \in A, Y \in B) = \iint_{A \times B} f_{X,Y}(x, y) dx dy,$$

da cui, per l'arbitrarietà di A e B , discende immediatamente l'eguaglianza

$$f_{Y|X}(y|x) f_X(x) = f_{X,Y}(x, y) \quad \text{per ogni } x, y \in \mathbb{R}$$

e dunque, dividendo ambo i membri per $f_X(x)$, che è stata supposta sempre strettamente positiva, si trae infine

$$f_{Y|X}(y|x) = \frac{f_{X,Y}(x, y)}{f_X(x)} \quad \text{per ogni } x, y \in \mathbb{R}. \quad (1.23)$$

È così provato il seguente risultato:

Teorema 1.2.25 *Se, su uno spazio probabilizzato (Ω, \mathcal{A}, P) , due variabili aleatorie X e Y ammettono una densità congiunta $f_{X,Y}$, allora esse ammettono una densità condizionale $f_{Y|X}$ e quest'ultima è data dalla (1.23).*

In particolare, se le due variabili aleatorie X e Y sono indipendenti, la densità condizionale di Y rispetto a X coincide con la densità di Y , cioè $f_{Y|X}(y|x) = f_Y(y)$.

Vediamo subito due esempi molto importanti riguardanti le leggi condizionali. Il primo di questi è un po' più tecnico e riguarda le leggi continue, mentre il secondo è più propriamente probabilistico, e riguarda invece le leggi discrete. Per motivi diversi, tuttavia, entrambi gli esempi sono molto significativi.

Esempio 1.2.26 Su uno spazio probabilizzato (Ω, \mathcal{A}, P) siano X e Y due variabili aleatorie, la cui densità congiunta sia la ripartizione uniforme sulla circonferenza di raggio unitario. Ricordando che l'area del disco unitario di equazione $x^2 + y^2 \leq 1$ è π , supponiamo che la loro densità congiunta sia

$$f_{X,Y}(x, y) = \begin{cases} \frac{1}{\pi} & \text{se } x^2 + y^2 \leq 1, \\ 0 & \text{altrimenti.} \end{cases}$$

Evidentemente, non ci si potrà attendere che le variabili aleatorie X e Y siano tra loro indipendenti, perché se una di queste ha un valore molto vicino a 1, l'altra deve avere un valore molto vicino a 0. Calcoliamo la densità marginale di X . A questo scopo osserviamo che, al variare di x nell'intervallo $[-1, 1]$, dall'equazione del disco segue che, per y , dev'essere $-\sqrt{1-x^2} \leq y \leq \sqrt{1-x^2}$ e di qui, integrando rispetto a y , si trae:

$$f_X(x) = \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} f_{X,Y}(x, y) dy = \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} \frac{1}{\pi} dy = \frac{2}{\pi} \sqrt{1-x^2} \quad \text{per } x \in [-1, 1].$$

La densità condizionale di Y rispetto a X è allora, fissato $x \in [-1, 1]$,

$$f_{Y|X}(y|x) = \frac{f_{X,Y}(x, y)}{f_X(x)} = \frac{\frac{1}{\pi}}{\frac{2}{\pi} \sqrt{1-x^2}} = \frac{1}{2\sqrt{1-x^2}} \quad \text{per } -\sqrt{1-x^2} \leq y \leq \sqrt{1-x^2}.$$

Come subito si riconosce, questa funzione non dipende da y e dunque, per ogni x nell'intervallo $[-1, 1]$, la legge condizionale di Y rispetto a $X = x$ è la ripartizione uniforme sull'intervallo $[-\sqrt{1-x^2}, \sqrt{1-x^2}]$

Esempio 1.2.27 Supponiamo che in un pollaio siano presenti un numero aleatorio di uova e supponiamo che questo numero si possa descrivere come una variabile aleatoria N , su un opportuno spazio probabilizzato (Ω, \mathcal{A}, P) , dotata di legge di Poisson $\mathcal{P}(\lambda)$. Indichiamo poi con X e Y le variabili aleatorie che rappresentano il numero di uova che sono fecondate e il numero di uova che, invece, non sono fecondate. Se si denota con p la probabilità che un singolo uovo sia fecondato, e se al solito si pone $q = 1 - p$, sarà naturale supporre che le variabili aleatorie X e Y abbiano come legge condizionale, rispetto a $N = n$, le leggi binomiali $\mathcal{B}(n, p)$ e $\mathcal{B}(n, q)$ rispettivamente. Inoltre, è naturale supporre che sia $X + Y = N$.

Vogliamo trovare la legge congiunta delle variabili aleatorie X e Y e stabilire se si tratti di due variabili aleatorie indipendenti. A questo scopo, siano i e j due numeri interi e cerchiamo la probabilità $P(X = i, Y = j)$. Per la formula della disintegrazione rispetto a N , si avrà:

$$\begin{aligned} P(X = i, Y = j) &= \sum_{n=0}^{\infty} P(X = i, Y = j | N = n) P(N = n) \\ &= P(X = i, Y = j | N = i + j) P(N = i + j) \\ &= P(X = i | N = i + j) P(N = i + j) \\ &= \frac{(i+j)!}{i! \cdot j!} p^i q^j \frac{\lambda^{i+j}}{(i+j)!} e^{-\lambda} = \frac{(\lambda p)^i}{i!} e^{-\lambda p} \cdot \frac{(\lambda q)^j}{j!} e^{-\lambda q}, \end{aligned}$$

dove la seconda eguaglianza segue dal fatto che, nella somma infinita, l'unico addendo non nullo è quello per cui $n = i + j$, mentre la terza eguaglianza segue dal fatto che, per definizione, sull'evento $\{N = i + j\}$ si ha $Y = (i + j) - X$ e dunque gli eventi $\{X = i\}$ e $\{Y = j\}$ coincidono.

È così provato che le variabili aleatorie X e Y sono indipendenti ed hanno leggi di Poisson rispettivamente $\mathcal{P}(\lambda p)$ e $\mathcal{P}(\lambda q)$.

1.2.8 Varianza e covarianza di una variabile aleatoria

Una variabile aleatoria integrabile X , definita su un opportuno spazio probabilizzato, si dice *centrata* se ha speranza nulla. Data una qualsiasi variabile aleatoria integrabile X , l'unica costante reale μ tale che la differenza $X - \mu$ sia centrata è evidentemente $\mu = \mathbb{E}[X]$; la differenza $X - \mathbb{E}[X]$ si chiama la variabile aleatoria centrata *associata* a X .

Definizione 1.2.28 Su uno spazio probabilizzato (Ω, \mathcal{A}, P) , sia X una variabile aleatoria di quadrato integrabile, ossia tale che X^2 sia una variabile aleatoria integrabile. Si denoti con μ la sua speranza. Si chiama la *varianza* di X il numero $\mathbf{Var}[X]$ così definito:

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

In altri termini, la varianza di X è uguale alla speranza del quadrato della variabile aleatoria centrata $X - \mu$ (cioè il momento del second'ordine di $X - \mu$), ovvero essa è la differenza tra il momento del second'ordine di X e il quadrato del suo momento del prim'ordine. Nella pratica, per trovare operativamente la varianza, si predilige questa seconda formula, riscrivibile anche nella forma

$$\mathbf{Var}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2. \quad (1.24)$$

Intuitivamente, si può dire che la varianza di X è una misura della dispersione di X attorno al proprio valor medio μ . Essa è nulla se e soltanto se la differenza $X - \mu$ è trascurabile, ossia se e soltanto se X è equivalente ad una costante (modulo P). Per ogni numero reale c , si ha poi:

$$\mathbf{Var}[X + c] = \mathbf{Var}[X], \quad \mathbf{Var}[cX] = c^2 \mathbf{Var}[X].$$

Il numero $\sigma[X] = \sqrt{\mathbf{Var}[X]}$ si chiama lo *scarto quadratico medio* (o la *deviazione standard*) di X .

Osservazione 1.2.29 Se X è una variabile aleatoria reale integrabile e dotata di densità, e se f_X è una densità per X , denotiamo con μ la speranza di X . La formula (1.21) permette allora di scrivere:

$$\mathbf{Var}[X] = \int_{\mathbb{R}} (x - \mu)^2 f_X(x) dx.$$

Ovviamente una formula analoga vale anche quando X è una variabile aleatoria discreta: indicando con E l'insieme finito o numerabile dei suoi valori assunti, sarà stavolta

$$\mathbf{Var}[X] = \sum_{n \in E} (x - \mu)^2 P(X = n).$$

Siano adesso X, Y due variabili aleatorie, definite su un opportuno spazio probabilizzato, integrabili e dotate di varianza finita. Si ponga $\mu_X = \mathbb{E}[X]$ e $\mu_Y = \mathbb{E}[Y]$. Si chiama allora la *covarianza* della coppia X, Y il numero reale $\mathbf{Cov}(X, Y)$ così definito:

$$\mathbf{Cov}(X, Y) = \mathbb{E}[(X - \mu_X)(Y - \mu_Y)] = \mathbb{E}[XY] - \mu_X\mu_Y.$$

A parole: la covarianza è la differenza tra la speranza del prodotto ed il prodotto delle speranze. Se risulta $\mathbf{Cov}(X, Y) = 0$, ossia $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$, si dice che le due variabili aleatorie X, Y sono tra loro *non correlate*. Per questo è sufficiente che X, Y siano tra loro indipendenti.

Teorema 1.2.30 *Su uno spazio probabilizzato, siano X, Y due variabili aleatorie integrabili e dotate di varianza finita. Si ha allora*

$$\mathbf{Var}[X + Y] = \mathbf{Var}[X] + \mathbf{Var}[Y] + 2\mathbf{Cov}(X, Y), \quad (1.25)$$

e quindi, affinché risulti

$$\mathbf{Var}[X + Y] = \mathbf{Var}[X] + \mathbf{Var}[Y],$$

occorre e basta che X, Y siano tra loro non correlate.

Dimostrazione Sia il primo membro, sia il secondo membro della relazione (1.25) non mutano se si sostituiscono X, Y con le variabili aleatorie centrate rispettivamente associate. Senza ledere la generalità, si potrà dunque supporre che X, Y siano entrambe centrate. Si ha allora:

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

Tanto basta per concludere. \square

Esempio 1.2.31 Calcoliamo la varianza delle variabili aleatorie dotate delle leggi discrete illustrate nel paragrafo 1.2.2.

Legge uniforme: sia X una variabile aleatoria dotata di legge $\mathcal{U}(N)$. Ricordando l'esempio 1.2.11, si ha $\mathbb{E}(X) = \frac{N+1}{2}$, mentre

$$\mathbb{E}(X^2) = \frac{1}{N} \sum_{k=1}^N k^2 = \frac{(N+1)(2N+1)}{6},$$

come si verifica agevolmente per induzione. Dunque

$$\mathbf{Var}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \frac{(N+1)(2N+1)}{6} - \frac{(N+1)^2}{4} = \frac{N^2-1}{12}.$$

Legge di Bernoulli: sia X una variabile aleatoria dotata di legge $\mathcal{B}(p)$. Ricordando l'esempio 1.2.11, si ha evidentemente $\mathbb{E}[X^2] = \mathbb{E}[X] = p$, e quindi:

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

Legge binomiale: sia ora X una variabile aleatoria con legge $\mathcal{B}(n, p)$. Si può allora supporre che X sia somma di n variabili aleatorie indipendenti (dunque a due a due non correlate), tutte dotate di legge di Bernoulli di parametro p . Ne segue $\mathbb{E}[X] = np$, come già sappiamo dall'esempio 1.2.11, e $\mathbf{Var}[X] = np(1 - p)$.

Legge geometrica: Se T è una variabile aleatoria con legge $\mathcal{G}(\lambda)$, in virtù dell'esempio 1.2.11 si ha $\mathbb{E}[T] = \lambda$; inoltre

$$\mathbb{E}[T^2] = \sum_{k=1}^{\infty} k P(T^2 = k).$$

D'altra parte T assume valori interi, quindi la somma si riduce ai valori $k = m^2$, con $m \in \mathbb{N}^+$. Pertanto

$$\begin{aligned} \mathbb{E}[T^2] &= \sum_{m=1}^{\infty} m^2 P(T^2 = m^2) = \sum_{m=1}^{\infty} m^2 P(T = m) = \sum_{m=1}^{\infty} m^2 p(1 - p)^{m-1} = \\ &= \sum_{m=1}^{\infty} (m^2 - m + m)p(1 - p)^{m-1} = \\ &= \sum_{m=1}^{\infty} m(m - 1)p(1 - p)^{m-1} + \mathbb{E}[T] = \\ &= p(1 - p) \sum_{m=1}^{\infty} m(m - 1)(1 - p)^{m-2} + \frac{1}{p} = \\ &= p(1 - p) \frac{d^2}{dp^2} \frac{1}{p} + \frac{1}{p} = \frac{2(1 - p)}{p^2} + \frac{1}{p} = \frac{2}{p^2} - \frac{1}{p}. \end{aligned}$$

Dunque

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

Legge di Poisson: sia infine X una variabile aleatoria con legge $\mathcal{P}(\lambda)$. Si ha allora, come risulta dall'esempio 1.2.11,

$$\mathbb{E}[X] = \lambda;$$

inoltre

$$\begin{aligned} \mathbb{E}[T^2] &= \sum_{m=1}^{\infty} m^2 P(T^2 = m^2) = \sum_{m=1}^{\infty} m^2 P(T = m) = \sum_{m=1}^{\infty} m^2 \frac{\lambda^m}{m!} e^{-\lambda} = \\ &= \sum_{m=1}^{\infty} m \frac{\lambda^m}{(m - 1)!} e^{-\lambda} = \sum_{m=1}^{\infty} (m - 1 + 1) \frac{\lambda^m}{(m - 1)!} e^{-\lambda} = \\ &= \sum_{m=2}^{\infty} \frac{\lambda^m}{(m - 2)!} e^{-\lambda} + \sum_{m=1}^{\infty} \frac{\lambda^m}{(m - 1)!} e^{-\lambda} = \lambda^2 + \lambda, \end{aligned}$$

da cui

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

In altri termini: per una legge di Poisson il parametro reale λ coincide con la media e con la varianza.

Come abbiamo già osservato, se due variabili aleatorie X e Y sono indipendenti, allora esse sono pure non correlate. In generale, invece, quando le variabili aleatorie X e Y non sono indipendenti, la loro covarianza è un importante indicatore della relazione che sussiste tra loro. Per rendercene conto, supponiamo assegnato uno spazio probabilizzato (Ω, \mathcal{A}, P) , e siano A e B due eventi. Indichiamo con X ed Y le indicatori di A e B rispettivamente, cioè le due variabili aleatorie bernoulliane X e Y così descritte:

$$X = \begin{cases} 1 & \text{se l'evento } A \text{ si realizza,} \\ 0 & \text{altrimenti,} \end{cases} \quad Y = \begin{cases} 1 & \text{se l'evento } B \text{ si realizza,} \\ 0 & \text{altrimenti,} \end{cases}$$

Si osservi innanzitutto che anche la variabile aleatoria XY è una funzione indicatrice e precisamente quella descritta dalle parole

$$XY = \begin{cases} 1 & \text{se } X = Y = 1, \\ 0 & \text{altrimenti.} \end{cases} .$$

Si ottiene così

$$\mathbf{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] = P(X = 1, Y = 1) - P(X = 1)P(Y = 1)$$

da cui si deduce, in particolare, che la condizione $\mathbf{Cov}(X, Y) > 0$ equivale a richiedere che sia $P(X = 1, Y = 1) > P(X = 1)P(Y = 1)$ e di qui, dividendo ambo i membri per $P(Y = 1)$, se ne deduce la disuguaglianza

$$P(X = 1|Y = 1) > P(X = 1).$$

Perciò la covarianza di X e Y è positiva se, sapendo che $\{Y = 1\}$ si è verificato, è più probabile che $\{X = 1\}$ si realizzi. In altre parole, la covarianza di X e Y è positiva se è noto che, allorché l'evento B si è realizzato, è più probabile che anche A si sia realizzato. La forza della relazione tra le due variabili aleatorie X e Y è misurata propriamente dal *coefficiente di correlazione*, che tiene conto anche delle deviazioni standard di X e Y . Precisamente, il coefficiente di correlazione, che si denota con $\rho(X, Y)$ è definito come

$$\rho(X, Y) = \frac{\mathbf{Cov}(X, Y)}{\sqrt{\mathbf{Var}[X]\mathbf{Var}[Y]}}.$$

1.2.9 Leggi continue

In questo paragrafo descriveremo le proprietà delle più importanti leggi sulla retta.

La ripartizione uniforme

Abbiamo già precedentemente illustrato (si veda l'esempio 1.2.15) la ripartizione uniforme su un segmento; riprendiamola brevemente. Per questo scopo, su di un fissato spazio probabilizzato (Ω, \mathcal{A}, P) , una variabile aleatoria reale X si dice *uniformemente ripartita* su un intervallo limitato e chiuso $[a, b]$ se la sua legge è la *ripartizione uniforme*, cioè se è dotata della densità:

$$f_X(x) = \frac{1}{b-a} \cdot I_{[a,b]}(x) = \begin{cases} \frac{1}{b-a} & \text{se } a \leq x \leq b, \\ 0 & \text{altrimenti.} \end{cases}$$

La funzione di ripartizione di una siffatta variabile aleatoria è, come subito si riconosce,

$$F_X(x) = \begin{cases} 0 & \text{per } x < a \\ \frac{x-a}{b-a} & \text{per } a \leq x \leq b \\ 1 & \text{per } x > b. \end{cases}$$

In particolare, se α e β sono due numeri reali, con $a \leq \alpha \leq \beta \leq b$, si avrà

$$P(\alpha \leq X \leq \beta) = \frac{\beta - \alpha}{b - a}.$$

Come peraltro abbiamo già visto nell'esempio 1.2.15, la speranza di X è:

$$\mathbb{E}[X] = \frac{1}{b-a} \int_a^b x \, dx = \frac{a+b}{2}.$$

Determiniamo anche la varianza di X . A questo scopo, sarà sufficiente calcolare il momento del second'ordine $\mathbb{E}[X^2]$ per applicare così la (1.24). Per questo, con un conto analogo a quello fatto per la speranza, usando la (1.21), si trova immediatamente

$$\mathbb{E}[X^2] = \frac{1}{b-a} \int_a^b x^2 \, dx = \frac{1}{3} \cdot \frac{b^3 - a^3}{b-a} = \frac{a^2 + ab + b^2}{3}.$$

La varianza di una variabile aleatoria uniformemente ripartita sull'intervallo $[a, b]$ è dunque

$$\mathbf{Var}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \frac{a^2 + ab + b^2}{3} - \left[\frac{a+b}{2} \right]^2 = \frac{(b-a)^2}{12}.$$

Esempio 1.2.32 Ad una certa fermata passa un autobus ogni 15 minuti a cominciare dalle 7:00 (quindi alle 7:00, alle 7:15, alle 7:30 e così via). Se un passeggero arriva alla fermata in un momento casuale con legge uniforme tra le 7:00 e le 7:30, si calcoli con che probabilità dovrà aspettare il prossimo autobus per meno di 5 minuti.

Per descrivere questo problema, su uno spazio probabilizzato (Ω, \mathcal{A}, P) , si consideri una variabile aleatoria reale X che rappresenti l'istante (espresso in minuti dopo le 7:00) in cui questa persona arriva alla fermata. Ovviamente si potrà supporre che questa abbia legge uniforme sull'intervallo $[0, 30]$. Siccome il passeggero deve aspettare meno di 5 minuti solo se egli arriva tra le 7:10 e le 7:15, oppure tra le 7:25 e le 7:30, la probabilità richiesta sarà data da

$$P(10 \leq X \leq 15) + P(25 \leq X \leq 30) = \frac{5}{30} + \frac{5}{30} = \frac{1}{3}.$$

La legge esponenziale

Abbiamo anche già introdotto la legge esponenziale $\mathcal{E}(\lambda)$ (v. esempio 1.2.12) che qui riprendiamo e studiamo più nel dettaglio. A questo scopo, su uno spazio probabilizzato (Ω, \mathcal{A}, P) , diremo che una variabile aleatoria X ha *legge esponenziale* di parametro λ

(in simboli $\mathcal{E}(\lambda)$) se essa è dotata della densità:

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

Come subito si riconosce, la funzione di ripartizione di una siffatta variabile aleatoria è

$$F_X(x) = \int_0^x \lambda e^{-\lambda t} dt = 1 - e^{-\lambda x} \quad \text{per } x \geq 0.$$

Nelle applicazioni pratiche, la legge esponenziale può rappresentare il tempo di attesa prima che si verifichi un certo evento casuale. Ad esempio, essa può rappresentare il tempo che trascorrerà (a partire da questo momento) fino al verificarsi di un terremoto, o allo scoppiare di un conflitto nucleare, oppure al giungere di una telefonata da parte di qualcuno che si ama.

Calcoliamo la speranza e la varianza di X . Nel primo caso, integrando per parti, si ha

$$\mathbb{E}[X] = \int_0^{+\infty} \lambda x e^{-\lambda x} dx = [-x e^{-\lambda x}]_0^{+\infty} + \int_0^{+\infty} e^{-\lambda x} dx = \frac{1}{\lambda}.$$

Per calcolare la varianza, applichiamo anche in questo caso la (1.24). Calcoliamo dunque il momento del second'ordine della variabile aleatoria X :

$$\mathbb{E}[X^2] = \int_0^{+\infty} \lambda x^2 e^{-\lambda x} dx = [-x e^{-\lambda x}]_0^{+\infty} + \int_0^{+\infty} 2x e^{-\lambda x} dx = \frac{2}{\lambda} \mathbb{E}[X] = \frac{2}{\lambda^2}.$$

Si ottiene così, finalmente,

$$\mathbf{Var}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \frac{1}{\lambda^2}.$$

La legge esponenziale gode di una proprietà equivalente a quella della legge geometrica nel caso discreto: *l'assenza di memoria*. Con questa locuzione s'intende che, se su uno spazio probabilizzato (Ω, \mathcal{A}, P) è data una variabile aleatoria positiva X , dotata di legge esponenziale $\mathcal{E}(\lambda)$, allora vale la proprietà:

$$P(X > s + t \mid X > t) = P(X > s) \quad \text{per ogni } s, t \geq 0. \quad (1.26)$$

Infatti, la condizione espressa dalla precedente uguaglianza si può riscrivere, per la definizione di probabilità condizionale, nella forma

$$\frac{P(X > s + t, X > t)}{P(X > t)} = P(X > s),$$

ovvero, moltiplicando ambo i membri per $P(X > t)$,

$$P(X > s + t) = P(X > s)P(X > t);$$

di qui, osservando che $P(X > x) = e^{-\lambda x}$, la tesi segue immediatamente osservando che questa uguaglianza è banalmente verificata, riducendosi all'identità $e^{-\lambda(s+t)} = e^{-\lambda s} e^{-\lambda t}$.

In realtà, si potrebbe addirittura dimostrare che questa proprietà caratterizza completamente la legge esponenziale, nel senso che qualsiasi variabile aleatoria positiva che goda dell'assenza di memoria ha necessariamente legge esponenziale per un opportuno parametro λ .

Per cercare di capire perché l'equazione (1.26) è detta *assenza di memoria*, si immagini che, su di un fissato spazio probabilizzato (Ω, \mathcal{A}, P) , la variabile aleatoria X rappresenti il tempo di funzionamento di un certo strumento prima che esso si guasti. Sapendo che un tale oggetto è già in funzione al tempo t e non si è ancora rotto, qual è la probabilità che esso continui a funzionare almeno per un ulteriore intervallo di tempo s ? Chiaramente, la probabilità richiesta è quella espressa dal primo membro della (1.26), ovvero $P(X > s + t \mid X > t)$. Infatti, dire che lo strumento in questione non si è ancora guastato al tempo t equivale a dire che il tempo in cui avverrà la rottura (indicato dalla variabile aleatoria X) è superiore a t . D'altro canto, dire che l'oggetto funzionerà per un ulteriore tempo s a partire dal tempo t significa che il tempo X dovrà essere maggiore di $s + t$. In questo senso, la (1.26) afferma che la legge del tempo di vita rimanente (*tempo residuo*) prima della rottura è la medesima sia nel caso in cui lo strumento stia funzionando già da un tempo t , sia nel caso in cui esso sia nuovo. In altri termini, se la (1.26) è soddisfatta, non vi è alcun bisogno di tenere presente l'età dell'oggetto, perché fintanto che esso funziona, si comporta esattamente come se fosse nuovo di zecca.

La legge gamma

Introduciamo adesso una nuova legge sulla retta. A questo scopo, ricordiamo che si chiama *funzione Γ di Eulero* la funzione Γ definita nell'insieme $(0, \infty)$ di tutti i numeri reali strettamente positivi a valori in \mathbb{R} da

$$\Gamma(x) = \int_0^{+\infty} t^{x-1} e^{-t} dt.$$

Non è difficile dimostrare che l'integrale che definisce la funzione Γ converge per ogni numero reale x strettamente positivo. Inoltre, essendo la funzione integranda positiva, la funzione Γ di Eulero è una funzione strettamente positiva e, come subito si riconosce, è $\Gamma(1) = 1$. Inoltre, integrando per parti, si se ne deduce che

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

Ragionando per induzione, ne consegue che, per ogni numero intero positivo n , vale l'eguaglianza $\Gamma(n) = (n - 1)!$ giacché, come abbiamo detto, $\Gamma(1) = 0! = 1$. Supposto, invece, per ipotesi induttiva, che sia $\Gamma(n) = (n - 1)!$, si ha

$$\Gamma(n + 1) = n\Gamma(n) = n(n - 1)! = n!.$$

Si può inoltre dimostrare che la funzione Γ di Eulero è *logaritmicamente convessa*, vale a dire $\log \Gamma$ è una funzione convessa. In altre parole, comunque si scelgano i numeri reali strettamente positivi x, y , e un numero $\lambda \in [0, 1]$, vale l'uguaglianza

$$\Gamma(\lambda x + (1 - \lambda)y) \leq \Gamma(x)^\lambda \Gamma(y)^{1-\lambda}.$$

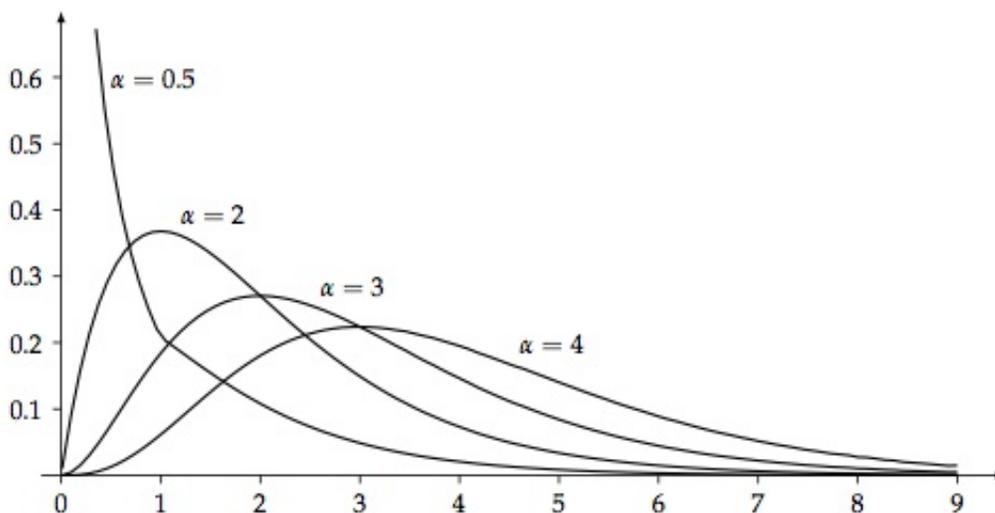
Le tre proprietà appena citate caratterizzano completamente la funzione Γ , nel senso che (*teorema di Bohr e Mollerup*) la funzione Γ è l'unica tra tutte le funzioni f definite sull'intervallo $(0, +\infty)$ che sia dotata delle tre proprietà seguenti:

1. $f(1) = 1$,
2. $f(x+1) = xf(x)$ per ogni $x > 0$,
3. $f(\lambda x + (1-\lambda)y) \leq f(x)^\lambda f(y)^{1-\lambda}$ per ogni $x, y > 0$ e $\lambda \in [0, 1]$.

Tornando al calcolo delle probabilità, fissati due numeri reali strettamente positivi α e λ e fissato uno spazio probabilizzato (Ω, \mathcal{A}, P) , diremo che una variabile aleatoria X ha *legge Ggamma* di parametri α, λ se essa è dotata della seguente densità:

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

La legge gamma si denoterà con il simbolo $\gamma(\alpha, \lambda)$.



. Grafico della densità di probabilità della legge gamma $\gamma(\alpha, 1)$ con $\alpha = 0.5, 2, 3, 4$.

Come subito si riconosce, la funzione appena definita è una densità di probabilità. Calcoliamo la speranza e la varianza di una variabile aleatoria dotata di questa densità. Con una semplicissima sostituzione, si riconosce subito che

$$\mathbb{E}[X] = \int_0^{+\infty} \frac{\lambda^\alpha}{\Gamma(\alpha)} x^\alpha e^{-\lambda x} dx = \frac{1}{\lambda \Gamma(\alpha)} \int_0^{+\infty} t^\alpha e^{-t} dt = \frac{\Gamma(\alpha+1)}{\lambda \Gamma(\alpha)} = \frac{\alpha \Gamma(\alpha)}{\lambda \Gamma(\alpha)} = \frac{\alpha}{\lambda}.$$

In maniera del tutto analoga si calcola anche il momento del second'ordine di X . Si ha infatti:

$$\mathbb{E}[X^2] = \int_0^{+\infty} \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha+1} e^{-\lambda x} dx = \frac{1}{\lambda^2 \Gamma(\alpha)} \int_0^{+\infty} t^{\alpha+1} e^{-t} dt = \frac{\Gamma(\alpha+2)}{\lambda^2 \Gamma(\alpha)} = \frac{\alpha^2 + \alpha}{\lambda^2}.$$

Si ottiene così, finalmente,

$$\mathbf{Var}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \frac{\alpha^2 + \alpha}{\lambda^2} - \left(\frac{\alpha}{\lambda}\right)^2 = \frac{\alpha}{\lambda^2}.$$

Un'importante proprietà delle variabili aleatorie dotate di legge gamma è espressa dal seguente teorema, la cui dimostrazione eccede gli scopi di questo corso.

Teorema 1.2.33 *Su un opportuno spazio probabilizzato (Ω, \mathcal{A}, P) , siano X_1, \dots, X_n variabili aleatorie indipendenti e dotate di leggi $\gamma(\alpha_1, \lambda), \dots, \gamma(\alpha_n, \lambda)$. Allora la variabile aleatoria $S = X_1 + \dots + X_n$ ha legge $\gamma(\alpha_1 + \dots + \alpha_n, \lambda)$. \square*

Da questo risultato, dopo aver osservato che, per $\alpha = 1$, la legge $\gamma(1, \lambda)$ coincide con la legge esponenziale $\mathcal{E}(\lambda)$, ne discende immediatamente che una variabile aleatoria X dotata di legge $\gamma(n, \lambda)$ è la somma di n variabili aleatorie indipendenti ed esponenziali di parametro λ .

Esempio 1.2.34 Supponiamo di avere un dispositivo che funzioni per mezzo di una batteria e supponiamo di avere a disposizione esattamente n batterie. Se la durata di vita di ciascuna di queste batterie si può esprimere con una legge esponenziale $\mathcal{E}(\lambda)$, per un certo parametro λ , allora il tempo totale di funzionamento del dispositivo avrà legge gamma $\gamma(n, \lambda)$.

1.2.10 La legge dei grandi numeri

Un problema che s'incontra molto frequentemente in statistica è il seguente: si osservano delle variabili aleatorie X_1, X_2, \dots, X_n indipendenti e tutte dotate della medesima legge, e se ne vuole stimare la speranza. Un importante risultato teorico che viene incontro a questo problema è la cosiddetta *legge dei grandi numeri*. Per dimostrare questo teorema, occorre dapprima introdurre il seguente risultato preliminare.

Teorema 1.2.35 (Disuguaglianza di Markov) *Consideriamo, su uno spazio probabilizzato (Ω, \mathcal{A}, P) , una variabile aleatoria strettamente positiva X , e sia ϵ un numero reale positivo. Allora*

$$P(X > \epsilon) \leq (1/\epsilon) \cdot \mathbb{E}[X]$$

Dimostrazione È sufficiente osservare che vale l'ovvia disuguaglianza $\epsilon I_{\{X > \epsilon\}} \leq X$. Di qui, infatti, prendendo la speranza di ambo i membri e sfruttando la proprietà di isotonia si ottiene immediatamente la disuguaglianza desiderata. \square

Come corollario di questo risultato, ricaviamo inoltre la seguente altra importantissima disuguaglianza.

Teorema 1.2.36 (Disuguaglianza di Chebyshev) *Nel quadro descritto dal teorema precedente, si supponga che X abbia media μ e varianza σ^2 . Allora, per ogni numero reale ϵ strettamente positivo, si ha*

$$P(|X - \mu| \geq \epsilon) \leq \frac{\sigma^2}{\epsilon^2}. \tag{1.27}$$

Dimostrazione Innanzitutto, osserviamo che il primo membro della disuguaglianza (1.27) coincide con $P(|X - \mu|^2 \geq \epsilon^2)$ e di qui, osservando che la variabile aleatoria $|X - \mu|^2$ è positiva, applicando la disuguaglianza di Markov con ϵ^2 si ottiene immediatamente

$$P(|X - \mu| \geq \epsilon) = P(|X - \mu|^2 \geq \epsilon^2) \leq \frac{1}{\epsilon^2} \mathbb{E}[|X - \mu|^2] = \frac{\sigma^2}{\epsilon^2},$$

perché per definizione $\sigma^2 = \mathbf{Var}[X] = \mathbb{E}[|X - \mu|^2]$. \square

L'importanza delle disuguaglianze di Markov e di Chebyshev sta nel fatto che esse permettono di limitare la probabilità di eventi che riguardano la probabilità di variabili aleatorie di cui si conoscono soltanto la media oppure la media e la varianza. Naturalmente, quando la distribuzione è nota, tali probabilità possono essere calcolate con esattezza, almeno in linea di principio, e non vi è necessità alcuna di ridursi all'utilizzo di una tal maggiorazione.

Esempio 1.2.37 Il numero di pezzi prodotti da una fabbrica durante una settimana di lavoro è mediamente di 50 pezzi. Si può stimare la probabilità che la fabbrica superi occasionalmente i 75 pezzi?

Evidentemente, il numero di pezzi prodotti dalla fabbrica si può descrivere come una variabile aleatoria discreta X , su un opportuno spazio probabilizzato (Ω, \mathcal{A}, P) , la cui legge non è nota, ma è nota la sua speranza $\mathbb{E}[X] = 50$. La probabilità richiesta è dunque $P(X \geq 75)$, che si può stimare attraverso la disuguaglianza di Markov, osservando che

$$P(X \geq 75) \leq \frac{1}{75} \cdot \mathbb{E}[X] = \frac{50}{75} = \frac{2}{3}.$$

Osserviamo che, se nella disuguaglianza di Chebyshev si pone $\epsilon = k\sigma$, essa si può scrivere nella forma equivalente

$$P(|X - \mu| > k\sigma) \leq 1/k^2.$$

In altri termini, la probabilità che una variabile aleatoria differisca dalla sua media per più di k volte la deviazione standard non può mai superare il valore $1/k^2$.

Grazie alla disuguaglianza di Chebyshev si può adesso dimostrare la *legge dei grandi numeri*, che afferma che la media aritmetica di n copie indipendenti di una variabile aleatoria converge verso il valor medio (la speranza) di quest'ultima per n che tende all'infinito. Tale convergenza si precisa dicendo che, scelto un numero reale ϵ strettamente positivo e piccolo a piacere, la media aritmetica si discosta dal valor medio per più di ϵ con probabilità che tende a 0, quando n tende all'infinito. Precisamente:

Definizione 1.2.38 Si consideri, su uno spazio probabilizzato (Ω, \mathcal{A}, P) , una successione $(X_n)_{n \in \mathbb{N}}$ di variabili aleatorie reali. Diremo che essa *converge in probabilità* verso una variabile aleatoria reale X se, comunque si scelga un numero reale ϵ maggiore di zero, si ha $P(|X_n - X| > \epsilon) \rightarrow 0$ al tendere di n all'infinito.

Possiamo ora enunciare la legge dei grandi numeri.

Teorema 1.2.39 (Legge dei grandi numeri) *Su uno spazio probabilizzato (Ω, \mathcal{A}, P) sia $(X_n)_{n \in \mathbb{N}}$ una successione di variabili aleatorie integrabili, aventi varianza finita, indipendenti e dotate della medesima legge. Allora, indicata con μ la comune speranza di queste variabili aleatorie, e posto $S_n = X_1 + \dots + X_n$, la successione $(S_n/n)_{n \in \mathbb{N}}$ converge in probabilità verso la costante μ .*

Dimostrazione Dalle proprietà della speranza seguono immediatamente le due relazioni seguenti:

$$\mathbb{E} \left[\frac{S_n}{n} \right] = \mathbb{E} \left[\frac{X_1 + \dots + X_n}{n} \right] = \mu, \quad \mathbf{Var} \left[\frac{S_n}{n} \right] = \mathbf{Var} \left[\frac{X_1 + \dots + X_n}{n} \right] = \frac{\sigma^2}{n}.$$

Dalla diseguaglianza di Chebyshev segue allora

$$P \left(\left| \frac{S_n}{n} - \mu \right| > \epsilon \right) \leq \frac{\sigma^2}{n\epsilon^2}.$$

Di qui, passando al limite per $n \rightarrow \infty$ e applicando il teorema dei carabinieri, segue immediatamente che $P(|S_n/n - \mu| > \epsilon) \rightarrow 0$, cioè che la successione $(S_n/n)_{n \geq 0}$ converge in probabilità verso la costante μ , come volevamo. \square

Osserviamo fin da subito che, nel caso particolare in cui la successione $(X_n)_{n \in \mathbb{N}}$ sia una successione di variabili aleatorie indipendenti e tutte dotate di legge di Bernoulli di parametro p , si ha $\mu = p$, mentre S_n rappresenta il numero di successi nelle prime n prove (e quindi S_n/n rappresenta la cosiddetta *frequenza* dei successi relativi alle prime n prove). In questo caso, dunque, il risultato precedente si può così leggere: la frequenza dei successi relativi alle prime n prove tende alla probabilità di successo in una singola prova. Cerchiamo di capire meglio il significato di quest'ultima affermazione. Supponiamo a questo scopo che le variabili aleatorie $(X_n)_{n \in \mathbb{N}}$ indichino prove successive e ripetute di un esperimento aleatorio, modellizzato dallo spazio probabilizzato (Ω, \mathcal{A}, P) , in cui si è interessati esclusivamente al realizzarsi di un determinato evento A ; in altri termini, supponiamo che sia

$$X_n = \begin{cases} 1 & \text{se l'evento } A \text{ si realizza all}'n\text{-esima prova,} \\ 0 & \text{se l'evento } A \text{ non si realizza all}'n\text{-esima prova.} \end{cases}$$

Dunque la variabile aleatoria $S_n = X_1 + \dots + X_n$ rappresenta il numero di volte in cui, nel corso delle prime n prove, l'evento A si è verificato. In altri termini, essa fornisce la frequenza con cui l'evento A si è realizzato. Poiché si ha $\mu = \mathbb{E}[X_i] = P(X_i = 1) = P(A)$, la legge dei grandi numeri ci assicura che la frazione delle n prove in cui si realizza l'evento A tende (nel senso della convergenza in probabilità) verso la probabilità $P(A)$.

Esempio 1.2.40 (Il metodo Montecarlo) Sia f una funzione reale limitata, definita sull'intervallo $[0, 1]$ e sia $(X_n)_{n \in \mathbb{N}}$ una successione di variabili aleatorie indipendenti, tutte dotate della stessa legge uniforme su $[0, 1]$. Allora, la successione $(f(X_n))_{n \in \mathbb{N}}$ è ancora formata da variabili aleatorie indipendenti, tutte di speranza eguale a $\mathbb{E}[f(X_1)]$. Per la legge dei grandi numeri, allora,

$$\frac{f(X_1) + \dots + f(X_n)}{n} \tag{1.28}$$

converge in probabilità verso il numero

$$\mathbb{E}[f(X_1)] = \int_0^1 f(x) dx.$$

Questa osservazione suggerisce un metodo probabilistico di calcolo numerico per l'integrale della funzione f sull'intervallo $[0, 1]$. Basterà disporre di un generatore aleatorio di numeri X_1, X_2, \dots con legge uniforme su $[0, 1]$ e quindi calcolare la media (1.28). Quando n è molto grande, questa quantità è un'approssimazione del numero $\int_0^1 f(x) dx$. Questo metodo di approssimazione, noto con il nome di *metodo Montecarlo*, non è particolarmente veloce, ma è molto semplice da implementare e per questo viene spesso utilizzato con profitto.

Esercizi del §1.2

1. Si lanciano tre monete equilibrate. Determinare la legge della variabile aleatoria che rappresenta il numero di teste uscite dopo il lancio delle monete, e calcolarne la speranza e la varianza.
2. Un individuo disponga di quattro urne numerate da 1 a 4 e di cinque palline. Egli metta a caso le palline all'interno delle urne, nel senso che ogni pallina abbia la stessa probabilità di cadere in una qualsiasi delle urne. Determinare la legge della variabile aleatoria che rappresenta il numero di palline all'interno della prima urna, e calcolarne la speranza e la varianza.
3. Un individuo disponga di tre urne numerate da 1 a 3 e di tre palline. Egli metta a caso le palline all'interno delle urne, nel senso che ogni pallina abbia la stessa probabilità di cadere in una qualsiasi delle urne. Determinare la legge della variabile aleatoria che rappresenta il numero di urne vuote, e calcolarne la speranza e la varianza.
4. Due individui lanciano un dado non truccato. Determinare la legge della variabile aleatoria che rappresenta la differenza (in valore assoluto) dei numeri ottenuti dai due giocatori.
5. Un individuo lancia due dadi non truccati. Determinare la legge della variabile aleatoria che rappresenta la somma dei due risultati.
6. Un collezionista ha già raccolto sessanta delle cento figurine di un album. Egli acquista una busta contenente ventiquattro figurine (supposte tutte differenti tra loro), tra le quali naturalmente ve ne possono essere anche alcune che egli già possiede. Calcolare la probabilità che tra le figurine appena acquistate ve ne siano più di venti di quelle che egli già possiede. In media, quante figurine nuove troverà il collezionista nella busta?
7. Da un'urna contenente tre palline rosse e due palline bianche si estrae una pallina. Se la pallina estratta è rossa, allora si lanciano due monete; altrimenti, se la pallina estratta è bianca, si lancia una sola moneta. Scrivere la legge della variabile aleatoria che rappresenta il numero di teste ottenute, e calcolarne la speranza. Calcolare poi la probabilità che sia uscita una pallina rossa, sapendo che si è ottenuta una sola testa.

8. Se si estrae senza reinserimento da una scatola contenente i numeri da 1 a 6 fino a quando non si ottiene il numero 4, in media quante estrazioni si debbono fare?
9. Si lanci una moneta per tre volte e si sommi 1 per ogni volta che esce testa, mentre si sottragga 1 per ogni volta che esce croce. Determinare la densità discreta della variabile aleatoria X che rappresenta il valore ottenuto dal risultato dei tre lanci.
10. Un sacchetto contiene 90 gettoni numerati da 1 a 90. Si estraggono i gettoni, uno dopo l'altro, senza reinserimento, fino a quando esce il numero 3. Se X designa la variabile aleatoria che rappresenta il numero di estrazioni effettuate, trovare la densità discreta di X . Si calcoli inoltre la speranza di X .
11. Si lancino due dadi a sei facce. Determinare la densità discreta della variabile aleatoria X che rappresenta il minimo tra i valori delle facce usciti. Determinare poi la densità discreta della variabile aleatoria Y che rappresenta il massimo.
12. In un test a risposta multipla ogni domanda presenta 5 possibili risposte, delle quali solo una è corretta. Ogni risposta esatta è valutata con 1, ogni risposta in bianco è valutata come 0. Come si dovrebbero valutare le risposte errate, in modo che chi risponde completamente a caso ad ogni domanda mediamente non prenda punti?
13. Si lancino due dadi e siano X e Y le variabili aleatorie che rappresentano i risultati dei due dadi. Sia poi Z la variabile aleatoria che vale 1 se $X = Y$ e 40 altrimenti.
 - Verificare che X, Z sono indipendenti.
 - Dedurre che anche Y, Z sono indipendenti.
 - Verificare che X, Y, Z non sono indipendenti.
14. Una moneta viene lanciata 5 volte. Sia X il numero di teste uscite, e si indichi con W la variabile aleatoria uguale a 1 se $X = 3$ e 0 altrimenti. Determinare la densità discreta congiunta di (X, W) .
15. Una lotteria ha un primo premio di 100 Euro, un secondo di 50 Euro ed un terzo di 25 Euro. Paghereste 1 Euro per un biglietto, se in vendita ci sono 100 biglietti? E se ce ne sono 500?
16. Su uno spazio probabilizzato (Ω, \mathcal{A}, P) , siano X una variabile aleatoria binomiale $\mathcal{B}(2, p)$ e $Y = 2 - X$. Posto $Z = XY$, determinare la densità congiunta di (X, Y) e di (X, Z) .
17. Su uno spazio probabilizzato (Ω, \mathcal{A}, P) , la variabile aleatoria X ha legge binomiale $\mathcal{B}(4, \frac{1}{2})$. Determinare la funzione di ripartizione di X e tracciarne un grafico.
18. Su uno spazio probabilizzato (Ω, \mathcal{A}, P) , sia X una variabile aleatoria con legge di Bernoulli di parametro $1/3$. Calcolare la speranza e la varianza della variabile aleatoria $Y = 2X - 1$.
19. Su uno spazio probabilizzato (Ω, \mathcal{A}, P) , sia X una variabile aleatoria discreta, a valori nell'insieme finito $E = \{0, 1, 2, 3\}$, dotata della densità discreta seguente:

x	0	1	2	3
$f(x)$	0.2	0.3	0.2	0.3

Determinare la speranza della variabile aleatoria $Y = 2X + 1$.

20. Su uno spazio probabilizzato (Ω, \mathcal{A}, P) , siano X, Y due variabili aleatorie indipendenti e dotate di legge di Bernoulli di parametri $1/2$ e $1/3$ rispettivamente. Determinare la legge delle seguenti variabili aleatorie: $X + Y$, $X - 2Y$, $|X - Y|$.
21. Su uno spazio probabilizzato (Ω, \mathcal{A}, P) , la variabile aleatoria X ha legge esponenziale $\mathcal{E}(2)$. Determinare la funzione di ripartizione di X e tracciarne un grafico.
22. Su uno spazio probabilizzato (Ω, \mathcal{A}, P) , sia X una variabile aleatoria con legge esponenziale. Dimostrare che, per ogni $x > 0$ e ogni $y > 0$, si ha $P(X > x + y | X > x) = P(X > y)$.
23. Su uno spazio probabilizzato (Ω, \mathcal{A}, P) , siano X e Y due variabili aleatorie con densità congiunta su $[0, 1]^2$ data da

$$f_{X,Y}(x, y) = \begin{cases} c & \text{se } y \leq x^2, \\ 0 & \text{altrimenti} \end{cases}$$

- Calcolare c .
 - Determinare le densità marginali di X e Y .
 - Calcolare $P(X \leq Y)$ e $P(Y \leq X)$.
24. Su uno spazio probabilizzato (Ω, \mathcal{A}, P) , la variabile aleatoria X ha densità:

$$f(x) = \begin{cases} cx^2 & \text{se } 0 \leq x \leq 3, \\ 0 & \text{altrimenti.} \end{cases}$$

Determinare il valore di c e calcolare la speranza e la varianza di X .

25. Se X e Y sono due variabili aleatorie, su uno spazio probabilizzato (Ω, \mathcal{A}, P) , che rappresentano i risultati del lancio di due dadi, determinare:
- la densità discreta di $X - 2Y$,
 - la speranza di $X - 2Y$,
 - la varianza di $X - 2Y$.
26. Su uno spazio probabilizzato (Ω, \mathcal{A}, P) sia X una variabile aleatoria uniformemente ripartita su $[-1, 1]$. Trovare la densità della variabile aleatoria $Y = X^2$.
27. È data la funzione seguente:

$$f(x) = \begin{cases} cx & \text{per } 0 \leq x < 3, \\ c(6 - x) & \text{per } 3 \leq x < 6, \\ 0 & \text{altrove.} \end{cases}$$

Determinare il numero reale c che rende f una densità di probabilità. Considerata poi, su uno spazio probabilizzato (Ω, \mathcal{A}, P) , una variabile aleatoria X dotata di densità f , calcolare le probabilità $P(X > 3)$ e $P(1.5 \leq X \leq 4.5)$. Calcolare, infine, la speranza di X .

28. Su uno spazio probabilizzato (Ω, \mathcal{A}, P) , è data una variabile aleatoria X avente legge uniforme sull'intervallo $[0, 1]$. Calcolare la legge della variabile aleatoria $Y = e^{\lambda X}$, dove λ è un qualsiasi numero reale, e calcolarne la speranza.
29. Si spezzi a caso un bastoncino di lunghezza unitaria; indi, a partire dai due segmenti ottenuti, si costruisca un triangolo rettangolo avente questi due segmenti come cateti. Determinare la legge della variabile aleatoria che rappresenta l'area di questo triangolo rettangolo.
30. Su uno spazio probabilizzato (Ω, \mathcal{A}, P) , se le variabili aleatorie X_1, X_2, X_3, X_4 rappresentano i lanci successivi di un dado, calcolare

$$\mathbb{E}[(X_1 + X_2)(X_3 + X_4)].$$

31. Dimostrare che, se X e Y sono due variabili aleatorie indipendenti, su uno spazio probabilizzato (Ω, \mathcal{A}, P) , tali che $\mathbb{E}[X] = \mathbb{E}[Y] = 0$, allora

$$\mathbb{E}[(X - Y)(X + Y)^2] = \mathbb{E}[(X - Y)^3].$$

32. Su uno spazio probabilizzato (Ω, \mathcal{A}, P) , si considerino due variabili aleatorie X e Y indipendenti e con la stessa legge. Si ponga poi $U = X - Y$ e $V = X + Y$. Trovare il coefficiente di correlazione tra U e V .
33. Su uno spazio probabilizzato (Ω, \mathcal{A}, P) , siano X e Y due variabili aleatorie di Poisson indipendenti di parametri 2 e 3 rispettivamente. Calcolare $\mathbb{E}[X(1 - X)Y]$.
34. Su uno spazio probabilizzato (Ω, \mathcal{A}, P) , sia X una variabile aleatoria che assume i valori a e $-a$ con egual probabilità. Trovare i valori del parametro a per cui la varianza di X sia 1. Se X_1, X_2 sono due ulteriori variabili aleatorie indipendenti, entrambe con la stessa legge di X , calcolare $P(|X_1 + X_2| < 1)$, usando il valore di a trovato in precedenza.
35. Su uno spazio probabilizzato (Ω, \mathcal{A}, P) , sia X una variabile aleatoria che assume i valori 1, 0, -1 con egual probabilità. Calcolare la varianza della variabile aleatoria $X + 1$. Se adesso X_1, X_2 sono due ulteriori variabili aleatorie indipendenti con la medesima legge di X , calcolare $P(X_1 + X_2 \leq 0)$.
36. Su uno spazio probabilizzato (Ω, \mathcal{A}, P) supponiamo assegnate due variabili aleatorie X_1, X_2 indipendenti e con legge di Poisson di parametro $\lambda > 0$. Determinare λ in modo tale che sia $\mathbb{E}[(X_1 - X_2)^2] = 1$. Utilizzando il valore di λ trovato, calcolare $P(X_1 + X_2 \geq 2)$ e $\mathbb{E}[g(X_1)]$, dove g è la funzione definita da

$$g(x) = \begin{cases} x & \text{per } x \leq 2, \\ 2 & \text{per } x > 2. \end{cases}$$

37. Su uno spazio probabilizzato (Ω, \mathcal{A}, P) , sia X una variabile aleatoria con legge esponenziale di parametro $\lambda > 0$ e poniamo $Y = e^X$.
- Stabilire per quali valori del parametro λ la speranza di Y è finita.
 - Stabilire per quali valori di λ la variabile aleatoria Y ammette varianza.
 - Calcolare il coefficiente di correlazione tra X e Y .

38. Su uno spazio probabilizzato (Ω, \mathcal{A}, P) , siano date due variabili aleatorie indipendenti X e Y , la prima con legge binomiale $\mathcal{B}(2, \frac{1}{3})$, la seconda con legge esponenziale $\mathcal{E}(2)$. Calcolare:
- $\mathbb{E}[X(Y - X)]$,
 - $P(XY \leq 1)$.
 - il coefficiente di correlazione tra X e XY .
39. Su uno spazio probabilizzato (Ω, \mathcal{A}, P) , siano X e Y due variabili aleatorie indipendenti, con X di legge esponenziale $\mathcal{E}(1)$ e Y a valori in $\{-1, 1\}$, tale che $P(Y = 1) = P(Y = -1) = \frac{1}{2}$. Posto allora $Z = XY$, calcolare:
- la funzione di ripartizione di Z ,
 - la densità di Z ,
 - la varianza di Z ,
 - il coefficiente di correlazione tra X e Z .
40. Su uno spazio probabilizzato (Ω, \mathcal{A}, P) , sia X una variabile aleatoria continua con legge uniforme sull'intervallo $[0, 1]$ e sia $Y = X(1 - X)$.
- Calcolare i momenti del primo, secondo, terzo e quarto ordine di X .
 - Calcolare $P(Y \leq \frac{5}{36})$.
 - Calcolare il coefficiente di correlazione tra X e Y .
41. Su uno spazio probabilizzato (Ω, \mathcal{A}, P) , sia X una variabile aleatoria con legge esponenziale di parametro $\lambda = 3$ e poniamo $Y = \frac{X}{1+X}$. Che valori può assumere Y ? Determinare la funzione di ripartizione e la densità di Y .
42. Su uno spazio probabilizzato (Ω, \mathcal{A}, P) , siano date due variabili aleatorie indipendenti X e Y , entrambe con legge binomiale $\mathcal{B}(2, \frac{1}{2})$, e si ponga $Z = XY$.
- Determinare la densità congiunta di X e Z .
 - Stabilire se Y e Z sono indipendenti.
 - Calcolare il coefficiente di correlazione tra X e Z .

1.3 Il teorema limite centrale

1.3.1 Le leggi normali

Descriviamo adesso una delle leggi più importanti della probabilità. Essa è stata introdotta per la prima volta dal matematico francese Abraham De Moivre, che la utilizzò per approssimare le probabilità associate alle variabili aleatorie binomiali $\mathcal{B}(n, p)$ quando il parametro n è grande. Il suo risultato fu poi esteso successivamente da Pierre-Simon de Laplace e successivamente trovò la sua formulazione più generale ad opera di Jarl W. Lindeberg (Helsinki, 1876–1932) e di Paul Lévy (Parigi, 1886–1971) che è oggi universalmente noto col nome di *teorema limite centrale*: si veda il paragrafo 1.3.4). Questo

appellativo è una traduzione piuttosto grossolana della ben più precisa dizione tedesca *zentraler Grenzwertsatz* (teorema centrale del limite). Prima di poter introdurre questo risultato, tuttavia, occorre introdurre una legge estremamente essenziale per il calcolo delle probabilità.

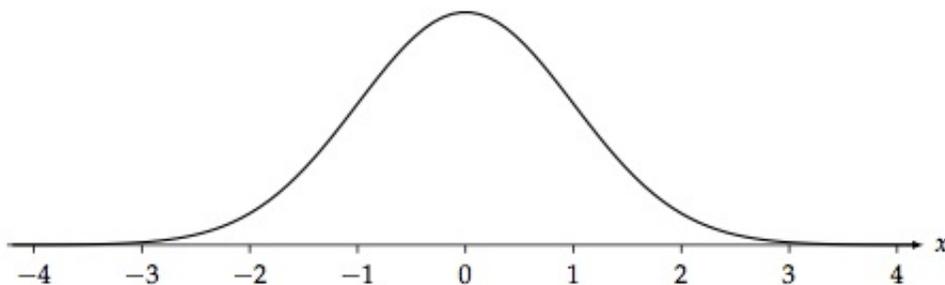
Definizione 1.3.1 Si chiama *legge normale ridotta*, e si denota con il simbolo $\mathcal{N}(0, 1)$, la legge di una variabile aleatoria X , definita su un opportuno spazio probabilizzato, dotata della densità:

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}. \quad (1.29)$$

Non è difficile riconoscere che X è una variabile aleatoria centrata ed ha varianza eguale a 1. Assegnata una siffatta variabile aleatoria X , e fissata una coppia μ, σ di numeri reali, con $\sigma > 0$, la variabile aleatoria $Y = \sigma X + \mu$ ha media μ e varianza σ^2 . La sua densità è data dalla funzione

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

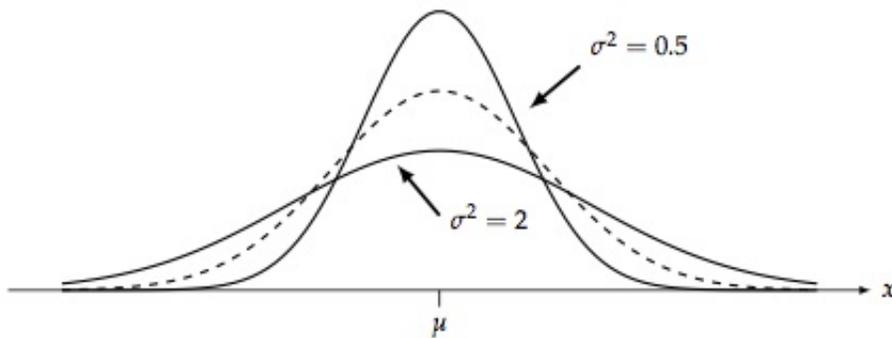
La legge di Y si chiama la *legge normale* di media μ e varianza σ^2 , e si denota con il simbolo $\mathcal{N}(\mu, \sigma^2)$. Inoltre, una variabile aleatoria dotata di legge normale si dice anche una variabile aleatoria *gaussiana*.



Due proprietà molto importanti delle leggi normali, delle quali vedremo alcune applicazioni nei paragrafi successivi, sono le seguenti.

Teorema 1.3.2 Sia Y una variabile aleatoria di legge $\mathcal{N}(\mu, \sigma^2)$ e sia α un numero reale. Allora la variabile aleatoria αY ha legge $\mathcal{N}(\alpha\mu, \alpha^2\sigma^2)$.

Teorema 1.3.3 Siano X, Y due variabili aleatorie indipendenti, la prima dotata di legge $\mathcal{N}(\mu_1, \sigma_1^2)$ e la seconda dotata di legge $\mathcal{N}(\mu_2, \sigma_2^2)$. Allora la variabile aleatoria $Z = X + Y$ è una variabile aleatoria gaussiana e la sua legge coincide con $\mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$.



. Confronto tra le densità normali per due diversi valori della varianza: $\sigma^2 = 2$ e $\sigma^2 = 0.5$. La curva tratteggiata rappresenta la densità della legge $\mathcal{N}(0, 1)$.

Notiamo che il teorema precedente non può valere senza l'ipotesi che X e Y siano tra loro indipendenti. Infatti, se X è una variabile aleatoria gaussiana di legge normale ridotta, e se si prende $Y = X$, allora la variabile aleatoria $Z = 2X$ ha legge $\mathcal{N}(0, 4)$ (per il teorema 1.3.2) e non $\mathcal{N}(0, 2)$ (come dovrebbe essere se valesse il teorema 1.3.3 senza l'ipotesi d'indipendenza).

Un'altra conseguenza fondamentale del teorema 1.3.3 è la seguente. Sia X una variabile aleatoria (su un opportuno spazio probabilizzato (Ω, \mathcal{A}, P)) di legge $\mathcal{N}(\mu, \sigma^2)$. Allora la variabile aleatoria

$$Z = \frac{X - \mu}{\sigma}$$

è una variabile aleatoria gaussiana di legge $\mathcal{N}(0, 1)$. Con una terminologia non troppo bella, ma entrata oramai nella consuetudine del calcolo delle probabilità, una tale variabile aleatoria è detta *standardizzata* e la funzione che ad X associa la variabile aleatoria $Z = (X - \mu)/\sigma$ si dice *standardizzazione* di X .

Osservazione 1.3.4 (La misura di una grandezza fisica) Come si è anticipato all'inizio del presente paragrafo, l'importanza delle leggi normali in statistica è dovuta a questo importantissimo risultato del calcolo delle probabilità che è il cosiddetto *teorema limite centrale* e che, in modo grossolano, si può così enunciare: una variabile aleatoria che si possa esprimere come la somma di un gran numero di variabili aleatorie indipendenti, tutte "abbastanza poco disperse", è approssimativamente normale.

Ad esempio, quando si effettua una misura di una grandezza fisica, si può supporre che il risultato dell'operazione sia una variabile aleatoria e che il risultato della misura sia la somma di un "valore vero" (ammesso che ciò abbia senso) più un termine casuale (l'errore di misura) che è dovuto alla risultante di molti effetti che perturbano gli strumenti utilizzati e le operazioni di lettura, ciascuno dei quali fornisce un piccolo contributo all'errore finale. In assenza di un errore sistematico, si può pensare dunque che il risultato della misura sia una variabile aleatoria della forma $\mu + Y$, dove μ è il vero valore da misurare, mentre Y denota l'errore assoluto che, secondo il teorema limite centrale, avrà approssimativamente legge normale di media 0 (poiché supponiamo che non vi sia un errore sistematico). Dunque, sarà del tutto naturale rappresentare il risultato della misura come una variabile aleatoria $X = \mu + Y$ dotata di legge normale $\mathcal{N}(\mu, \sigma^2)$, dove i parametri μ e σ^2 saranno l'oggetto della ricerca nell'ambito della teoria della misura

e dell'errore: fornire i metodi per questo sarà compito della statistica, ma il teorema limite centrale (cioè il fatto che un effetto casuale, che sia la risultante di molti piccoli effetti, segua necessariamente una legge approssimativamente normale) viene spesso utilizzato per giustificare a priori che le variabili aleatorie che rappresentano la misura di una grandezza fisica seguono una legge normale.

1.3.2 La legge del chi-quadro e la legge di Student

Una legge che può essere definita a partire dalla legge normale e che risulta particolarmente importante per la statistica è la cosiddetta “legge del chi-quadro”. Allo scopo d'introdurla, sia n un numero intero e siano Z_1, \dots, Z_n una n -upla di variabili aleatorie definite su uno spazio probabilizzato (Ω, \mathcal{A}, P) , indipendenti e tutte dotate di legge normale ridotta $\mathcal{N}(0, 1)$. La legge della variabile aleatoria $Y = Z_1^2 + \dots + Z_n^2$ si chiama la *legge del chi-quadro* ad n gradi di libertà, e si denota con il simbolo $\chi^2(n)$. Per tutti gli usi che faremo in seguito delle leggi del chi-quadro non sarà necessario conoscere l'espressione esplicita della densità. Sarà comodo, però, conoscerne la speranza e la varianza. Poiché, per ciascun indice j , risulta $\mathbb{E}[Z_j^2] = \mathbf{Var}[Z_j] = 1$, si ricava subito

$$\mathbb{E}[Y] = \mathbb{E}[Z_1^2] + \dots + \mathbb{E}[Z_n^2] = n.$$

In altri termini, per una variabile aleatoria con legge $\chi^2(n)$ la speranza coincide con il numero di gradi di libertà.

Per la varianza si ha, essendo Z_i^2 e Z_j^2 indipendenti per $i \neq j$,

$$\begin{aligned} \mathbf{Var}[Y] &= \mathbb{E}[(Z_1^2 + \dots + Z_n^2)^2] - n^2 = \sum_{i=1}^n \sum_{j=1}^n \mathbb{E}[Z_i^2 Z_j^2] - n^2 = \\ &= \sum_{h=1}^n \mathbb{E}[Z_h^4] + \sum_{i \neq j} \mathbb{E}[Z_i^2] \mathbb{E}[Z_j^2] - n^2, \end{aligned}$$

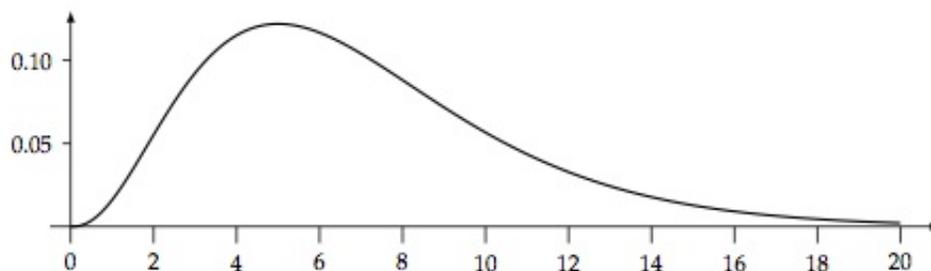
e poiché, integrando due volte per parti,

$$\mathbb{E}[Z_h^4] = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} z^4 e^{-\frac{z^2}{2}} dz = 3,$$

si ottiene subito

$$\mathbf{Var}[Y] = 3n + (n^2 - n) - n^2 = 2n.$$

La figura sottostante descrive l'andamento di una densità $\chi^2(n)$.

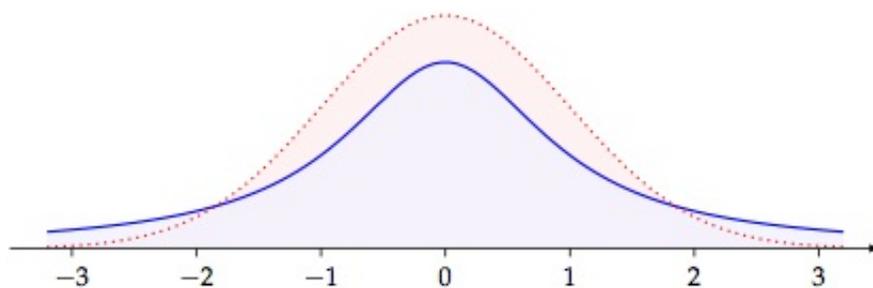


Andamento di una densità $\chi^2(7)$. Questo è l'andamento tipico delle densità $\chi^2(n)$ per $n \geq 2$. Il massimo si trova sempre un po' prima della media, che è eguale al numero di gradi di libertà (7 in questo caso).

Siano ora Z una variabile aleatoria di legge normale ridotta $\mathcal{N}(0, 1)$ e sia Y una variabile aleatoria con legge del chi-quadro $\chi^2(n)$. Si chiama allora la *legge di Student* (dallo pseudonimo di William Sealy Gosset, che si faceva chiamare “Student”) a n gradi di libertà, e si denota con il simbolo $t(n)$, la legge della variabile aleatoria

$$T = \frac{Z\sqrt{n}}{\sqrt{Y}}.$$

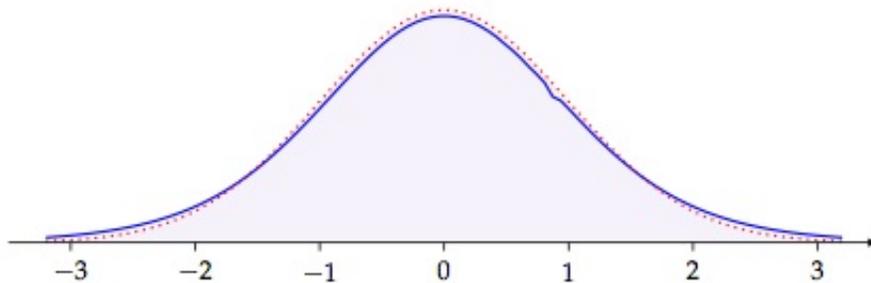
Come per le leggi $\chi^2(n)$ non è molto importante conoscere l’espressione della densità della legge $t(n)$. Nelle figure sottostanti sono rappresentate le densità delle leggi di Student con $n = 1$ e $n = 9$ gradi di libertà, confrontando il loro andamento con quello della normale ridotta che ha una forma analoga.



In realtà si può dimostrare che, al crescere di n , la densità $t(n)$ converge verso la legge normale ridotta $\mathcal{N}(0, 1)$, nel senso che se, su uno spazio probabilizzato (Ω, \mathcal{A}, P) , per ogni intero positivo n , indichiamo con T_n una variabile aleatoria con legge $t(n)$ e con Z una variabile aleatoria normale, allora, per ogni evento A , si ha $P(T_n \in A) \rightarrow P(Z \in A)$ al tendere di n all’infinito. Per capirne il motivo, ragionando però approssimativamente con un argomento che non è né completo né formale, osserviamo che, se Y_n ha legge $\chi^2(n)$ allora si può scrivere nella forma $Y_n = Z_1^2 + \dots + Z_n^2$, dove le variabili aleatorie Z_1, \dots, Z_n sono indipendenti ed hanno legge $\mathcal{N}(0, 1)$. Dunque, per la legge dei grandi numeri, la variabile aleatoria

$$\frac{Y_n}{n} = \frac{Z_1^2 + \dots + Z_n^2}{n}$$

converge verso la costante $\mathbb{E}[Z_1^2] = \dots = \mathbb{E}[Z_n^2] = 1$. Di conseguenza, la variabile aleatoria $T_n = Z/\sqrt{Y_n/n}$ avrà approssimativamente la legge di Z .



Confronto tra la densità $\mathcal{N}(0, 1)$ (tratteggiata) e $t(9)$. Come si vede l’andamento delle due densità è simile, ma la densità di Student tende a 0 per $|x| \rightarrow +\infty$ più lentamente della $\mathcal{N}(0, 1)$. Si può dimostrare che la differenza tra le densità $t(n)$ e $\mathcal{N}(0, 1)$ tende a 0 quando $n \rightarrow \infty$.

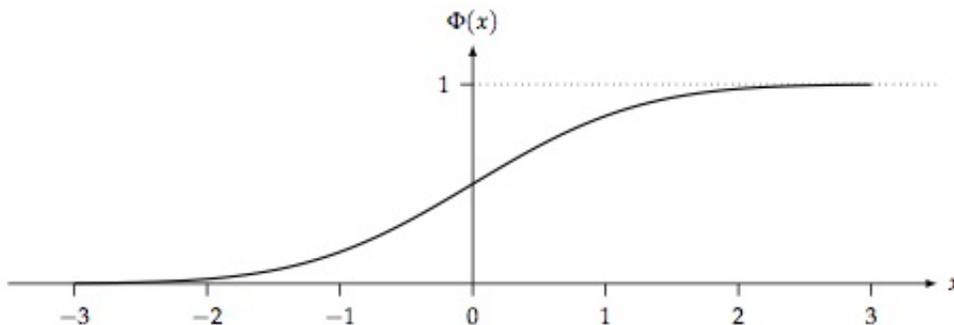
1.3.3 Uso della funzione di ripartizione, i quantili

Utilizzando le leggi normali e quelle da esse derivate (la legge del chi-quadro e la legge di Student in particolare), diventa assai importante l'uso delle funzioni di ripartizione. Vista la complessità (dal punto di vista della teoria delle funzioni) delle densità di queste leggi e l'impossibilità di condurre calcoli esatti con le rispettive funzioni di ripartizione, prima tra tutte la legge normale, la cui funzione di ripartizione non si può scrivere esplicitamente per mezzo delle funzioni elementari, si rende necessario l'uso di tabelle nelle quali siano riportati i loro valori numerici approssimati. L'uso di queste tabelle è assai semplice, ma il loro utilizzo richiede un minimo di cautela per adattare le moltissime situazioni che si possono trovare nella pratica ai ben più modesti dati tabulati che queste forniscono. Nel resto di questo paragrafo, dunque, saranno approfondite alcune semplici nozioni riguardanti la funzione di ripartizione della legge normale ridotta, per vedere come le sue proprietà formali possano essere utilizzate per risolvere alcuni problemi piuttosto caratteristici per il calcolo di talune probabilità.

A questo scopo, supponiamo assegnato uno spazio probabilizzato (Ω, \mathcal{A}, P) e, su di esso, una variabile aleatoria gaussiana Z dotata di legge normale ridotta e denotiamo con Φ la sua funzione di ripartizione. In altri termini, sia Φ la funzione, definita, per ogni numero reale z , da:

$$\Phi(z) = P(Z \leq z) = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx. \quad (1.30)$$

Essa prende anche il nome di *funzione d'errore* e il suo grafico, che si può agilmente tracciare studiando la funzione integrale, è rappresentato qui sotto.



Come è stato detto all'inizio del paragrafo, questa funzione può essere utilizzata per calcolare, per mezzo delle apposite tavole, la probabilità di qualsiasi variabile aleatoria con legge normale, riconducendoci a quella ridotta per mezzo della "standardizzazione". Per riconoscerlo, su uno spazio probabilizzato (Ω, \mathcal{A}, P) , sia X una variabile aleatoria gaussiana di legge $\mathcal{N}(\mu, \sigma^2)$ e supponiamo di voler calcolare la probabilità $P(X \leq c)$. È sufficiente, allora, standardizzare la variabile aleatoria X per ottenere

$$Z = \frac{X - \mu}{\sigma} \leq \frac{c - \mu}{\sigma},$$

e di qui, ricordato che la variabile aleatoria $Z = (X - \mu)/\sigma$ ha legge $\mathcal{N}(0, 1)$, si trae immediatamente

$$P(X \leq x) = P\left(\frac{X - \mu}{\sigma} < \frac{c - \mu}{\sigma}\right) = \Phi\left(\frac{c - \mu}{\sigma}\right).$$

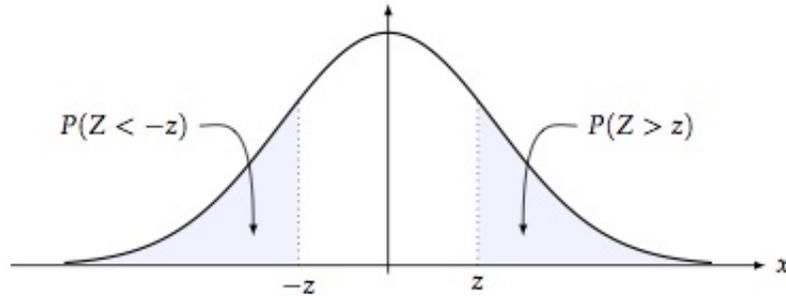
Analogamente, per calcolare la probabilità $P(a \leq X \leq b)$, con $a < b$, si avrà:

$$\begin{aligned} P(a \leq X \leq b) &= P\left(\frac{a - \mu}{\sigma} \leq \frac{X - \mu}{\sigma} \leq \frac{b - \mu}{\sigma}\right) \\ &= P\left(\frac{X - \mu}{\sigma} \leq \frac{b - \mu}{\sigma}\right) - P\left(\frac{X - \mu}{\sigma} \leq \frac{a - \mu}{\sigma}\right) \\ &= \Phi\left(\frac{b - \mu}{\sigma}\right) - \Phi\left(\frac{a - \mu}{\sigma}\right). \end{aligned}$$

In entrambi i casi, ci siamo ricondotti al calcolo di un valore di Φ attraverso la standardizzazione della variabile aleatoria gaussiana X . D'altro canto, come abbiamo detto, l'integrale dell'equazione (1.30) che definisce questa funzione non si può risolvere analiticamente e dunque il calcolo effettivo dei valori di Φ dev'essere fatto approssimativamente utilizzando ad esempio i valori tabulati nell'appendice oppure con l'ausilio del calcolatore.

Le tabelle della funzione di ripartizione della normale ridotta, tuttavia, riportano soltanto i valori di $\Phi(z)$ per i valori positivi di z . Questo perché è possibile ottenere $\Phi(-z)$ sfruttando la simmetria della densità della legge $\mathcal{N}(0, 1)$. Infatti, poiché la densità della legge normale ridotta è simmetrica (la funzione (1.29) è pari), se, su un assegnato spazio probabilizzato (Ω, \mathcal{A}, P) , Z denota una variabile aleatoria reale dotata di legge $\mathcal{N}(0, 1)$, anche $-Z$ ha la medesima legge, nel senso che, lo ricordiamo, per ogni evento A si ha $P(Z \in A) = P(-Z \in A)$. Sia dunque z un qualsiasi numero reale positivo. Si ha allora (si veda la figura sottostante):

$$\Phi(-z) = P(Z < -z) = P(Z > z) = 1 - P(Z \leq z) = 1 - \Phi(z). \quad (1.31)$$



Esempio 1.3.5 Su uno spazio probabilizzato (Ω, \mathcal{A}, P) , sia X una variabile aleatoria gaussiana di media 3 e varianza $\sigma^2 = 16$. Si voglia calcolare la probabilità $P(X < 11)$. A questo scopo, poniamo come al solito $Z = (X - \mu)/\sigma$ e osserviamo che (utilizzando le tavole):

$$P(X < 11) = P\left(\frac{X - 3}{4} < \frac{11 - 3}{4}\right) = P(Z < 2) = \Phi(2) \approx 0.9972.$$

Si voglia calcolare adesso la probabilità $P(X > -1)$. In modo del tutto analogo, si otterrà:

$$P(X > -1) = P\left(\frac{X - 3}{4} > \frac{-1 - 3}{4}\right) = P(Z > -1) = P(Z < 1) = \Phi(1) \approx 0.8413.$$

Infine, se si volesse calcolare invece la probabilità $P(2 < X < 7)$, si avrebbe:

$$\begin{aligned} P(2 < X < 7) &= P\left(\frac{2-3}{4} < \frac{X-3}{4} < \frac{7-3}{4}\right) \\ &= P(-1/4 < Z < 1) \\ &= \Phi(1) - \Phi(-0.25) \\ &= \Phi(1) - [1 - \Phi(0.25)] \approx 0.4400, \end{aligned}$$

dove nella penultima uguaglianza è stata sfruttata la (1.31).

Introduciamo adesso un'ulteriore nozione che semplificherà molte formule legate alla statistica che vedremo in seguito. Diamo dapprima la definizione nel caso più generale, ma c'interesseremo fin da subito soltanto del caso delle tre leggi più importanti ai fini della loro applicazione alla statistica: le leggi normali, la legge del chi-quadro e la legge di Student.

Definizione 1.3.6 Su un assegnato spazio probabilizzato (Ω, \mathcal{A}, P) , sia X una variabile aleatoria reale. Per ogni numero reale α , con $0 < \alpha < 1$, chiameremo *quantile di ordine α* relativo alla variabile aleatoria X il numero

$$x_\alpha = \inf\{x \in \mathbb{R} : P(X \leq x) \geq \alpha\}.$$

Dalla definizione segue abbastanza facilmente che

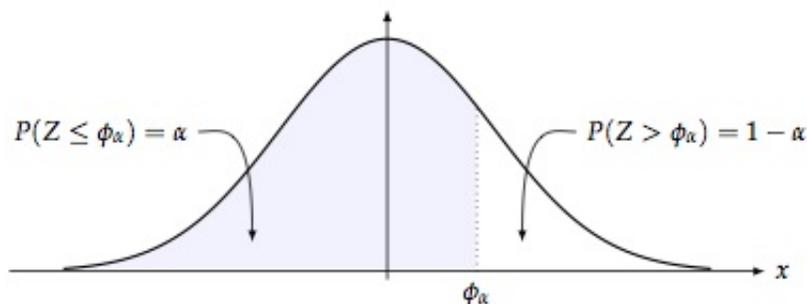
$$P(X \leq x_\alpha) \geq \alpha, \quad P(X < x_\alpha) \leq \alpha.$$

Tuttavia, se X è una variabile aleatoria continua, essa è diffusa; in questo caso, essendo $P(X = x_\alpha) = 0$, si ottiene

$$P(X \leq x_\alpha) = \alpha.$$

Nel caso speciale della legge normale ridotta, il quantile di ordine α sarà indicato con ϕ_α . Così, indicata come al solito Z una variabile aleatoria gaussiana di legge $\mathcal{N}(0, 1)$ su un opportuno spazio probabilizzato (Ω, \mathcal{A}, P) , si avrà (vedi figura sottostante):

$$P(Z \leq \phi_\alpha) = \Phi(\phi_\alpha) = \alpha.$$



In modo analogo a quello della legge normale, con i simboli $t_\alpha(n)$ e $\chi_\alpha^2(n)$ si indicano rispettivamente i quantili di ordine α delle variabili aleatorie di legge $t(n)$ e $\chi^2(n)$. In altri termini: i numeri $t_\alpha(n)$ e $\chi_\alpha^2(n)$ sono definiti dalle relazioni

$$\begin{aligned} P(Y \leq \chi_\alpha^2(n)) &= \alpha, \\ P(T \leq t_\alpha(n)) &= \alpha, \end{aligned}$$

dove Y e T sono due variabili aleatorie di legge $\chi^2(n)$ e $t(n)$ rispettivamente. Dal fatto che la legge di Student è simmetrica rispetto all'asse delle ordinate, proprio come nel caso delle leggi normali, se T è una variabile aleatoria con legge $t(n)$, anche $-T$ avrà legge $t(n)$, cosicché

$$P(T \leq -t_\alpha(n)) = P(-T > t_\alpha(n)) = P(T > t_\alpha(n)) = 1 - P(T \leq t_\alpha(n)) = 1 - \alpha,$$

e di qui segue immediatamente la relazione

$$P(|T| \geq t_{1-\alpha/2}(n)) = \alpha :$$

infatti, a questo scopo sarà sufficiente osservare che, affinché risulti $|T| \geq t_{1-\alpha/2}(n)$, dev'essere $T \leq -t_{1-\alpha/2}(n)$ oppure $T \geq t_{1-\alpha/2}(n)$ e dunque:

$$\begin{aligned} P(|T| \geq t_{1-\alpha/2}(n)) &= P(T \leq -t_{1-\alpha/2}(n)) + P(T \geq t_{1-\alpha/2}(n)) \\ &= 2(1 - (1 - \alpha/2)) = 2(\alpha/2) = \alpha. \end{aligned}$$

1.3.4 Il teorema limite centrale

Abbiamo già accennato nel paragrafo 1.3.1, sia pure senza enunciarlo, al teorema limite centrale e ad alcune sue conseguenze. Andiamo ora ad enunciarlo con esattezza, ed a dedurne un'altra conseguenza che ci permetterà d'introdurre a sua volta alcuni tra i più importanti, ed utili, risultati di approssimazione del calcolo delle probabilità.

Teorema 1.3.7 (Teorema limite centrale) *Consideriamo, su uno spazio probabilizzato (Ω, \mathcal{A}, P) , una successione $(X_n)_{n \in \mathbb{N}}$ di variabili aleatorie indipendenti, tutte dotate della stessa legge, con speranza μ e varianza finita σ^2 . Poniamo*

$$S_n = X_1 + \cdots + X_n, \quad T_n = \frac{S_n - \mathbb{E}[S_n]}{\sqrt{\text{Var}[S_n]}} = \frac{S_n - n\mu}{\sigma\sqrt{n}}. \quad (1.32)$$

Si denoti poi con F_n la funzione di ripartizione di T_n . Allora, per ogni numero reale t , si ha

$$\lim_{n \rightarrow \infty} F_n(t) = \Phi(t).$$

In altri termini, il teorema afferma che per ogni successione $(X_n)_{n \in \mathbb{N}}$ di variabili aleatorie reali, indipendenti e tutte dotate della medesima legge, sulla quale non si fa alcuna ipotesi purché la speranza μ e la varianza σ^2 siano finite, la variabile aleatoria

$$S_n = X_1 + \cdots + X_n$$

ha approssimativamente, per n abbastanza grande, legge normale di media $n\mu$ e varianza $n\sigma^2$.

Un'applicazione tipica del teorema limite centrale è la seguente: supponiamo di voler calcolare la probabilità $P(S_n \leq c)$, dove X_1, \dots, X_n sono variabili aleatorie indipendenti e tutte dotate della stessa legge di speranza μ e varianza σ^2 (entrambe finite), e dove si sia posto $S_n = X_1 + \dots + X_n$. Allora, per n grande, si può approssimare il numero

$$P(S_n \leq c) = P\left(\frac{S_n - n\mu}{\sigma\sqrt{n}} \leq \frac{c - n\mu}{\sigma\sqrt{n}}\right)$$

con il numero

$$\Phi\left(\frac{c - n\mu}{\sigma\sqrt{n}}\right)$$

(calcolabile numericamente mediante le tavole della legge normale). Questa approssimazione è comunemente chiamata *approssimazione normale*.

Esempio 1.3.8 È stato stimato dal produttore che il tempo medio di vita di una determinata marca di lampadine è di 10 giorni, con una deviazione standard di 3 giorni. Qual è la probabilità che 40 lampadine siano sufficienti per un anno intero?

Formuliamo un modello del problema: consideriamo una sequenza X_1, \dots, X_{40} di $n = 40$ variabili aleatorie indipendenti, tutte dotate della medesima legge di speranza $\mu = 10$ e deviazione standard $\sigma = 3$ e tali che, per ogni indice i , la variabile aleatoria X_i rappresenti la durata dell' i -esima lampadina. La variabile aleatoria $S = X_1 + \dots + X_{40}$ rappresenterà dunque la durata del sistema di tutte le 40 lampadine, e dunque si tratterà di calcolare la probabilità $P(S > 365)$. Per il teorema limite centrale, ovvero sfruttando l'approssimazione normale, essa avrà approssimativamente legge normale di media $n\mu = 40 \cdot 10 = 400$ giorni e deviazione standard $\sqrt{n}\sigma = \sqrt{40} \cdot 3 \approx 19$ giorni. Indicata allora con Z una variabile aleatoria reale di legge normale ridotta, si ha:

$$\begin{aligned} P(S > 365) &= P\left(\frac{S - 400}{19} > \frac{365 - 400}{19}\right) \\ &= P(Z > -1.84) = 1 - \Phi(-1.84) = \Phi(1.84) \approx 0.9671. \end{aligned}$$

Una delle più dirette applicazioni dell'approssimazione normale riguarda le variabili aleatorie binomiali. Precisamente, se X_1, \dots, X_n sono variabili aleatorie indipendenti e tutte dotate della stessa legge di Bernoulli $\mathcal{B}(p)$, la variabile aleatoria S_n ha legge binomiale $\mathcal{B}(n, p)$. D'altra parte, il calcolo delle probabilità della forma $P(S_n \leq c)$ è abbastanza complicato: è preferibile dunque utilizzare l'approssimazione normale. Poiché risulta $\mathbb{E}[S_n] = np$ e $\mathbf{Var}[S_n] = npq$ (dove si sia posto $q = 1 - p$), l'approssimazione normale, in questo caso, diviene:

$$P(S_n \leq c) \approx \Phi\left(\frac{c - np}{\sqrt{npq}}\right). \quad (1.33)$$

Esempio 1.3.9 Una popolazione contiene in proporzioni eguali due tipi d'individui (tipo A e tipo B). Da essa viene estratto un campione di 100 individui. Qual è la probabilità che il campione contenga almeno 65 individui di tipo A ?

Se al solito poniamo, per ciascun indice i compreso tra 1 e 100,

$$X_i = \begin{cases} 1 & \text{se l}'i\text{-esimo individuo nel campione è di tipo } A, \\ 0 & \text{altrimenti,} \end{cases}$$

allora il numero totale d'individui di tipo A nel campione è $S_{100} = X_1 + \dots + X_{100}$ e sappiamo che questa variabile aleatoria segue una legge binomiale $\mathcal{B}(100, 1/2)$. La probabilità richiesta è dunque:

$$P(S_{100} \geq 65) = \sum_{k=65}^{100} \binom{100}{k} \frac{1}{2^{100}}. \quad (1.34)$$

Questo calcolo è evidentemente molto laborioso e difficilmente realizzabile senza l'uso di un calcolatore. Invece, utilizzando l'approssimazione normale (1.33), si ottiene:

$$P(S_{100} \geq 65) = 1 - P(S_{100} < 65) \approx 1 - \Phi\left(\frac{64 - 50}{\sqrt{25}}\right) = 1 - \Phi(2.8) \approx 0.0026.$$

Il calcolo esatto della (1.34) avrebbe dato come risultato $P(S_{100} \geq 65) \approx 0.00176$. Osserviamo però che, poiché S_n assume soltanto valori interi, si ha $P(S_{100} \geq 65) = P(S_{100} \geq 64.5)$ e dunque, ripetendo i calcoli, si ottiene

$$P(S_{100} \geq 64.5) \approx 1 - \Phi\left(\frac{64.5 - 25}{\sqrt{25}}\right) = 1 - \Phi(2.9) \approx 0.00186,$$

che è una migliore approssimazione del valore vero 0.00176.

Con riferimento a quanto abbiamo stabilito nell'esempio precedente, possiamo estendere quanto abbiamo fatto osservando, in linea del tutto generale, che è noto empiricamente che, per variabili aleatorie a valori interi, si ottiene sempre una migliore approssimazione calcolando, per ogni numero intero c , la probabilità $P(S_n \leq c + 1/2)$ piuttosto che $P(S_n \leq c)$. Infatti, ogni evento della forma $\{S_n = k\}$ coincide con l'evento $\{k - 0.5 < S_n < k + 0.5\}$ e dunque la riunione di tutti questi eventi, per $k \leq c$, permette di dedurre che $\{S_n \leq c\}$ coincide con $\{S_n \leq c + 0.5\}$. Analogamente, l'evento $\{S_n > c\}$ coincide con $\{S_n \geq c + 0.5\}$. Questa modifica prende il nome di *correzione di continuità*.

Esempio 1.3.10 Il numero ideale di studenti per il primo anno di un corso in una certa università è di 150. L'università, sapendo dall'esperienza passata che soltanto il 30% degli studenti ammessi segue le lezioni, adotta la politica di accettare le iscrizioni di 450 studenti. Si calcoli la probabilità che più di 150 studenti del primo anno frequentino le lezioni.

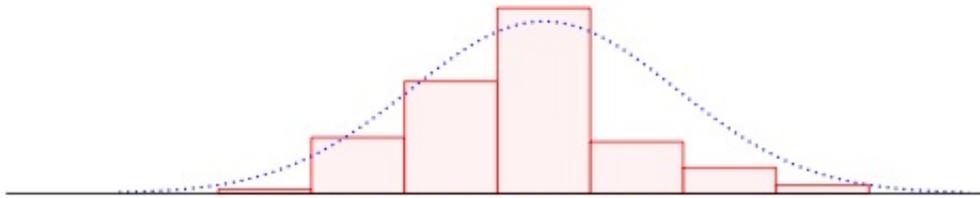
A questo scopo, su un opportuno spazio probabilizzato (Ω, \mathcal{A}, P) , si denoti con X il numero degli studenti che frequentano. Se si assume che ogni studente ammesso decida

o meno di seguire le lezioni indipendentemente da tutti gli altri, allora si potrà ragionevolmente supporre che X abbia legge binomiale di parametri $n = 450$ e $p = 0.3$. Poiché il calcolo con la legge binomiale è troppo complesso, si utilizzerà piuttosto l'approssimazione normale. Indicando a questo scopo con Z una variabile aleatoria, sempre definita sullo spazio probabilizzato (Ω, \mathcal{A}, P) , dotata di legge normale ridotta ed effettuando la *correzione di continuità*, si otterrà:

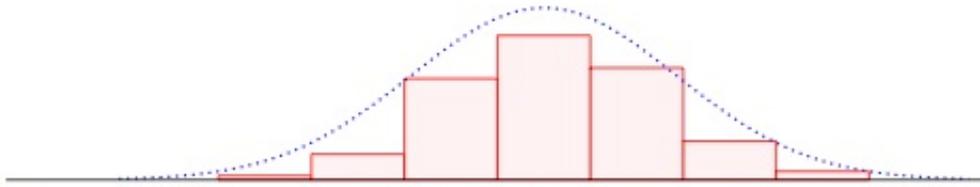
$$\begin{aligned} P(X > 150.5) &= P\left(\frac{X - 450 \cdot 0.3}{\sqrt{450 \cdot 0.3 \cdot 0.7}} > \frac{150.5 - 450 \cdot 0.3}{\sqrt{450 \cdot 0.3 \cdot 0.7}}\right) \\ &= P(Z > 1.59) = 1 - \Phi(1.59) \approx 0.06. \end{aligned}$$

Ancora non abbiamo sollevato la questione di quanto debba essere grande n perché l'approssimazione normale possa applicarsi. Tradizionalmente si considera che la soglia di applicabilità sia $n = 30$ (altri richiedono $n = 50$). In realtà, non vi sono risultati teorici che giustifichino una siffatta scelta, che si basa piuttosto sull'esperienza pratica. Anzi, si può mostrare con degli esempi che, qualunque sia l'intero n , anche molto grande, si possono trovare delle variabili aleatorie X_1, \dots, X_n per le quali la legge di T_n (data da (1.32)) sia lontana dalla legge $\mathcal{N}(0, 1)$. Per esempio, se le variabili aleatorie X_1, \dots, X_n hanno legge di Bernoulli $\mathcal{B}(p)$ si può vedere che, affinché l'approssimazione normale sia soddisfacente, dev'essere $np \geq 5$ e $n(1-p) \geq 5$. Quindi, per valori di p estremi, cioè molto vicini ad 1 oppure a 0, il valore di n necessario può essere molto grande. Osserviamo che questi valori estremi di p corrispondono a delle leggi molto asimmetriche. Per questo, i valori di n indicati precedentemente (30 oppure 50) devono dunque considerarsi come validi per la maggior parte delle leggi che s'incontrano nella pratica, ma vanno aumentati in presenza di leggi molto asimmetriche.

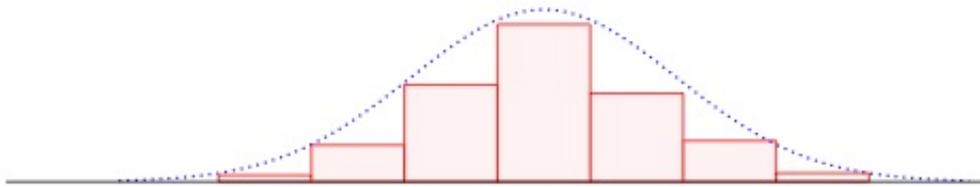
La figura sottostante illustra graficamente come la densità binomiale di parametri n e p si possa approssimare con una legge normale, purché il parametro p non sia troppo estremo, dando origine alle forti asimmetrie cui ci siamo appena riferiti. Si vede, infatti, che soltanto il grafico (c) fornisce un'approssimazione piuttosto buona, mentre il primo, (a), mostra una forte asimmetria nella parte destra, dovuta ad una scelta del parametro $p = 0.05$ che viene in parte compensata, nel caso (b), da una scelta di n più grande.



(a) Istogramma di 200 simulazioni di T_n per delle leggi di Bernoulli con $p = 0.05$ e $n = 50$. Si nota una certa discrepanza tra il grafico e l'istogramma (che è un po' asimmetrico). In questo caso $np = 2.5$, valore troppo basso.



(b) Istogramma di 200 simulazioni di T_n per delle leggi di Bernoulli con $p = 0.05$ e $n = 200$. Ora è $np = 10$ e l'accordo tra il grafico e l'istogramma è buono.



(c) Istogramma di 200 simulazioni di T_n per delle leggi di Bernoulli con $p = 0.5$ e $n = 50$. Ora è $np = n(1-p) = 25$, dunque un valore largamente superiore a 5.

Esercizi del §1.3

1. Su uno spazio probabilizzato (Ω, \mathcal{A}, P) , sia X una variabile aleatoria normale di media 0 e varianza 4. Calcolare le probabilità seguenti:

$$(a) P(-1 \leq X \leq 1), \quad (b) P(X \leq -3), \\ (c) P(|X - 1| \leq 2), \quad (d) P(3 \leq X \leq 6).$$

2. Su uno spazio probabilizzato (Ω, \mathcal{A}, P) , sia X una variabile aleatoria normale di media 8 e varianza 1.3. Calcolare le probabilità seguenti:

$$(a) P(X \leq 9.3), \quad (b) P(X \geq 10), \\ (c) P(6.5 \leq X \leq 7.5).$$

3. Su uno spazio probabilizzato (Ω, \mathcal{A}, P) , sia X una variabile aleatoria gaussiana. Sapendo che $P(X \geq 35) = 0.20$ e $P(X \geq 38) = 0.15$, determinare la media e la varianza di X .

4. Su uno spazio probabilizzato (Ω, \mathcal{A}, P) , sia X una variabile aleatoria gaussiana. Sapendo che $P(X \leq 21) = 0.4$ e $P(X \geq 23) = 0.3$, determinare la media e la varianza di X .

5. Su uno spazio probabilizzato (Ω, \mathcal{A}, P) , sia X una variabile aleatoria gaussiana con media e varianza entrambe eguali a 2. Calcolare la probabilità $P(|X - 2| \leq 2)$.

6. Su uno spazio probabilizzato (Ω, \mathcal{A}, P) , sia X una variabile aleatoria gaussiana con media $\mu > 0$ e varianza $\sigma^2 = f(\mu)$. Determinare la funzione f in maniera tale che la probabilità $P(X \leq 0)$ non dipenda da μ .
7. Su uno spazio probabilizzato (Ω, \mathcal{A}, P) , siano X e Y due variabili aleatorie indipendenti, con X gaussiana di legge $\mathcal{N}(0, 2)$ e Y a valori in $[-1, 1]$ con legge la ripartizione uniforme.
- Calcolare $P(|X| > 1)$.
 - Calcolare la varianza di Y .
 - Calcolare $\mathbb{E} \left[\left(\frac{X}{\sqrt{2}} - Y \right)^2 \right]$.
8. Su uno spazio probabilizzato (Ω, \mathcal{A}, P) , siano assegnate due variabili aleatorie X, Y indipendenti, la prima delle quali abbia legge normale $\mathcal{N}(1, 4)$ e la seconda legge bernoulliana di parametro $p = \frac{1}{4}$. Si definisca poi la variabile aleatoria W come $W = -X$ sull'evento $\{Y = 0\}$ e $W = 2X$ sull'evento $\{Y = 1\}$.
- Determinare la funzione di ripartizione di W .
 - Ricavare la densità di W .
 - Calcolare $P((Y + 1)W \leq 1)$.
9. Su uno spazio probabilizzato (Ω, \mathcal{A}, P) , siano X una variabile aleatoria gaussiana di legge $\mathcal{N}(0, 4)$ e sia $Y = X^2$.
- Valutare $P(Y > 1)$.
 - Determinare la funzione di ripartizione di Y .
 - Determinare la densità di Y .
10. In una certa partita costituita da casse piene di balle di riso, le casse hanno un peso medio di 50 kg, con scarto quadratico medio 4 kg. Le balle vengono scaricate in un magazzino di stoccaggio e vendute una per una. Qual è la probabilità che acquistandone una il suo peso non sia inferiore o eguale a 26 kg?
11. Un distributore di caffè è tarato in maniera tale da fornire 25 cm^3 di caffè con una varianza di 4 cm^6 . Determinare la probabilità che il distributore fornisca una tazza con più di 29 cm^3 di caffè.
12. Lo scorso anno, nel corso di Analisi matematica 2, gli studenti hanno riportato una media di 26 con uno scarto quadratico medio di 2.3. Qual è la probabilità che uno studente di quel corso abbia riportato una valutazione compresa tra 23 e 27?
13. Il peso medio di una confezione di pasta è di 1 kg con uno scarto quadratico medio di 30 g. Qual è la probabilità che un lotto di 40 confezioni pesi più di 39.6 kg?
14. Il 60% di un tipo di automobile ha un difetto al tergicristallo posteriore. Una concessionaria ha trattato la vendita di 50 di queste autovetture. Determinare la probabilità che essa abbia venduto più di 34 automobili con questo difetto.

15. Un medicinale contiene un principio attivo la cui efficacia dipende dalla quantità assunta in diverse somministrazioni. Da un controllo emerge che il contenuto di principio attivo di ogni pasticca preparata è 0.8 mg con uno scarto quadratico medio di 0.2 mg. Considerato che una scatola contiene 40 pasticche e che, inoltre, affinché una scatola venga commercializzata essa deve contenere non meno di 30 mg di principio attivo, determinare la percentuale di confezioni commercializzabili.
16. Un test di matematica è costituito da trenta domande alle quali si può rispondere soltanto “sì” oppure “no”. Per superare l’esame, il candidato deve rispondere correttamente ad almeno 18 domande. Calcolare la probabilità che, rispondendo a caso a tutte le domande, uno studente superi l’esame.
17. Un test di matematica è costituito da cinquanta domande a risposta multipla. Per ciascuna domanda vengono proposte tre risposte di cui soltanto una è quella corretta. Per passare l’esame, il candidato deve rispondere esattamente ad almeno venticinque domande. Uno studente, che non conosca la risposta alle domande, decida di rispondere a caso. Calcolare la probabilità che egli superi l’esame.
18. Un insegnante propone un test con trenta domande a risposta multipla, con cinque risposte per domanda, di cui una sola esatta. L’insegnante ritiene che non debbano superare il test gli studenti che conoscono meno di dieci risposte esatte. Quale limite di sufficienza deve porre perché chi conosce solo nove risposte esatte e risponde a caso alle altre, abbia una probabilità di circa il 10% di superare la prova?
19. Si lancia 120 volte una coppia di dadi. Se per risultato del generico lancio s’intende la somma dei numeri usciti nel corso di quel lancio, calcolare la probabilità che il risultato 7 si presenti almeno 26 volte.
20. Se una popolazione di individui è composta al 30% da individui dotati di una determinata caratteristica, qual è la probabilità che, scegliendo a caso duecento individui, almeno cinquanta di essi possiedano la caratteristica in questione?
21. Tra i 900 studenti di una scuola, si è calcolato che i $\frac{2}{7}$ consumano il pranzo di mezzogiorno a scuola. Quanti pasti devono essere predisposti se si vuole una probabilità del 99% che i pasti preparati siano sufficienti per tutte le richieste?
22. Si generano 250 numeri casuali compresi tra 0 e 1. Determinare la probabilità che la media di questi numeri casuali sia compresa tra 0.4 e 0.6.
23. Dentro le confezioni di un prodotto alimentare ci sono dei piccoli premi per bambini. La percentuale di confezioni con un premio è del 70%. Determinare la probabilità che, comprando 40 scatole del prodotto ci siano più di 15 scatole senza premio.
24. Su uno spazio probabilizzato (Ω, \mathcal{A}, P) , siano X_1, X_2, \dots, X_{200} duecento variabili aleatorie fra loro indipendenti e tutte dotate della medesima legge $\chi^2(2)$ con due gradi di libertà. Si ponga

$$M = X_1 + X_2 + \dots + X_{200}.$$

- Calcolare $\mathbb{E}[M]$.
- Calcolare $\mathbb{E}[M^2]$.
- Stimare la probabilità $P(M \geq 440)$.

25. Su uno spazio probabilizzato (Ω, \mathcal{A}, P) , siano X e Y due variabili aleatorie indipendenti, con leggi rispettivamente $\chi^2(43)$ e di Bernoulli di parametro $\lambda = \frac{1}{2}$.
- Calcolare $\mathbb{E}[XY]$.
 - Calcolare $\mathbb{E}[(X - Y)^2]$.
 - Stimare $P(|X - 3Y| \leq 70.62)$.

Capitolo 2

Elementi di statistica inferenziale

2.1 La stima parametrica

2.1.1 Introduzione

Nelle scienze sperimentali (chimiche, fisiche, biologiche, ecc.) il ricercatore si trova spesso in presenza di dati, provenienti da rilevazioni oppure da misure fisiche, che egli ha il compito di elaborare allo scopo di darne un'interpretazione nell'ambito della teoria che sta studiando. Egli dovrà allora munirsi di opportuni strumenti teorici che gli consentano di costruire un modello matematico dove organizzare questi dati e quindi di trarne le opportune conclusioni, mediante “inferenze” all'interno del modello matematico che ha scelto di adottare. Poiché questo genere di esperimenti presenta intrinsecamente un certo grado d'incertezza dovuto alla natura del problema stesso, è chiaro il ricercatore si troverà di fronte ad un esperimento aleatorio e dunque il modello matematico sarà stabilito naturalmente nell'ambito del calcolo delle probabilità. Di conseguenza, il compito della statistica non sarà tanto quello di fornire il modello matematico, quanto quello di organizzare i dati effettivamente raccolti e soprattutto di costruire delle opportune *regole d'inferenza* che permettano di trarre le deduzioni a partire dai dati raccolti e organizzati. Tradizionalmente la statistica si divide in due parti: la statistica *descrittiva* e la statistica *inferenziale*. La prima di queste si occupa di organizzare e riassumere in modo significativo i dati raccolti, preparandoli ad una successiva analisi inferenziale, e qui termina il suo compito: ne parleremo brevemente nell'ultimo paragrafo. La seconda, invece, utilizzando i metodi e le nozioni del calcolo delle probabilità, si occupa di costruire delle regole d'inferenza da applicare ai dati raccolti.

Il problema generale della statistica è dunque quello in cui si voglia studiare un insieme molto grande, detto *popolazione*, di oggetti a cui sono associate quantità misurabili. L'approccio statistico a questo problema consiste allora nel selezionare un sottoinsieme ridotto di questi oggetti, che viene chiamato un *campione*, e analizzarlo sperando di essere in grado, da questo, di trarre delle conclusioni valide per l'intera popolazione.

Esempio 2.1.1 (Controllo di qualità) Una popolazione è composta da due tipi d'individui: quelli di tipo *A* e quelli di tipo *B* (si veda l'esempio 1.2.8). Supponiamo però di non conoscere il rapporto effettivo tra il numero d'individui di tipo *A* (e dunque di

quelli di tipo B) e il numero totale degli individui della popolazione, e consideriamo il solito esperimento aleatorio consistente nello scegliere dalla popolazione n individui.

Le osservazioni di questo esperimento sono allora delle quantità casuali x_1, \dots, x_n che possono assumere soltanto i valori 0 oppure 1, secondo il codice che abbiamo già precedentemente stabilito nell'esempio 1.2.8. Si potrà dunque pensare che i dati raccolti x_1, \dots, x_n siano i valori assunti da n variabili aleatorie X_1, \dots, X_n definite su un opportuno spazio probabilizzabile (Ω, \mathcal{A}) . Non conoscendo, tuttavia, il rapporto effettivo tra il numero di individui di tipo A e il numero totale degli individui, non siamo in grado di scegliere una misura di probabilità P sullo spazio probabilizzabile (Ω, \mathcal{A}) . D'altra parte, sarà naturale richiedere che queste variabili aleatorie siano tutte indipendenti e bernoulliane, per un opportuno parametro θ . Dunque, si potrà considerare sullo spazio probabilizzabile (Ω, \mathcal{A}) una famiglia di misure di probabilità $\{P^\theta, \theta \in [0, 1]\}$ in maniera tale che, per ogni scelta del parametro sconosciuto θ , le variabili aleatorie X_1, \dots, X_n siano indipendenti ed abbiano legge di Bernoulli $\mathcal{B}(1, \theta)$ secondo la misura di probabilità P^θ .

Esempio 2.1.2 (Misure ripetute di una grandezza fisica) Per effettuare una misura, con un determinato strumento, si esegue un certo numero di misurazioni ottenendo così n risultati x_1, \dots, x_n . Per quello che abbiamo detto riguardo alle variabili aleatorie gaussiane (si veda l'osservazione 1.3.4), questi dati raccolti si potranno pensare come i valori assunti da n variabili aleatorie X_1, \dots, X_n definite su un opportuno spazio probabilizzabile (Ω, \mathcal{A}) . Anche in questo caso non si è in grado di scegliere un'opportuna misura di probabilità su questo spazio. D'altra parte sarà naturale richiedere che queste variabili aleatorie siano indipendenti ed abbiano legge normale, come ci suggerisce il teorema limite centrale, pur non essendo note, a priori, la speranza e la varianza. Si potrà quindi scegliere su (Ω, \mathcal{A}) una famiglia di misure di probabilità $\{P^{\mu, \sigma} : (\mu, \sigma) \in \mathbb{R} \times \mathbb{R}^+\}$ in modo tale che, per ogni scelta dei parametri μ e σ , le variabili aleatorie X_1, \dots, X_n siano indipendenti e gaussiane di legge normale $\mathcal{N}(\mu, \sigma^2)$ secondo la misura di probabilità $P^{\mu, \sigma}$. In questo caso, inoltre, lo scopo stesso della misurazione sarà proprio quello di "stimare" quali siano i comuni valori della speranza e della varianza di queste variabili aleatorie.

Come mostrato negli esempi, in un problema di statistica ci si trova in presenza di un esperimento aleatorio al quale possiamo associare facilmente uno spazio probabilizzabile (Ω, \mathcal{A}) ; tuttavia non siamo in grado, con esattezza, di scegliere a priori su di esso una misura di probabilità P che descriva correttamente come si ritenga di voler distribuire il proprio grado di fiducia tra i possibili eventi legati a questo esperimento aleatorio; tutto quello che siamo in grado di fare in questo spazio è, invece, stabilire che la misura di probabilità P appartiene ad una determinata famiglia \mathcal{P} più o meno grande di misure di probabilità. Al limite, quando non si abbia alcun tipo di informazione addizionale, si potrà addirittura scegliere l'insieme costituito da tutte le misure di probabilità sullo spazio probabilizzabile (Ω, \mathcal{A}) .

Ci si trova così di fronte ad un problema che, per molti aspetti, precede quello affrontato nel calcolo delle probabilità, ovvero quello di voler determinare, sulla base dei dati raccolti qual è la misura di probabilità P , nell'insieme \mathcal{P} , che si ritiene rappresenti più

fedelmente delle altre la distribuzione del nostro grado di fiducia nei confronti dei diversi eventi dell'esperimento aleatorio in esame. Da questo punto di vista, in un certo senso, l'esperimento aleatorio della statistica è fatto proprio allo scopo di raccogliere informazioni su questa misura di probabilità. È così giustificata la seguente definizione.

Definizione 2.1.3 Dati un insieme non vuoto Ω , una tribù \mathcal{A} di parti di Ω e una famiglia \mathcal{P} di misure di probabilità sullo spazio probabilizzabile (Ω, \mathcal{A}) , la terna $(\Omega, \mathcal{A}, \mathcal{P})$ prende il nome di *modello statistico*. Inoltre, se esistono un numero intero d e un sottoinsieme Θ di \mathbb{R}^d tale che la famiglia \mathcal{P} si possa parametrizzare attraverso Θ , nel senso che esiste un'applicazione biunivoca $\theta \mapsto P^\theta$ di Θ in \mathcal{P} , il modello statistico $(\Omega, \mathcal{A}, \mathcal{P})$ si dice *parametrico* e si denoterà preferibilmente con il simbolo

$$(\Omega, \mathcal{A}, (P^\theta)_{\theta \in \Theta})$$

In caso contrario, il modello statistico sarà detto *non parametrico*.

Con il linguaggio appena introdotto, possiamo così riassumere le considerazioni fatte finora:

Compito preliminare per lo studio di un problema statistico sarà associare all'esperimento aleatorio un modello statistico $(\Omega, \mathcal{A}, \mathcal{P})$.

In tutto il seguito, utilizzeremo prevalentemente modelli statistici parametrici; dunque, se non sarà specificato altrimenti, con la locuzione *modello statistico* intenderemo sempre un modello statistico parametrico. Prima di poter descrivere i principali strumenti della statistica, occorre introdurre alcune utili definizioni e notazioni. A questo scopo, diamo immediatamente la seguente definizione.

Definizione 2.1.4 Fissato un modello statistico $(\Omega, \mathcal{A}, (P^\theta)_{\theta \in \Theta})$, una variabile aleatoria X definita sullo spazio probabilizzabile (Ω, \mathcal{A}) è detta una *statistica*.

A prima vista, potrà sembrare che tale denominazione non sia altro che un inutile cambiamento di vocabolario. In realtà, questo nuovo vocabolo ha il compito di tradurre il fatto che, in questo contesto, non è fissata una misura di probabilità sullo spazio probabilizzabile (Ω, \mathcal{A}) , ma un'intera famiglia $(P^\theta)_{\theta \in \Theta}$. Esso dunque esprime, a livello intuitivo, l'idea che, pur essendo ben stabilito lo spazio probabilizzabile in cui la variabile aleatoria è definita, senza ulteriori specificazioni non siamo in grado di dire niente circa la sua legge e quindi, in particolare, non si possono fare calcoli di natura probabilistica, salvo che non sia stato scelto, in qualche modo, un valore per il parametro sconosciuto θ . È chiaro, infatti, che non appena sia stato fissato un valore θ_0 per il parametro sconosciuto, l'esperimento aleatorio sarà descritto dallo spazio probabilizzato $(\Omega, \mathcal{A}, P^{\theta_0})$ e, dunque, il problema statistico si trasformerà in un problema probabilistico.

Definizione 2.1.5 Sul modello statistico $(\Omega, \mathcal{A}, (P^\theta)_{\theta \in \Theta})$, una sequenza di statistiche si dice *indipendente* se è indipendente rispetto a ciascuna delle misure di probabilità P^θ . Analogamente, un'assegnata statistica Y si dice *integrabile* se tale è secondo ciascuna delle misure di probabilità P^θ . Se, dunque, la statistica Y è integrabile, fissato un

elemento θ di Θ , la sua speranza secondo la misura di probabilità P^θ si denoterà con il simbolo

$$\mathbb{E}^\theta[Y]$$

(da leggersi: “la speranza di Y per il valore θ del parametro”) e così, analogamente, indicheremo con $\mathbf{Var}^\theta[Y]$ la sua varianza rispetto alla misura di probabilità P^θ .

Come abbiamo detto all’inizio del paragrafo e mostrato negli esempi, tuttavia, uno sperimentatore che compia un dato esperimento aleatorio opererà selezionando da una popolazione un sottoinsieme di oggetti, le cui caratteristiche saranno sottoposte a misura, in modo tale da ottenerne dei risultati numerici per mezzo dei quali egli potrà trarne le dovute conclusioni e rispondere alle domande che si era posto. Per schematizzare questa procedura, diamo la seguente definizione.

Definizione 2.1.6 Sia $(\mathcal{L}(\theta))_{\theta \in \Theta}$ un’assegnata famiglia di leggi di probabilità. Una sequenza finita X_1, \dots, X_n di statistiche definite su di un assegnato modello statistico $(\Omega, \mathcal{A}, (P^\theta)_{\theta \in \Theta})$ si chiama un *campione statistico* di *taglia* n , estratto da una popolazione, di legge $\mathcal{L}(\theta)$, se le statistiche sono indipendenti e, per ogni indice θ , tutte dotate di legge $\mathcal{L}(\theta)$.

Osserviamo fin da subito che, comunque si scelga una famiglia $(\mathcal{L}(\theta))_{\theta \in \Theta}$ di leggi di probabilità, utilizzando lo schema delle prove indipendenti (teorema 1.2.7) si può sempre costruire un modello statistico e, su di esso, un campione statistico di taglia n , estratto da una popolazione, che abbia legge $\mathcal{L}(\theta)$. Talvolta ci si riferisce a questo modello statistico chiamandolo semplicemente il *modello statistico campionario* di legge $(\mathcal{L}(\theta))_{\theta \in \Theta}$.

Concludiamo questo paragrafo descrivendo anche brevemente quali sono i principali metodi d’inferenza propri della statistica. Come abbiamo ampiamente detto, infatti, quando si studia un esperimento aleatorio dal punto di vista statistico, il vero obiettivo è quello di porsi delle domande sul parametro sconosciuto θ e considerare il fenomeno aleatorio che ne dipende (nel caso dell’esempio 2.1.1, l’estrazione degli n individui dalla popolazione) semplicemente come un esperimento che si compie al solo scopo di trarne qualche indicazione sul parametro θ . In quest’ottica, si possono riconoscere in particolare tre problemi.

1. Il problema della *stima puntuale* del parametro sconosciuto. Esso consiste nella scelta di uno *stimatore*, cioè di un’applicazione T di Ω in Θ . Quest’applicazione rappresenterà la strategia seguente: ci s’impegna, qualunque sarà la realizzazione ω dell’esperimento, ad attribuire convenzionalmente al parametro sconosciuto θ il “valore stimato” $T(\omega)$. Dunque, il problema della stima consisterà nello scegliere lo stimatore T in modo da minimizzare l’errore commesso nell’attribuire al parametro sconosciuto il valore stimato, ossia minimizzando certe quantità, di natura probabilistica, legate appunto allo stimatore ed espresse mediante le probabilità P^θ .
2. Il problema della *stima insiemistica*. Essa consiste nello scegliere un’applicazione \mathcal{S} di Ω in $\mathcal{P}(\Theta)$ che rappresenterà la strategia seguente: ci s’impegna, qualunque sarà la realizzazione ω dell’esperimento, a stimare il vero valore del parametro sconosciuto come appartenente all’insieme $\mathcal{S}(\omega)$ (*insieme di fiducia*).

3. Il problema dei *test d'ipotesi*. Sia data un'*ipotesi* concernente il valore del parametro sconosciuto, cioè l'ipotesi che consiste nell'affermare che il vero valore del parametro appartenga ad una parte fissata Θ_0 di Θ ; il problema consiste nello scegliere in maniera "ragionevole" un *test* che permetta di verificare o di confutare questa ipotesi, cioè costruire una partizione (D, D^c) di Ω che rappresenterà la strategia seguente: ci s'impegna, qualunque sia la realizzazione ω dell'esperimento, a rifiutare l'ipotesi se ω appartiene a D , e ad accettarla se ω appartiene a D^c .

Alla luce di tutto quanto abbiamo detto fino a questo momento, dunque, emerge la differenza fondamentale tra il probabilista puro e lo statistico: quest'ultimo, infatti, non può contentarsi di contemplare le cose e di constatare che la conoscenza del vero valore del parametro θ gli sarà preclusa per l'eternità. Lo statistico è obbligato a passare all'azione, cioè a prendere in ogni caso una decisione. Il suo problema consiste nello studiare, per ciascun valore possibile del parametro, le conseguenze di ciascuna delle sue possibili azioni, e nello scegliere una *regola di decisione* (o *strategia d'azione*), in modo da minimizzare certe conseguenze che sia ragionevole considerare come nocive. Ma occorre sottolineare il fatto che, qualunque sia la strategia che alla fine lo statistico sceglierà, essa dovrà essere una "*regola a priori*", del tipo seguente: ci s'impegna *a priori*, cioè *prima* di compiere l'esperimento e di osservarne il risultato ω , ad agire in un modo che sia univocamente determinato da ω .

2.1.2 Teoria della stima: la nozione di stimatore

Sia $(\Omega, \mathcal{A}, (P^\theta)_{\theta \in \Theta})$ un assegnato modello statistico e, su di esso, supponiamo assegnato un campione X_1, \dots, X_n di taglia n . Come abbiamo detto, il primo dei problemi dello statistico è quello consistente nel ricavare, dalle osservazioni x_1, \dots, x_n del campione, alcune informazioni sul parametro θ o, più in generale, su una funzione $\psi(\theta)$ del parametro, dove ψ è un'assegnata funzione di Θ in \mathbb{R} . A questo scopo, iniziamo col dare la seguente definizione.

Definizione 2.1.7 Nel quadro sopra descritto, supponiamo fissata una funzione reale ψ definita su Θ . Si chiama uno *stimatore* di $\psi(\theta)$ ogni statistica della forma

$$T = t(X_1, \dots, X_n).$$

a valori nell'immagine di $\psi(\theta)$.

Intuitivamente, assegnare uno stimatore $T = t(X_1, \dots, X_n)$ di $\psi(\theta)$ significa fissare la regola secondo la quale, se i dati raccolti dalle osservazioni sono x_1, \dots, x_n ci s'impegna a stimare la quantità sconosciuta $\psi(\theta)$ con il numero $t(x_1, \dots, x_n)$ detto, appunto, la *stima* di $\psi(\theta)$. Inoltre, poiché uno stimatore dipende dalla taglia n del campione estratto, solitamente, si costruisce una successione $(T_n)_{n \geq 1}$ dove, per ciascun indice n , la statistica $T_n = t_n(X_1, \dots, X_n)$ è uno stimatore del parametro sconosciuto $\psi(\theta)$.

Osserviamo, comunque, che il valore assunto dallo stimatore è un'approssimazione del parametro $\psi(\theta)$. In effetti, uno stimatore è una variabile aleatoria (cioè una funzione delle osservazioni) e dunque non assumerà quasi mai (eccetto in casi particolari e

particolarmente semplici) il valore $\psi(\theta)$ da stimare, anche se, naturalmente, si spera comunque che prenda valori non troppo distanti dal parametro che si vuole stimare.

Notiamo subito che, per la definizione che abbiamo appena dato, *qualunque* funzione delle osservazioni è uno stimatore. Occorre quindi disporre anche di qualche criterio con cui stabilire quali di queste funzioni delle osservazioni siano dei “buoni” stimatori e, anche nell’ambito di quelli che si considerano “buoni” stimatori, quali siano i “migliori”. Non entreremo troppo nel dettaglio di questo argomento, perché ci limiteremo qui a studiare soltanto gli stimatori più “naturali” dei semplici problemi che affronteremo (soprattutto legati alla media e alla varianza della legge del campione). Diamo comunque un cenno a questa problematica ed incominciamo con la definizione seguente.

Definizione 2.1.8 Nel quadro sopra descritto, si dice che T è uno stimatore *corretto* (o *non distorto*, in inglese *unbiased*) del parametro $\psi(\theta)$ se risulta

$$\mathbb{E}^\theta[T] = \psi(\theta) \quad \text{per ogni } \theta \in \Theta. \quad (2.1)$$

In caso contrario, esso si dirà uno stimatore *distorto*.

In altri termini, la statistica T può prendere valori diversi da $\psi(\theta)$, ma, se è uno stimatore corretto, il valor medio dei suoi valori dev’essere comunque $\psi(\theta)$, qualunque sia il valore del parametro θ . D’altra parte, se non si ragiona in media, è chiaro che la sostituzione del valore di θ con il suo stimatore T comporta sempre un *costo* (una *perdita*), ed uno stimatore sarà tanto migliore quanto minore sarà questo costo. Dunque si tratta di chiarire cosa s’intende per “costo”. In generale, il costo proveniente dalla sostituzione di $\psi(\theta)$ con il numero reale a è una funzione positiva $(\theta, a) \mapsto C(\theta, a)$. Poiché, se ω rappresenta l’esito dell’esperimento, è stato stabilito di sostituire $\psi(\theta)$ con $T(\omega)$, per ogni θ otteniamo la variabile aleatoria $\omega \mapsto C(\theta, T(\omega))$. Possiamo dunque dare la seguente definizione:

Definizione 2.1.9 Si chiama *rischio* dello stimatore T il suo costo medio, ovvero la funzione reale \mathcal{R}_T , definita su Θ , da

$$\mathcal{R}_T(\theta) = \mathbb{E}^\theta[C(\theta, T)].$$

Generalmente, viene utilizzata come costo la funzione $C(\theta, a) = |\psi(\theta) - a|^2$, che prende il nome di *costo quadratico*. In corrispondenza di questa scelta, il rischio ad essa associato, cioè la funzione $\mathcal{R}_T(\theta) = \mathbb{E}^\theta[|\psi(\theta) - T|^2]$, prende il nome di *rischio quadratico*. Inoltre, se T è uno stimatore corretto di $\psi(\theta)$, tenuto conto della (2.1), si ha $\mathcal{R}_T(\theta) = \mathbf{Var}^\theta[T]$. Infatti,

$$\mathcal{R}_T(\theta) = \mathbb{E}^\theta [|\psi(\theta) - T|^2] = \psi(\theta)^2 - 2\psi(\theta)\mathbb{E}^\theta[T] + \mathbb{E}^\theta[T^2] = \mathbb{E}^\theta[T^2] - \mathbb{E}^\theta[T]^2 = \mathbf{Var}^\theta[T].$$

Infine, introdotte le definizioni di stimatore corretto e di rischio, si potranno scegliere quali sono dei “buoni” stimatori (quelli corretti) e quali siano i “migliori”. A questo scopo, se T è uno stimatore del parametro $\psi(\theta)$, diremo che esso è *preferibile* ad un altro stimatore S se, per ogni θ in Θ , si ha $\mathcal{R}_T(\theta) \leq \mathcal{R}_S(\theta)$. Inoltre, muniti di questa sorta di “ordinamento” tra gli stimatori, se si considera una famiglia \mathcal{D} di stimatori per il parametro $\psi(\theta)$ e se T appartiene a questa famiglia, si potrà stabilire che T è *ottimale* (rispetto a \mathcal{D}) se è preferibile ad ogni altro stimatore della famiglia \mathcal{D} .

2.1.3 La media empirica e la varianza empirica

Fino a questo momento abbiamo visto in estrema generalità la nozione di stimatore. È giunto dunque il momento di fare qualche esempio per vedere quali siano le reali applicazioni di quanto abbiamo stabilito nel paragrafo precedente. A questo scopo, consideriamo una popolazione di elementi, a ciascuno dei quali sia associata una grandezza numerica. Su un opportuno modello statistico $(\Omega, \mathcal{A}, (P^\theta)_{\theta \in \Theta})$, indichiamo dunque con X_1, \dots, X_n un campione di taglia n estratto da questa popolazione. Tanto per fare un esempio, si potrà supporre che il campione sia il risultato di n misure di una medesima grandezza fisica, nel qual caso (come abbiamo stabilito nell'osservazione 1.3.4), si tratterà di un campione statistico estratto da una popolazione di legge $\mathcal{N}(\mu, \sigma^2)$. In ogni caso, supporremo sempre che il campione sia costituito da variabili aleatorie reali che siano integrabili ed abbiano varianza finita. Indichiamo dunque con μ e con σ^2 la loro speranza e la loro varianza, che prenderanno il nome di *media della popolazione* e *varianza della popolazione*. Supporremo, inoltre, che né μ , né σ^2 siano note e dunque che il parametro sconosciuto sia $\theta = (\mu, \sigma^2)$ (e, conseguentemente, sarà $\Theta = \mathbb{R} \times \mathbb{R}_+$). Diamo immediatamente la seguente definizione.

Definizione 2.1.10 Nel quadro sopra descritto, si chiama *media empirica* (o *media campionaria*) lo stimatore di μ così definito:

$$\bar{X} = \frac{X_1 + \dots + X_n}{n}.$$

Osserviamo prima di tutto che si tratta di uno stimatore corretto. Per riconoscerlo, basta osservare che, dalle proprietà della speranza, si ha:

$$\mathbb{E}^{\mu, \sigma} [\bar{X}] = \mathbb{E}^{\mu, \sigma} \left[\frac{X_1 + \dots + X_n}{n} \right] = \frac{\mathbb{E}^{\mu, \sigma} [X_1] + \dots + \mathbb{E}^{\mu, \sigma} [X_n]}{n} = \frac{n\mu}{n} = \mu.$$

Inoltre, per quanto riguarda la varianza, si ha (ricordando che le X_i sono indipendenti)

$$\mathbf{Var}^{\mu, \sigma} [\bar{X}] = \mathbf{Var}^{\mu, \sigma} \left[\frac{X_1 + \dots + X_n}{n} \right] = \frac{\mathbf{Var}^{\mu, \sigma} [X_1] + \dots + \mathbf{Var}^{\mu, \sigma} [X_n]}{n^2} = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}.$$

Dunque, la media empirica è uno stimatore corretto la cui varianza è ridotta, rispetto alla varianza della popolazione, di un fattore n . Dunque il rischio quadratico dello stimatore \bar{X} è infinitesimo per $n \rightarrow +\infty$, e in definitiva \bar{X} ha una variabilità che si riduce sempre di più, all'aumentare di n , rendendo la stima puntuale di μ mediante \bar{X} sempre più precisa all'aumentare della taglia del campione. Inoltre, possiamo approssimare la legge della media empirica: a questo scopo, basta osservare che, con le notazioni appena introdotte, l'enunciato del teorema limite centrale (si veda il paragrafo 1.3.4) si può parafrasare dicendo che la statistica

$$Z = \frac{\bar{X} - \mu}{\sigma} \sqrt{n}$$

ha approssimativamente legge normale ridotta rispetto a ciascuna delle probabilità $P^{\mu, \sigma}$, il che significa che \bar{X} ha approssimativamente legge normale $\mathcal{N}(\mu, \sigma^2/n)$.

Occupiamoci adesso di trovare uno stimatore anche per la varianza della popolazione. A questo scopo, prendiamo ancora un campione X_1, \dots, X_n di taglia n estratto da una popolazione e continuiamo a supporre che ciascuna statistica del campione abbia speranza e varianza finite, che indicheremo ancora con μ e σ^2 . Sappiamo allora che \bar{X} è la sua media empirica.

Definizione 2.1.11 Nel quadro sopra descritto, si chiama *varianza empirica* (oppure *varianza campionaria*), lo stimatore S^2 di σ^2 così definito:

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2.$$

La sua radice quadrata, ovvero la statistica $S = \sqrt{S^2}$, si chiamerà invece la *deviazione standard empirica* (o campionaria).

Osserviamo subito che la varianza empirica è anch'essa uno stimatore corretto della varianza. Per riconoscerlo, ricordata l'eguaglianza $\sum_{i=1}^n X_i = n\bar{X}$, si ha:

$$\sum_{i=1}^n (X_i - \bar{X})^2 = \sum_{i=1}^n X_i^2 - 2\bar{X} \sum_{i=1}^n X_i + \sum_{i=1}^n \bar{X}^2 = \sum_{i=1}^n X_i^2 - n\bar{X}^2.$$

Inoltre, tenuto conto di questa relazione, si potrà agevolmente scrivere la varianza empirica nella sua forma equivalente:

$$S^2 = \frac{1}{n-1} \left(\sum_{i=1}^n X_i^2 - n\bar{X}^2 \right) \quad (2.2)$$

ovvero

$$(n-1)S^2 = \sum_{i=1}^n X_i^2 - n\bar{X}^2.$$

Prendendo adesso la speranza di quest'ultima eguaglianza e ricordando che, per ogni variabile aleatoria reale Y integrabile e dotata di varianza finita, il suo momento del second'ordine si potrà scrivere nella forma $\mathbb{E}^{\mu, \sigma}[Y^2] = \mathbf{Var}^{\mu, \sigma}[Y] + \mathbb{E}^{\mu, \sigma}[Y]^2$, si trae:

$$\begin{aligned} (n-1)\mathbb{E}^{\mu, \sigma}[S^2] &= \mathbb{E}^{\mu, \sigma} \left[\sum_{i=1}^n X_i^2 \right] - \mathbb{E}^{\mu, \sigma} [n\bar{X}^2] \\ &= n\mathbb{E}^{\mu, \sigma} [X_1^2] - n\mathbb{E}^{\mu, \sigma} [\bar{X}^2] \\ &= n\mathbf{Var}^{\mu, \sigma}[X_1] + n\mathbb{E}^{\mu, \sigma}[X_1]^2 - n\mathbf{Var}^{\mu, \sigma}[\bar{X}] - n\mathbb{E}^{\mu, \sigma}[\bar{X}]^2 \\ &= n\sigma^2 + n\mu^2 - n \cdot \frac{\sigma^2}{n} - n\mu^2 = (n-1)\sigma^2 \end{aligned}$$

da cui, finalmente,

$$\mathbb{E}^{\mu, \sigma}[S^2] = \sigma^2.$$

È piuttosto complicato invece calcolare direttamente la varianza di S^2 : tuttavia, come nel caso della media empirica, la quantità $\mathbf{Var}^{\mu, \sigma}[S^2]$ tende a 0 al tendere di n all'infinito, come vedremo fra poco. Dunque anche la stima puntuale di σ^2 mediante lo stimatore S^2 è sempre più precisa all'aumentare della taglia del campione.

Esempio 2.1.12 I dati che seguono sono il risultato di 16 misure indipendenti del punto di fusione del piombo espressi in gradi Celsius:

330.0	328.6	324.4	334.0	337.5	341.0	343.3	329.5
322.0	331.0	340.4	326.5	327.3	340.0	331.0	332.3

Assumendo che questi dati possano essere pensati come un campione normale di taglia $n = 16$, si determini una stima della media e della deviazione standard del campione. Poiché, come sappiamo, la media empirica e la varianza empirica sono due stimatori corretti della media e della varianza della popolazione, si avrà semplicemente:

$$\bar{x} = \frac{x_1 + \cdots + x_{16}}{16} \approx 333.6^\circ\text{C}, \quad s = \sqrt{\frac{(x_1 - \bar{x})^2 + \cdots + (x_{16} - \bar{x})^2}{15}} \approx 6.4^\circ\text{C}.$$

Per quanto abbiamo detto nel paragrafo 2.1.2, i valori \bar{x} e s potranno essere assunti come una stima della media μ e della deviazione standard σ . In altre parole, il punto di fusione è stimato con una variabile aleatoria gaussiana dotata di legge normale $\mathcal{N}(\bar{x}, s^2)$.

Nel caso della legge della media empirica abbiamo visto che il teorema limite centrale permette di trovarla, almeno approssimativamente, in qualsiasi caso. Per studiare adesso la legge della varianza empirica, limitiamoci al caso in cui il campione X_1, \dots, X_n sia estratto da una popolazione che abbia una legge normale $\mathcal{N}(\mu, \sigma^2)$, intendendo così che le variabili aleatorie che compongono il campione siano (ovviamente) indipendenti ed abbiano legge normale di media μ e varianza σ^2 . A questo scopo, sarà utile il seguente risultato che enunceremo con il linguaggio della probabilità soltanto per alleggerire la notazione.

Teorema 2.1.13 (di Cochran) *Su uno spazio probabilizzato (Ω, \mathcal{A}, P) sia X_1, \dots, X_n un campione statistico di taglia n e di legge $\mathcal{N}(\mu, \sigma^2)$. Indichiamo, come al solito, con \bar{X} e S^2 la media e la varianza empirica. Allora \bar{X} e S^2 sono tra loro indipendenti. Inoltre, \bar{X} ha sempre legge normale $\mathcal{N}(\mu, \sigma^2/n)$. Infine, se si pone*

$$W = \frac{(n-1)S^2}{\sigma^2}, \quad T = \frac{\bar{X} - \mu}{S} \sqrt{n},$$

allora la variabile aleatoria W ha legge $\chi^2(n-1)$, mentre T ha legge $t(n-1)$. \square

Il teorema 2.1.13 non soltanto fornisce le leggi di \bar{X} e S^2 per i campioni gaussiani, ma stabilisce anche un'importante proprietà (unica delle leggi normali): che questi due stimatori sono tra loro indipendenti. Inoltre, la conoscenza delle leggi dei due stimatori \bar{X} e S^2 (che, chiaramente, dipendono dai parametri sconosciuti μ e σ) permette di fare delle previsioni probabilistiche sia sul valor medio che sulla varianza.

Ma c'è di più: ricordando che la variabile aleatoria W ha legge $\chi^2(n-1)$ e dunque ha varianza uguale a $2(n-1)$, si ricava subito

$$\text{Var}[S^2] = \frac{\sigma^4}{(n-1)^2} \text{Var}[W] = \frac{2\sigma^2}{n-1}.$$

Dunque, come abbiamo anticipato in precedenza, la stima puntuale della varianza σ^2 di un campione statistico gaussiano mediante lo stimatore S^2 è sempre più precisa all'aumentare della taglia del campione.

Esempio 2.1.14 Il tempo di vita, in ore, di un tipo di lampadine ha media 500 e deviazione standard 80. Preso un campione di taglia $n = 16$ ed assumendo che abbia legge normale, quale sarà la probabilità che la media empirica sia maggiore di 525 ore? Per risolvere il problema, su uno spazio probabilizzato (Ω, \mathcal{A}, P) , prendiamo 16 variabili aleatorie X_1, \dots, X_{16} indipendenti e gaussiane di legge $\mathcal{N}(500, 6400)$. Dunque, ricordato che \bar{X} ha legge normale di media $\mu = 500$ e deviazione standard $\sigma/\sqrt{n} = 80/4 = 20$, se ne trae immediatamente:

$$P(\bar{X} > 525) = P\left(\frac{\bar{X} - 500}{20} > \frac{525 - 500}{20}\right) = 1 - \Phi(1.25) \approx 0.11.$$

Esempio 2.1.15 Il tempo impiegato da un microprocessore ad eseguire alcuni processi è rappresentabile come una variabile aleatoria normale con media 30ns (nanosecondi) e deviazione standard 3ns. Se si osserva l'esecuzione di 16 processi, qual è la probabilità che la varianza empirica risultante sia maggiore di 15ns?

Per rispondere a questa domanda, su uno spazio probabilizzato (Ω, \mathcal{A}, P) , prendiamo 16 variabili aleatorie X_1, \dots, X_{16} indipendenti e gaussiane di legge $\mathcal{N}(30, 9)$. Siccome l'ampiezza del campione è $n = 16$ e $\sigma^2 = 9$, grazie al teorema 2.1.13 si potrà scrivere, utilizzando la tavola della legge $\chi^2(15)$:

$$P(S^2 > 15) = P\left(\frac{(n-1)S^2}{\sigma^2} > 15 \cdot \frac{15}{9}\right) = P(W > 25) \approx 1 - 0.95 = 0.05.$$

2.1.4 Stimatori di massima verosimiglianza

Supponiamo che un individuo disponga di un'urna contenente cento palline, alcune delle quali bianche e le altre nere e supponiamo che l'individuo non conosca il numero di palline bianche nell'urna, ma egli sappia che ve ne sono 99 di un colore e 1 dell'altro. Egli estragga poi dieci palline in sequenza, rimettendo ogni volta la pallina estratta nell'urna, e supponiamo che ciascuna delle palline estratte sia bianca. Egli è allora portato a dedurre che all'interno dell'urna vi siano 99 palline bianche e solo 1 pallina nera. Questo perché, intuitivamente, è più *verosimile* estrarre per dieci volte una pallina bianca da un'urna contenente 99 palline bianche ed una nera, piuttosto che viceversa. Ritorniamo adesso al caso generale di un modello statistico $(\Omega, \mathcal{A}, (P^\theta)_{\theta \in \Theta})$ nel quale sia definito un campione statistico X_1, \dots, X_n di taglia n e supponiamo di voler trovare uno stimatore per tutte le componenti del parametro sconosciuto $\theta = (\theta_1, \dots, \theta_d)$. Cerchiamo di formalizzare il ragionamento fatto nel precedente semplicissimo esempio. A questo scopo, siano x_1, \dots, x_n le quantità osservate. Supponiamo che il campione sia dotato di densità (discreta oppure continua) e denotiamo con F_θ la densità congiunta del campione secondo P^θ . Poiché le variabili aleatorie che compongono il campione sono indipendenti, tale densità congiunta sarà il prodotto delle densità marginali, tutte eguali a una data funzione f_θ , e dunque $F_\theta(x_1, \dots, x_n) = f_\theta(x_1) \cdots f_\theta(x_n)$. Chiameremo *funzione di verosimiglianza* la funzione

$$\mathcal{M}(\theta | x_1, \dots, x_n) = F_\theta(x_1, \dots, x_n) = f_\theta(x_1) \cdots f_\theta(x_n),$$

definita su Θ ed avente x_1, \dots, x_n come parametri.

Per motivare il nome dato a questa funzione, osserviamo che, nel caso in cui le variabili aleatorie X_1, \dots, X_n siano discrete, la funzione di verosimiglianza coincide con la probabilità $P^\theta(X_1 = x_1, \dots, X_n = x_n)$, che è proprio la probabilità che il campione assuma i valori x_1, \dots, x_n . Se interpretiamo allora la funzione \mathcal{M} come la *verosimiglianza* (ovvero la plausibilità, la credibilità) che si realizzino i dati misurati x_1, \dots, x_n quando θ è assunto come parametro, sembra ragionevole adottare come stima di θ quel valore $\hat{\theta}$ che rende massima la verosimiglianza per i dati osservati: in tal caso infatti si ha

$$P^{\hat{\theta}}(X_1 = x_1, \dots, X_n = x_n) \geq P^\theta(X_1 = x_1, \dots, X_n = x_n) \quad \forall \theta \in \Theta;$$

dunque scegliendo $\theta = \hat{\theta}$, ossia scegliendo la misura di probabilità $P^{\hat{\theta}}$, la realizzazione dei dati osservati appare più probabile. In altri termini, la stima di massima verosimiglianza $\hat{\theta}$ è definita come: *il valore di θ che rende massima la verosimiglianza $\mathcal{M}(\theta | x_1, \dots, x_n)$, quando i valori osservati sono x_1, \dots, x_n .*

Definizione 2.1.16 Nel quadro sopra descritto, supponiamo che la funzione \mathcal{M} , per ogni scelta dei parametri x_1, \dots, x_n , ammetta un unico punto di massimo assoluto, che nel seguito denoteremo con il simbolo

$$\hat{\theta}(x_1, \dots, x_n) = (\hat{\theta}_1(x_1, \dots, x_n), \dots, \hat{\theta}_d(x_1, \dots, x_n)).$$

Le statistiche

$$T_1 = \hat{\theta}_1(X_1, \dots, X_n), \quad \dots, \quad T_d = \hat{\theta}_d(X_1, \dots, X_n),$$

sono stimatori delle componenti $\theta_1, \dots, \theta_d$, e il vettore $\hat{\theta}(x_1, \dots, x_n)$ si chiama lo *stimatore di massima verosimiglianza* di $\theta_1, \dots, \theta_d$.

In generale, uno stimatore di massima verosimiglianza può non esistere (questo accade, ad esempio, se in corrispondenza di una certa scelta dei parametri x_1, \dots, x_n la funzione definita su Θ da $\theta \mapsto \mathcal{M}(\theta | x_1, \dots, x_n)$ non possiede un punto di massimo assoluto), oppure esso può non essere unico (se i punti di massimo assoluto sono più di uno). Tuttavia, esistono dei teoremi che assicurano che, se n è abbastanza grande, questi problemi non compaiono, sicché, nella pratica il problema dell'esistenza dello stimatore di massima verosimiglianza si presenta raramente. Inoltre, poiché in genere l'insieme Θ è un sottoinsieme di \mathbb{R}^d e la funzione di verosimiglianza è abbastanza regolare, la ricerca dello stimatore di massima verosimiglianza si può compiere con i soliti metodi dell'analisi (ricerca degli zeri della derivata).

Osservazione 2.1.17 Nella pratica, invece di calcolare i punti di massimo della funzione \mathcal{M} , conviene calcolare i punti di massimo della funzione logaritmica $\mathcal{L} = \ln \mathcal{M}$. Evidentemente, infatti, queste due funzioni hanno gli stessi punti di massimo e minimo, giacché, per le ben note regole di derivazione, si ha

$$\frac{\partial \mathcal{L}}{\partial \theta_j} = \frac{1}{\mathcal{M}} \cdot \frac{\partial \mathcal{M}}{\partial \theta_j}.$$

Esempio 2.1.18 (Stima del parametro della legge di Poisson) Andiamo a calcolare lo stimatore di massima verosimiglianza per un campione X_1, \dots, X_n di legge di Poisson di parametro sconosciuto λ . Per questo, fissiamo i parametri x_1, \dots, x_n e scriviamo la funzione di verosimiglianza:

$$\begin{aligned}\mathcal{M}(\lambda) &= P^\lambda(X_1 = x_1, \dots, X_n = x_n) \\ &= e^{-\lambda} \frac{\lambda^{x_1}}{x_1!} \dots e^{-\lambda} \frac{\lambda^{x_n}}{x_n!} \\ &= e^{-n\lambda} \cdot \prod_{i=1}^n \frac{\lambda^{x_i}}{x_i!}.\end{aligned}$$

Consideriamo adesso la funzione

$$\mathcal{L}(\lambda) = \ln \mathcal{M}(\lambda) = -n\lambda + \ln \lambda \sum_{i=1}^n x_i - \sum_{i=1}^n \ln(x_i!)$$

e deriviamo quest'ultima rispetto al parametro λ :

$$\frac{\partial \mathcal{L}}{\partial \lambda} = -n + \frac{1}{\lambda} \sum_{i=1}^n x_i.$$

Si riconosce subito che questa derivata si annulla nel punto

$$\hat{\lambda} = \frac{1}{n} \sum_{i=1}^n x_i$$

e dunque lo stimatore di massima verosimiglianza per il parametro λ della legge di Poisson coincide con la media empirica del campione che, com'è noto, è uno stimatore corretto.

Esempio 2.1.19 Nel 1998, a Berkeley, in California, il numero di incidenti stradali in 10 giornate senza pioggia scelte a caso è stato

4	0	6	5	2	1	2	0	4	3
---	---	---	---	---	---	---	---	---	---

Attraverso questi dati, si stimi per quell'anno la probabilità di giornate senza pioggia in cui ci sono stati al più 2 incidenti.

Siccome vi è un elevato numero di automobilisti, ciascuno dei quali ha solo una piccola probabilità di essere coinvolto in un incidente stradale, è ragionevole assumere che il numero d'incidenti quotidiani sia una variabile aleatoria di Poisson. Visto che uno stimatore del parametro della legge di Poisson è la media empirica, questo può essere stimato come

$$\bar{\lambda} = \frac{x_1 + \dots + x_{10}}{10} = 2.7.$$

Poiché, come abbiamo detto nel paragrafo 2.1.2, una volta scelto lo stimatore per il parametro ci si deve impegnare a utilizzare il valore ottenuto come stima di esso, se,

sullo stesso modello statistico, indichiamo con X una variabile aleatoria (indipendente dal campione) con legge di Poisson di parametro λ , che rappresenta il numero di incidenti stradali in un giorno senza pioggia, si otterrà che la probabilità desiderata è data da:

$$P^{\bar{\lambda}}(X \leq 2) = \frac{\bar{\lambda}^0}{0!}e^{-\bar{\lambda}} + \frac{\bar{\lambda}^1}{1!}e^{-\bar{\lambda}} + \frac{\bar{\lambda}^2}{2!}e^{-\bar{\lambda}} = \left[1 + 2.7 + \frac{(2.7)^2}{2}\right] e^{-2.7} \approx 0.4936.$$

Quindi, secondo la nostra stima, la probabilità che vi siano al più due incidenti in una giornata senza pioggia è di poco meno del 50%.

Esempio 2.1.20 (Stima del parametro della legge geometrica) Andiamo a calcolare lo stimatore di massima verosimiglianza per un campione di legge geometrica di parametro θ . Dette x_1, \dots, x_n le osservazioni del campione, la funzione di massima verosimiglianza è:

$$\begin{aligned} \mathcal{M}(\theta) &= P^\theta(X_1 = x_1, \dots, X_n = x_n) \\ &= \theta(1 - \theta)^{x_1-1} \dots \theta(1 - \theta)^{x_n-1} \\ &= \theta^n (1 - \theta)^{\sum_{i=1}^n x_i - n} \end{aligned}$$

Consideriamo adesso la funzione

$$\mathcal{L}(\theta) = \ln \mathcal{M}(\theta) = n \ln \theta + \left(\sum_{i=1}^n x_i - n \right) \cdot \ln(1 - \theta)$$

e deriviamo quest'ultima rispetto al parametro θ :

$$\frac{\partial \mathcal{L}}{\partial \theta} = \frac{n}{\theta} - \frac{\sum_{i=1}^n x_i - n}{1 - \theta} = \frac{n(1 - \theta) + n\theta - \theta \sum_{i=1}^n x_i}{\theta(1 - \theta)} = \frac{n - \theta \sum_{i=1}^n x_i}{\theta(1 - \theta)}.$$

Si riconosce subito che questa derivata si annulla nel punto

$$\hat{\theta} = \frac{n}{\sum_{i=1}^n x_i}$$

e dunque lo stimatore di massima verosimiglianza per il parametro di un campione con legge geometrica sarà:

$$T = \frac{n}{\sum_{i=1}^n X_i}.$$

Esempio 2.1.21 (Stima del parametro della legge esponenziale) Calcoliamo lo stimatore di massima verosimiglianza per un campione X_1, \dots, X_n di legge esponenziale di parametro sconosciuto λ . Indicando al solito con x_1, \dots, x_n i possibili valori dell'osservazione, la funzione di massima verosimiglianza sarà:

$$\mathcal{M}(\lambda) = f_\lambda(x_1) \cdots f_\lambda(x_n) = \prod_{i=1}^n \lambda e^{-\lambda x_i} = \lambda^n e^{-\lambda(x_1 + \dots + x_n)}.$$

Consideriamo adesso la funzione

$$\mathcal{L}(\lambda) = \ln \mathcal{M}(\lambda) = n \ln \lambda - \lambda \sum_{i=1}^n x_i$$

e deriviamo rispetto a λ :

$$\frac{\partial \mathcal{L}}{\partial \lambda} = \frac{n}{\lambda} - \sum_{i=1}^n x_i.$$

Si riconosce subito che questa derivata si annulla nel punto

$$\hat{\lambda} = \frac{n}{\sum_{i=1}^n x_i}$$

e dunque lo stimatore di massima verosimiglianza per il parametro di un campione con legge esponenziale sarà:

$$\Lambda = \frac{n}{\sum_{i=1}^n X_i}.$$

Esempio 2.1.22 (Stima per la media e la varianza della legge normale) Dato un campione X_1, \dots, X_n di legge normale $\mathcal{N}(\mu, \sigma^2)$, calcoliamo gli stimatori di massima verosimiglianza per μ e per σ^2 . Denotati con x_1, \dots, x_n i parametri, la funzione di massima verosimiglianza è:

$$\mathcal{M}(\mu, \sigma) = \frac{1}{(\sqrt{2\pi}\sigma)^n} \prod_{i=1}^n \exp \left[-\frac{(x_i - \mu)^2}{2\sigma^2} \right],$$

e dunque il suo logaritmo è:

$$\mathcal{L}(\mu, \sigma) = -n \ln \sqrt{2\pi} - n \ln \sigma - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2.$$

Deriviamo dapprima rispetto alla variabile μ :

$$\frac{\partial \mathcal{L}}{\partial \mu} = \frac{1}{\sigma^2} \sum_{i=1}^n (x_i - \mu) = \frac{1}{\sigma^2} \left[\sum_{i=1}^n x_i - n\mu \right].$$

Si riconosce subito che questa derivata si annulla nel punto

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^n x_i.$$

Deriviamo ora rispetto alla variabile σ :

$$\frac{\partial \mathcal{L}}{\partial \sigma} = -\frac{n}{\sigma} + \frac{1}{\sigma^3} \sum_{i=1}^n (x_i - \mu)^2.$$

Sostituendo a μ il valore $\hat{\mu}$, si trova subito che la derivata rispetto a σ si annulla nel punto

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \hat{\mu})^2.$$

Le due derivate di $\mathcal{L}(\mu, \sigma)$ rispetto a μ ed a σ si annullano dunque per $\hat{\mu} = \frac{1}{n} \sum_{i=1}^n x_i$ e $\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \hat{\mu})^2$. Notiamo che, per la media, lo stimatore di massima verosimiglianza coincide con la media empirica (che è uno stimatore corretto). Per lo stimatore della varianza, invece, osserviamo che esso coincide con lo stimatore $\Sigma^2 = \frac{n-1}{n} S^2$, ove S^2 è la varianza empirica e quindi è uno stimatore distorto, perché $\mathbb{E}^{\mu, \sigma}[\Sigma^2] = \frac{n-1}{n} \mathbb{E}^{\mu, \sigma}[S^2] = \frac{n-1}{n} \sigma^2$.

Negli esempi che abbiamo fatto, gli stimatori di massima verosimiglianza non sono risultati molto differenti da quelli che già conoscevamo, tranne per lo stimatore del parametro della legge esponenziale e quello della varianza per un campione normale, dove Σ^2 è diverso dalla varianza empirica S^2 che già conoscevamo. In particolare Σ^2 è uno stimatore distorto. In generale, infatti, non è detto che gli stimatori di massima verosimiglianza siano corretti. Tuttavia, si può dimostrare che essi godono comunque di buone proprietà “asintotiche” (cioè, quando la taglia n del campione tende all’infinito, essi approssimano il parametro da stimare meglio di tutti gli altri). Anche lo stimatore Σ^2 stesso, in effetti, differisce da S^2 solo per una quantità che tende a 0 per $n \rightarrow \infty$, cioè quando la taglia del campione tende all’infinito. Non è dunque il caso di preoccuparsi troppo se uno stimatore risulterà distorto, poiché non è detto che lo stimatore migliore (nel senso di quello che minimizzi il rischio) sia necessariamente uno stimatore non distorto!

2.1.5 Lo stimatore dei momenti

In questo paragrafo vogliamo descrivere un ulteriore metodo per costruire uno stimatore del parametro $\psi(\theta)$ che, sebbene non troppo preciso, è però spesso piuttosto elementare e fornisce risultati che nella pratica sono piuttosto utili. A questo scopo, sia X_1, \dots, X_n un campione di taglia n di legge di densità f_θ (continua o discreta), dove il parametro θ varia in un insieme Θ . Indichiamo con $m_r(\theta)$ il momento di ordine r di una (e quindi ciascuna) delle X_i , ovvero:

$$m_r(\theta) = \mathbb{E}^\theta [X_i^r].$$

Supponiamo inoltre che il parametro da stimare $\psi(\theta)$ si possa scrivere in funzione dei primi k momenti delle X_i , cioè che esista una funzione g tale che:

$$\psi(\theta) = g(m_1(\theta), \dots, m_k(\theta)). \quad (2.3)$$

L’esempio più semplice, e che abbiamo già incontrato spesso, è quello in cui $\psi(\theta)$ sia la media. In questo caso, la relazione precedente diventa semplicemente $\psi(\theta) = m_1(\theta)$ e conosciamo già la media empirica come uno stimatore corretto. Facendoci guidare da questo caso particolare, invece, definiamo in generale i *momenti empirici* di ordine r del campione X_1, \dots, X_n come

$$M_r = \frac{X_1^r + \dots + X_n^r}{n}.$$

Il momento empirico di ordine r è uno stimatore corretto del momento $m_r(\theta)$. Infatti, ricordando che le variabili aleatorie X_1, \dots, X_n che compongono il campione hanno tutte la stessa legge, per ciascun indice i si avrà $\mathbb{E}^\theta[X_i^r] = m_r(\theta)$ e di qui si trae immediatamente:

$$\mathbb{E}^\theta[M_r] = \mathbb{E}^\theta \left[\frac{X_1^r + \dots + X_n^r}{n} \right] = \frac{1}{n} \{ \mathbb{E}^\theta[X_1^r] + \dots + \mathbb{E}^\theta[X_n^r] \} = \frac{1}{n} \cdot n m_r(\theta) = m_r(\theta).$$

L'idea del *metodo dei momenti* consiste semplicemente nel sostituire nel termine di destra della relazione (2.3) i momenti empirici a quelli teorici, ovvero nel considerare come stimatore di $\psi(\theta)$ la statistica

$$T = g(M_1, \dots, M_r),$$

che prende il nome di *stimatore dei momenti* di $\psi(\theta)$.

Come subito si riconosce, se si utilizza questo metodo prendendo come $\psi(\theta)$ la media μ , lo stimatore dei momenti non è altro che la media empirica $M_1 = \bar{X}$. Invece, per stimare la varianza σ^2 , osserviamo che

$$\sigma^2 = m_2(\theta) - m_1(\theta)^2$$

e dunque lo stimatore dei momenti di σ^2 è:

$$M_2 - M_1^2 = \frac{1}{n} \sum_{i=1}^n X_i^2 - \left(\frac{1}{n} \sum_{i=1}^n X_i \right)^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 = \frac{n-1}{n} S^2,$$

ovvero, come subito si riconosce, il medesimo stimatore che si trova con il metodo della massima verosimiglianza per le leggi normali.

Esempio 2.1.23 (Stima del parametro della legge esponenziale) Si assegni un campione statistico X_1, \dots, X_n di taglia n e di legge esponenziale $\mathcal{E}(\lambda)$. Sappiamo allora che $m_1 = 1/\lambda$, da cui, risolvendo rispetto a λ , segue immediatamente che è $\lambda = 1/m_1$. Dunque, lo stimatore dei momenti del parametro della legge esponenziale è $\Lambda = 1/M_1 = 1/\bar{X}$. Come si vede, si tratta dello stesso stimatore ottenuto con il metodo della massima verosimiglianza.

Esempio 2.1.24 (Stima dei parametri della legge gamma) Si consideri un campione statistico X_1, \dots, X_n di taglia n e di legge $\gamma(\alpha, \lambda)$. Sappiamo allora che:

$$m_1 = \frac{\alpha}{\lambda}, \quad m_2 = \frac{\alpha^2 + \alpha}{\lambda^2}$$

e di qui, risolvendo in α e λ , si trovano immediatamente le relazioni

$$\alpha = \frac{m_1^2}{m_2 - m_1^2}, \quad \lambda = \frac{m_1}{m_2 - m_1^2}.$$

Di conseguenza, due stimatori dei parametri α e λ sono rispettivamente:

$$A = \frac{M_1^2}{M_2 - M_1^2} = \frac{(\sum_{i=1}^n X_i)^2}{n \sum_{i=1}^n X_i^2 - (\sum_{i=1}^n X_i)^2}, \quad \Lambda = \frac{M_1}{M_2 - M_1^2} = \frac{n \sum_{i=1}^n X_i}{n \sum_{i=1}^n X_i^2 - (\sum_{i=1}^n X_i)^2}.$$

Volendo scrivere questi due stimatori in termini della media empirica e della deviazione standard empirica, basterà osservare che è $\bar{X} = M_1$ e $S^2 = \frac{n}{n-1}(M_2 - M_1^2)$ per ottenere immediatamente:

$$A = \frac{n\bar{X}^2}{(n-1)S^2}, \quad \Lambda = \frac{n\bar{X}}{(n-1)S^2}.$$

2.1.6 Gli intervalli di fiducia

Torniamo adesso al caso generale di un campione X_1, \dots, X_n di taglia n estratto da una popolazione di legge $(\mathcal{L}(\theta))_{\theta \in \Theta}$ su un assegnato modello statistico $(\Omega, \mathcal{A}, (P^\theta)_{\theta \in \Theta})$. La teoria degli stimatori puntuali permette di *approssimare* il valore del parametro sconosciuto $\psi(\theta)$, introducendo un opportuno stimatore $T = t(X_1, \dots, X_n)$ ed utilizzando le osservazioni raccolte x_1, \dots, x_n , semplicemente con il numero $\tau = t(x_1, \dots, x_n)$. Per esempio, nel caso in cui il campione sia estratto da una popolazione gaussiana di legge $\mathcal{N}(\mu, \sigma^2)$, abbiamo stabilito che la media empirica \bar{X} è uno stimatore corretto per μ , dunque un'approssimazione della media μ è il valor medio $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$ delle osservazioni.

Ciò non significa, tuttavia, che possiamo aspettarci che il valore approssimato che abbiamo trovato sia esattamente uguale al parametro sconosciuto, né talvolta interessa tanto il valore esatto del parametro sconosciuto, quanto più avere la sicurezza che questo sia maggiore (o minore) di una soglia critica. Per esempio, se il campione statistico si riferisce alle misure della temperatura di un certo componente tecnologico, per il quale sappiamo che esso si usura oltre una certa temperatura critica, sarà importante poter stabilire se la temperatura a regime del componente è sempre inferiore a quella critica. Per risolvere questo tipo di problemi, rispetto ad uno stimatore puntuale, è a volte preferibile poter produrre un intervallo (o, più in generale, un insieme) per il quale abbiamo una certa *fiducia*, che il parametro da stimare vi appartenga. Per ottenere un tale *intervallo di fiducia*, è necessario utilizzare la legge dello stimatore scelto. Per questo, diamo dapprima una definizione generale e, subito dopo, focalizziamo la nostra attenzione su due casi particolari e particolarmente importanti.

Definizione 2.1.25 Fissato un modello statistico $(\Omega, \mathcal{A}, (P^\theta)_{\theta \in \Theta})$, sia ψ una funzione reale definita su Θ ed indichiamo con E la sua immagine. Un'applicazione \mathcal{S} di Ω in $\mathcal{P}(E)$ si dice un *insieme aleatorio* se, per ogni elemento θ di Θ , l'insieme

$$\{\psi(\theta) \in \mathcal{S}\} = \{\omega \in \Omega : \psi(\theta) \in \mathcal{S}(\omega)\}$$

è un evento, ossia appartiene alla tribù \mathcal{A} . In tal caso, la probabilità $\kappa(\theta, \psi(\theta)) = P^\theta(\psi(\theta) \in \mathcal{S})$ si chiama la *curva di fiducia* dell'insieme aleatorio \mathcal{S} per il parametro sconosciuto $\psi(\theta)$.

Nel quadro descritto dalla definizione, osserviamo immediatamente che, se T_1 e T_2 sono due statistiche, l'applicazione \mathcal{S} definita per ogni ω in Ω da $\mathcal{S}(\omega) = [T_1(\omega), T_2(\omega)]$ è un insieme aleatorio, giacché, come subito si riconosce, vale l'uguaglianza

$$\{\psi(\theta) \in \mathcal{S}\} = \{T_1 \leq \psi(\theta) \leq T_2\} = \{T_1 \leq \psi(\theta)\} \cap \{T_2 < \psi(\theta)\}^c.$$

Analogamente, come subito si riconosce, fissato un numero reale positivo δ , se T è una statistica, sono intervalli aleatori anche le applicazioni \mathcal{S}_δ , \mathcal{S}_- e \mathcal{S}_+ definite per ogni ω in Ω da

$$\mathcal{S}_\delta(\omega) = [T(\omega) - \delta, T(\omega) + \delta], \quad (2.4)$$

$$\mathcal{S}_-(\omega) = (-\infty, T(\omega)], \quad (2.5)$$

$$\mathcal{S}_+(\omega) = [T(\omega), +\infty). \quad (2.6)$$

Tutto ciò premesso, possiamo finalmente dare la seguente definizione.

Definizione 2.1.26 Su un fissato modello statistico $(\Omega, \mathcal{A}, (P^\theta)_{\theta \in \Theta})$, sia α un numero reale compreso tra 0 e 1 e sia \mathcal{S} un insieme aleatorio. Si dice che \mathcal{S} è un *insieme di fiducia* per $\psi(\theta)$ di *livello* $(1 - \alpha)$ se, per ogni θ nell'insieme Θ , si ha

$$P^\theta(\psi(\theta) \in \mathcal{S}) \geq 1 - \alpha. \quad (2.7)$$

In particolare, se l'insieme aleatorio \mathcal{S} è della forma (2.4), esso si dice un *intervallo di fiducia bilatero*, mentre se è della forma (2.5) oppure della forma (2.6), esso si dice un *intervallo di fiducia unilatero*.

In altri termini, un insieme di fiducia di livello $(1 - \alpha)$ per il parametro sconosciuto $\psi(\theta)$ è un insieme aleatorio la cui curva di fiducia per il parametro sconosciuto $\psi(\theta)$ è sempre maggiore di un fissato livello $(1 - \alpha)$.

Nella pratica, supponiamo assegnato un opportuno modello statistico $(\Omega, \mathcal{A}, (P^\theta)_{\theta \in \Theta})$ e, su di esso, un campione statistico X_1, \dots, X_n di taglia n . Supponiamo inoltre che $T = t(X_1, \dots, X_n)$ sia uno stimatore del parametro sconosciuto $\psi(\theta)$. Si voglia ad esempio trovare un intervallo di fiducia bilatero del parametro sconosciuto. A questo scopo, osserviamo che la definizione (2.7) si può riscrivere in questo caso nella forma seguente:

$$P^\theta (|t(X_1, \dots, X_n) - \psi(\theta)| < \delta) \geq 1 - \alpha.$$

Se è nota (anche in approssimazione) la legge della statistica $T = t(X_1, \dots, X_n)$, sarà possibile anche conoscere la legge della statistica $U = |T - \psi(\theta)|$ e di qui, come subito si riconosce, per trovare l'intervallo di fiducia richiesto basterà scegliere opportunamente il parametro δ . Si otterrà, così facendo, un intervallo della forma

$$[t(X_1, \dots, X_n) - \delta, t(X_1, \dots, X_n) + \delta].$$

Se dunque indichiamo con x_1, \dots, x_n le osservazioni, ossia i dati raccolti nell'esperimento aleatorio, si potrà dire che, ad un livello di fiducia $(1 - \alpha)$, il valore del parametro sconosciuto $\psi(\theta)$ appartiene all'intervallo numerico

$$[t(x_1, \dots, x_n) - \delta, t(x_1, \dots, x_n) + \delta]. \quad (2.8)$$

È importante osservare che quest'ultimo intervallo perde il significato di probabilità: non è vero, infatti, che la probabilità che $\psi(\theta)$ appartenga a questo intervallo è pari a $(1 - \alpha)$; per questo si parla piuttosto di *fiducia* e non di *probabilità*: prima di introdurre le osservazioni, il problema era effettivamente aleatorio, ma nell'intervallo numerico (2.8) non c'è più niente di aleatorio.

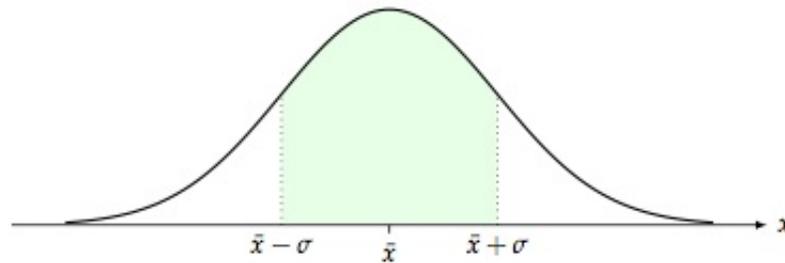
Esempio 2.1.27 (Stima del valore di una misura singola) Supponiamo di voler effettuare una singola misura di una grandezza fisica con uno strumento. Se indichiamo con σ la sensibilità dello strumento, si potrà pensare alla misura da effettuare come a un campione statistico X di taglia 1 su un opportuno modello statistico $(\Omega, \mathcal{A}, (P^\mu)_{\mu \in \mathbb{R}})$ di legge $\mathcal{N}(\mu, \sigma^2)$ rispetto alla misura di probabilità P^μ . Effettuata la misura, che indicheremo con \bar{x} , essa sarà una stima puntuale della media μ . Ora, per costruire un intervallo di fiducia per la media, ossia per il valore che si vuole misurare, basterà stimare la probabilità

$$P^\mu (|X - \mu| \leq \delta).$$

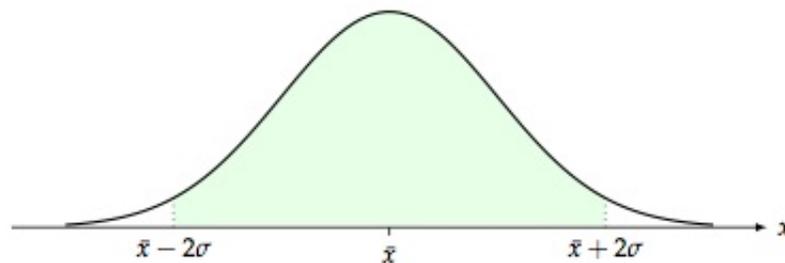
D'altra parte, poiché la statistica X è gaussiana, indicata con Z la sua standardizzata, cioè la variabile aleatoria $Z = (X - \mu)/\sigma$, sarà

$$P^\mu (|X - \mu| \leq \delta) = P^\mu \left(|Z| \leq \frac{\delta}{\sigma} \right) = 2\Phi \left(\frac{\delta}{\sigma} \right) - 1$$

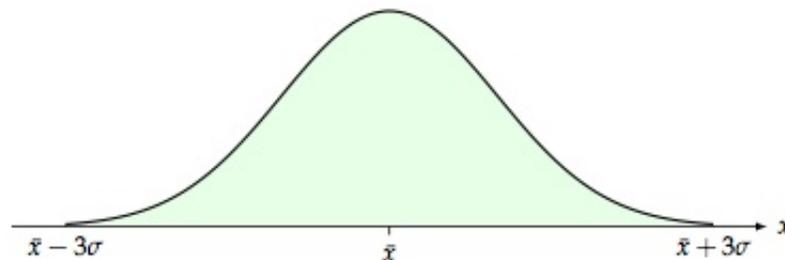
che, come subito si riconosce, non dipende da μ . Se scegliamo $\delta = k\sigma$ per un opportuno valore di k , un intervallo di fiducia della media sarà allora della forma $[\bar{x} - k\sigma, \bar{x} + k\sigma]$ ed il suo livello sarà $2\Phi(k) - 1$. In particolare, per $k = 1, 2, 3$ si ottengono tre intervalli di fiducia di livelli rispettivamente 68%, 95% e 99%, come mostrato nella figura sottostante.



(a) Quando si prende $k = 1$, si ottiene un intervallo di confidenza di livello 68 %.



(b) Quando si prende $k = 2$, si ottiene un intervallo di confidenza di livello 95 %.



(c) Quando si prende $k = 3$, si ottiene un intervallo di confidenza di livello 99 %.

Torniamo adesso al caso generale: su un opportuno modello statistico $(\Omega, \mathcal{A}, (P^\theta)_{\theta \in \Theta})$ sia dato un campione statistico X_1, \dots, X_n di taglia n e sia $T = t(X_1, \dots, X_n)$ uno stimatore puntuale del parametro sconosciuto $\psi(\theta)$. Siano poi x_1, \dots, x_n i dati raccolti dall'esperimento aleatorio. Abbiamo detto che un intervallo di fiducia bilatero al livello $(1 - \alpha)$ sarà in genere della forma

$$[t(x_1, \dots, x_n) - \delta, t(x_1, \dots, x_n) + \delta]$$

per un'opportuna scelta del parametro δ , eventualmente dipendente dai dati raccolti x_1, \dots, x_n . Osserviamo che tanto più quest'intervallo è "piccolo", tanto più sarà "precisa" l'approssimazione del parametro sconosciuto $\psi(\theta)$ con il suo valore stimato $\tau = t(x_1, \dots, x_n)$ attraverso lo stimatore scelto, giacché, con una fiducia di $(1 - \alpha)$, il valore del parametro sconosciuto appartiene a questo intervallo. Chiameremo allora la quantità numerica

$$\eta = \left| 1 - \frac{2\delta}{\tau} \right|$$

la *precisione relativa* del parametro $\psi(\theta)$ al livello $(1 - \alpha)$.

Esempio 2.1.28 Riprendendo l'esempio 2.1.27, osserviamo che la precisione relativa dell'intervallo di fiducia a livello $2\Phi(k) - 1$ è $\eta = \left| 1 - 2k \frac{\sigma}{\bar{x}} \right|$. Per esempio, se dalla misura abbiamo ottenuto il valore $\bar{x} = 5.4$ (in un'opportuna unità di misura) e se la sensibilità dello strumento utilizzato è $\sigma = 0.5$, al livello del 68% ($k = 1$) si avrà

$$\eta_{68\%} = 0.92 \approx 90\%$$

di precisione relativa; invece al livello del 95% ($k = 2$) si avrà una precisione relativa

$$\eta_{95\%} = 0.61 \approx 60\%.$$

2.1.7 Intervalli di fiducia per la media e per la varianza nei campioni gaussiani

Come abbiamo detto in molte occasioni, i campioni gaussiani hanno un'importanza particolare perché sono legati all'atto del misurare e, dunque, sono fondamentali nella pratica di laboratorio (sia esso di fisica, chimica, biologia, ecc.). Il resto di questo paragrafo sarà dunque dedicato a costruire gli intervalli di fiducia per la media e la varianza di un campione estratto da una legge gaussiana, sia pur limitatamente agli intervalli bilateri. Consideriamo a questo scopo un campione di taglia n di variabili aleatorie gaussiane X_1, \dots, X_n .

Un primo semplicissimo caso (ma raramente utile nelle applicazioni pratiche) è quello in cui si suppone che le osservazioni X_1, \dots, X_n abbiano tutte legge $\mathcal{N}(\mu, \sigma^2)$, dove la deviazione standard σ è conosciuta. Questo può essere, ad esempio, il caso in cui si stimi la deviazione standard con la sensibilità dello strumento di misura, ma si tratta di una stima che solo raramente ha significato perché presuppone, ad esempio, che le misure ripetute siano comunque molto precise. In questo caso, comunque, per il teorema di

Cochran (teorema 2.1.13), la media empirica \bar{X} ha legge normale di media μ e deviazione standard σ/\sqrt{n} . Dunque la variabile aleatoria

$$Z = \frac{\bar{X} - \mu}{\sigma} \sqrt{n}$$

avrà legge normale ridotta $\mathcal{N}(0, 1)$. Scelto allora un numero reale α compreso tra 0 e 1, cerchiamo un intervallo di fiducia di livello $(1 - \alpha)$ per μ . Basterà per questo osservare che è, similmente a quanto visto nell'esempio 2.1.27:

$$P^\mu (|\bar{X} - \mu| \leq \delta) = P^\mu \left(|Z| \leq \frac{\delta\sqrt{n}}{\sigma} \right) = 1 - \alpha$$

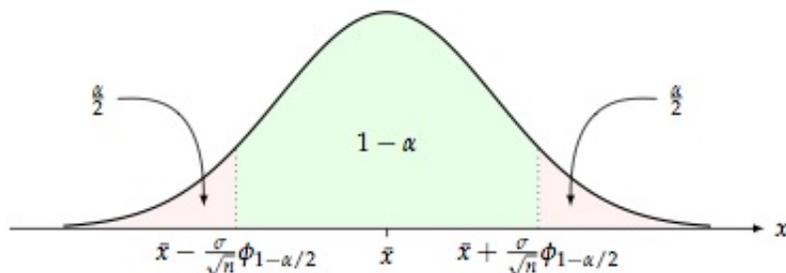
da cui si deduce che

$$\frac{\delta\sqrt{n}}{\sigma} = \phi_{1-\alpha/2}$$

e dunque

$$\left[\bar{X} - \frac{\sigma}{\sqrt{n}} \phi_{1-\alpha/2}, \bar{X} + \frac{\sigma}{\sqrt{n}} \phi_{1-\alpha/2} \right] \quad (2.9)$$

è un intervallo di fiducia di livello $(1 - \alpha)$ per μ (si veda la figura sottostante).



Intervallo di confidenza di livello $(1 - \alpha)$ per un campione gaussiano con varianza nota σ^2 .

In pratica, però, la deviazione standard σ è raramente nota o approssimabile con qualche dato certo dell'esperimento aleatorio, e quindi non è possibile calcolare esplicitamente l'intervallo (2.9). È allora ragionevole domandarsi se non si possa sostituire al posto di σ^2 il valore della varianza empirica S^2 , che è appunto uno stimatore corretto di σ^2 . A questo scopo, sempre con l'ausilio del teorema di Cochran, possiamo velocemente trovare un intervallo di fiducia per μ anche quando non sia nota la deviazione standard. Supponiamo dunque, in questo caso, che siano ignote sia μ che σ e consideriamo, come al solito, la media empirica \bar{X} e la varianza empirica S^2 come stimatori puntuali per la media e la varianza rispettivamente. Si ha allora, dato che la variabile aleatoria $T = \frac{\bar{X} - \mu}{S} \sqrt{n}$ ha legge di Student $t(n - 1)$:

$$1 - \alpha = P^{\mu, \sigma} (|T| \leq t_{1-\alpha/2}(n - 1)) = P^{\mu, \sigma} \left(\left| \frac{\bar{X} - \mu}{S} \sqrt{n} \right| \leq t_{1-\alpha/2}(n - 1) \right),$$

e ciò vuole dire che

$$\left[\bar{X} - \frac{S}{\sqrt{n}} t_{1-\alpha/2}(n - 1), \bar{X} + \frac{S}{\sqrt{n}} t_{1-\alpha/2}(n - 1) \right] \quad (2.10)$$

è un intervallo di fiducia per μ di livello $(1 - \alpha)$. Confrontando questo intervallo con quello dato dalla (2.9) (che si riferiva alla semplice situazione in cui la varianza σ^2 era conosciuta), si vede che, effettivamente, l'idea di sostituire alla varianza σ^2 il suo stimatore S^2 era una buona idea, a patto però di sostituire ai quantili della legge normale quelli della legge di Student $t(n - 1)$ che sono un po' più grandi, ma che, comunque, per campioni di taglia molto grande tendono ad assomigliarsi.

Esempio 2.1.29 La tabella seguente riporta cento misurazioni della velocità della luce nell'aria, effettuate dal fisico sperimentale Albert Abraham Michelson (1852–1931) tra il 5 giugno ed il 2 luglio 1879 (fonte: S.M. Stigler, *The Annals of Statistics* **5**, 1055–1098, 1977). I dati si devono intendere espressi in km/s, la velocità della luce è stimata come 299000 più il valore indicato.

850	740	900	1070	930	850	950	980	980	880	1000	980	930	650	760
810	1000	1000	960	960	960	940	960	940	880	800	850	880	900	840
830	790	810	880	880	830	800	790	760	800	880	880	880	860	720
720	620	860	970	950	880	910	850	870	840	840	850	840	840	840
890	810	810	820	800	770	760	740	750	760	910	920	890	860	880
720	840	850	850	780	890	840	780	810	760	810	790	810	820	850
870	870	810	740	810	940	950	800	810	870					

Domandiamoci qual è un intervallo di fiducia per la velocità della luce nell'aria al livello $1 - \alpha = 0.95$ sulla base di queste misurazioni. Come abbiamo già detto, le misure ripetute di una stessa quantità sperimentale sono un caso tipico in cui si assume che i valori ottenuti si possano modellizzare con un campione di legge normale. Possiamo quindi applicare i risultati di questo paragrafo: basta calcolare media e varianza empiriche, che risultano essere

$$\bar{X} = 852.40 \text{ km/s}, \quad S^2 = 6242.67 \text{ km/s}^2$$

e quindi l'intervallo di fiducia, al livello $1 - \alpha = 0.95$, è, con $\frac{S}{10} t_{0.975}(99) = 15.7$,

$$[852.4 - 15.7, 852.4 + 15.7] = [836.7, 868.1]$$

che si può anche esprimere, utilizzando la simbologia propria della pratica di laboratorio in fisica, dicendo che la velocità della luce nell'aria è

$$(299852.4 \pm 15.7) \text{ km/s}.$$

Oggi si sa che la velocità della luce nell'aria è 299711.347 km/s con un errore inferiore a 1 m/s, dunque le misurazioni di Michelson tendevano a sovrastimare, perché il valore più moderno è più piccolo dell'estremo inferiore dell'intervallo di fiducia costruito.

Concludiamo questo paragrafo calcolando un intervallo di fiducia anche per la varianza di un campione gaussiano. Costruiamo dapprima un intervallo unilatero. A questo scopo, iniziamo con l'osservare che la varianza σ^2 è un numero positivo, così come il suo stimatore S^2 . Inoltre, per il teorema di Cochran la variabile aleatoria $(n - 1)S^2/\sigma^2$ ha legge $\chi^2(n - 1)$. Si ha dunque, per definizione di quantile:

$$\begin{aligned} \alpha &= P^{\mu, \sigma} \left(\frac{n - 1}{\sigma^2} S^2 \leq \chi_\alpha^2(n - 1) \right) = \\ &= P^{\mu, \sigma} \left(\sigma^2 \geq \frac{(n - 1)S^2}{\chi_\alpha^2(n - 1)} \right) = 1 - P^{\mu, \sigma} \left(\sigma^2 \leq \frac{(n - 1)S^2}{\chi_\alpha^2(n - 1)} \right), \end{aligned}$$

da cui

$$1 - \alpha = P^{\mu, \sigma} \left(\sigma^2 \leq \frac{(n-1)S^2}{\chi_{\alpha}^2(n-1)} \right);$$

ciò significa che un intervallo di fiducia di livello $\beta = 1 - \alpha$ per la varianza è:

$$\left[0, \frac{n-1}{\chi_{\alpha}^2(n-1)} S^2 \right].$$

Veniamo adesso a costruire un intervallo bilatero. Si potrà ripetere facilmente il ragionamento fatto per l'intervallo unilatero, osservando però stavolta che è:

$$\begin{aligned} 1 - \alpha &= 1 - \frac{\alpha}{2} - \frac{\alpha}{2} \\ &= P^{\mu, \sigma} \left(\frac{(n-1)S^2}{\sigma^2} \leq \chi_{1-\alpha/2}^2(n-1) \right) - P^{\mu, \sigma} \left(\frac{(n-1)S^2}{\sigma^2} \leq \chi_{\alpha/2}^2(n-1) \right) \\ &= P^{\mu, \sigma} \left(\chi_{\alpha/2}^2(n-1) \leq \frac{(n-1)S^2}{\sigma^2} \leq \chi_{1-\alpha/2}^2(n-1) \right) \\ &= P^{\mu, \sigma} \left(\frac{(n-1)S^2}{\chi_{1-\alpha/2}^2(n-1)} \leq \sigma^2 \leq \frac{(n-1)S^2}{\chi_{\alpha/2}^2(n-1)} \right) \end{aligned}$$

da cui si ottiene che

$$\left[\frac{(n-1)S^2}{\chi_{1-\alpha/2}^2(n-1)}, \frac{(n-1)S^2}{\chi_{\alpha/2}^2(n-1)} \right] \quad (2.11)$$

è un intervallo di fiducia di livello $1 - \alpha$ per σ^2 .

Questi intervalli di fiducia sono sicuramente più desueti rispetto a quelli per la media; ciononostante essi hanno una loro utilità pratica molto importante. Solo per fare un semplice esempio, se si decide di stimare la deviazione standard di una misura con la sensibilità dello strumento, in modo tale da poter considerare il campione gaussiano come avente varianza nota, si potrà ad esempio verificare di non aver commesso una leggerezza, controllando se il valore della sensibilità appartiene a uno di questi intervalli.

Esercizi del §2.1

1. Per ogni numero reale θ , si consideri la funzione reale

$$f_{\theta}(x) = \begin{cases} c_{\theta} x^{\theta} & \text{per } x \in (0, 2), \\ 0 & \text{altrimenti.} \end{cases}$$

- (a) Determinare per quali valori del parametro θ la funzione f_{θ} può essere una densità di probabilità e, per questi valori di θ , calcolare la costante c_{θ} .
- (b) Calcolare la media e la varianza di una variabile aleatoria X , definita su un opportuno modello statistico $(\Omega, \mathcal{A}, (P^{\theta})_{\theta \in \Theta})$, dotata di densità f_{θ} .
- (c) Stimare con il metodo dei momenti il valore del parametro θ , quando siano stati campionati i seguenti dati:

1.6	0.5	0.6	1.7	1.4
0.8	1.2	1.1	1.3	0.9

2. Su un opportuno modello statistico $(\Omega, \mathcal{A}, (P^\theta)_{\theta \in \Theta})$, la statistica X abbia la seguente densità:

$$f_\theta(x) = \begin{cases} c_\theta(2x + \theta) & \text{per } x \in (0, 1), \\ 0 & \text{altrimenti.} \end{cases}$$

- Determinare per quali valori del parametro θ la funzione f_θ è una densità di probabilità e per tali valori calcolare la costante c_θ .
- Trovare la legge di X rispetto a P^θ .
- Calcolare la media e la varianza di X .
- Trovare una stima del parametro θ col metodo dei momenti, quando siano stati campionati i seguenti dati:

0.34	0.21	0.32	0.61
0.46	0.55	0.61	0.23

3. Per ogni numero reale θ maggiore di -3 , si consideri la funzione f_θ definita da

$$f_\theta(x) = \begin{cases} \frac{c_\theta}{x^{4+\theta}} & \text{per } x > 1, \\ 0 & \text{altrimenti.} \end{cases}$$

- Determinare il valore di c_θ che rende questa funzione una densità di probabilità.
- Utilizzando il metodo dei momenti, costruire uno stimatore per il parametro θ e stimarlo a partire dai seguenti dati:

3	5	4	11	8
7	14	5	4	2

4. Per ogni numero reale θ , si consideri la funzione reale

$$f_\theta(x) = \begin{cases} c_\theta x(1-x) + \theta & \text{per } x \in [0, 1], \\ 0 & \text{altrimenti.} \end{cases}$$

Determinare per quali valori del parametro θ , la funzione f_θ può essere una densità di probabilità e, per questi valori di θ , calcolare la costante c_θ .

- Si determini uno stimatore T di θ col metodo dei momenti.
- Calcolare la media di T .
- Applicando lo stimatore T appena definito, stimare il parametro θ , quando siano stati campionati i seguenti dati:

0.3	0.7	0.4	0.3
0.9	0.6	0.4	0.5

5. Si considerino i dati x_1, \dots, x_n che si ipotizzano essere realizzazioni di un campione statistico X_1, \dots, X_n definito su un opportuno modello statistico $(\Omega, \mathcal{A}, (P^\theta)_{\theta \in \Theta})$, dotato di legge di densità

$$f_\theta(x) = \begin{cases} c_\theta(\theta - |x|) & \text{per } -1 \leq x \leq 1, \\ 0 & \text{altrimenti,} \end{cases}$$

con θ numero reale positivo.

- (a) Determinare il valore di c_θ che renda la funzione f_θ una densità di probabilità.
 (b) Trovare uno stimatore T di θ con il metodo dei momenti.
 (c) Stimare il parametro θ , quando siano stati campionati i seguenti dati:

-0.03	0.25	0.16	-0.73
-0.87	0.05	0.02	0.01

6. Un'urna contiene un numero ignoto di monete n . Per stabilirne il numero si effettua l'esperimento aleatorio consistente nel lanciare ciascuna delle monete e contare il numero di teste uscite durante i lanci. Ripetendo questo esperimento, si ottengono i seguenti risultati sul numero di teste:

692	695	665	674	719
680	686	658	691	645

Si stimi il numero di monete presenti nell'urna.

7. Per ogni numero reale θ compreso tra 0 e 1, si consideri la seguente funzione reale

$$f_\theta(x) = \begin{cases} c_\theta x^{-\theta} & \text{per } 0 \leq x \leq 1, \\ 0 & \text{altrimenti.} \end{cases}$$

- (a) Si determini il valore di c_θ che rende questa funzione una densità di probabilità.
 (b) Si determini lo stimatore di massima verosimiglianza per θ .
 (c) Si calcoli una stima di θ se si sono registrate le seguenti osservazioni:

0.3	0.7	0.5	0.1	0.1
-----	-----	-----	-----	-----

8. Su un opportuno modello statistico $(\Omega, \mathcal{A}, (P^\theta)_{\theta \in \Theta})$, sia X_1, \dots, X_n un campione estratto da una popolazione con legge di densità:

$$f_\theta(x) = c_\theta e^{-\theta|x|}$$

- (a) Determinare per quali valori del parametro θ la funzione f_θ è una densità di probabilità, e per tali valori calcolare la costante c_θ .
 (b) Calcolare la media e la varianza del campione.
 (c) Trovare lo stimatore di massima verosimiglianza del parametro θ e stimare il parametro quando siano stati campionati i seguenti dati:

0.34	-0.13	0.27	-0.03
-0.11	0.83	0.36	-0.18

9. Si supponga che l'intensità del segnale emesso da un *access point* e percepito dalla scheda *wireless* di un *laptop* segua la statistica di Rayleigh, definita dalla densità

$$f_\rho(x) = \begin{cases} \frac{r}{\pi\rho} e^{-\frac{r^2}{2\pi\rho}} & \text{per } r > 0, \\ 0 & \text{altrimenti.} \end{cases}$$

- (a) Trovare lo stimatore di massima verosimiglianza R del parametro ρ .

(b) Stimare il parametro ρ quando la scheda registra i seguenti valori:

7	6	21	12
17	14	25	13

10. Su un opportuno modello statistico $(\Omega, \mathcal{A}, (P^\theta)_{\theta \in \Theta})$ sia X_1, \dots, X_n un campione estratto da una popolazione con legge di densità

$$f_\theta(x) = \begin{cases} \frac{\theta}{\sqrt{x}} e^{-\pi\theta^2 x} & \text{per } x > 0, \\ 0 & \text{altrimenti.} \end{cases}$$

Determinare lo stimatore di massima verosimiglianza di θ e trovare una stima di θ , quando si siano campionati i seguenti dati:

0.47	2.91	1.68	0.66	9.34	0.21
------	------	------	------	------	------

11. Si determini lo stimatore di massima verosimiglianza del parametro θ di un campione con legge di densità

$$f_\theta(x) = \begin{cases} e^{\theta-x} & \text{per } x \geq \theta, \\ 0 & \text{altrimenti.} \end{cases}$$

Si dia quindi una stima di θ , quando si siano campionati i seguenti dati:

4.2	6.3	7.1	5.8	8.3
-----	-----	-----	-----	-----

12. Determinare lo stimatore di massima verosimiglianza del parametro θ di un campione con legge di densità

$$f_\theta(x) = \begin{cases} \theta x^{\theta-1} & \text{per } x \in [0, 1], \\ 0 & \text{altrimenti.} \end{cases}$$

Si dia quindi una stima di θ , quando si siano campionati i seguenti dati:

0.23	0.46	0.78	0.63	0.90
------	------	------	------	------

13. Su un opportuno modello statistico $(\Omega, \mathcal{A}, (P^\theta)_{\theta \in \Theta})$, sia X_1, \dots, X_n un campione estratto da una popolazione con legge di densità:

$$f_\theta(x) = \begin{cases} c_\theta x & \text{per } x \in (0, \theta), \\ 0 & \text{altrimenti.} \end{cases}$$

- (a) Determinare il valore di c_θ che rende questa funzione una densità di probabilità.
 (a) Determinare lo stimatore di massima verosimiglianza di θ .
 (b) Determinare uno stimatore di θ col metodo dei momenti.
 (c) Dare una stima di θ per mezzo di entrambi gli stimatori, quando siano stati campionati i seguenti dati:

13.7	14.1	9.4	10.8	7.7	9.7
------	------	-----	------	-----	-----

14. Un campione di 14 individui estratto da una popolazione gaussiana ha media empirica $\bar{x} = 13$ e varianza empirica $s^2 = 2$. Sulla base dei dati, dare una stima di un ragionevole valore massimo rilevabile in una misurazione ad un livello di fiducia del 90%. Dare poi una stima di un ragionevole valore minimo allo stesso livello di fiducia.
15. Un campione di 16 individui estratto da una popolazione gaussiana ha media empirica $\bar{x} = 11$ e varianza empirica $s^2 = 4$. Sulla base dei dati, dare la stima di un ragionevole valore massimo rilevabile in una misurazione ad un livello di fiducia del 95%.
16. Un campione di 22 individui estratto da una popolazione gaussiana ha media empirica $\bar{x} = 9$ e varianza (nota) $\sigma^2 = 3$. Sulla base dei dati, dare una stima di un ragionevole valore minimo rilevabile in una misurazione, ad un livello di fiducia del 99%.
17. Un campione di 12 individui estratto da una popolazione gaussiana ha media empirica $\bar{x} = 7$ e varianza empirica $s^2 = 1$. Costruire un intervallo di fiducia bilatero per la media e uno per la varianza, entrambi al livello 95%.
18. Un campione estratto da una popolazione gaussiana ha media empirica $\bar{x} = 23$ e varianza empirica $s^2 = 5$. Quale dovrebbe essere, a parità di valori campionari, la taglia del campione per ottenere una precisione relativa di 10^{-3} , con un livello di fiducia del 98%?
19. Una bilancia elettronica ha sensibilità $\sigma = 0.001$ g. Si pesa un oggetto e si raccolgono i seguenti dati espressi in grammi:

3.142	3.163	3.155	3.150	3.141
-------	-------	-------	-------	-------

- (a) Trovare un intervallo di fiducia al 95% per il peso dell'oggetto, assumendo come deviazione standard delle misurazioni la sensibilità della bilancia.
 - (b) Determinare, a parità di media empirica, quante volte l'oggetto dovrebbe essere pesato, per ottenere un intervallo di fiducia al 95% che determini il peso con una precisione relativa di 10^{-3} .
 - (c) Trovare un intervallo di fiducia unilatero destro di livello del 95% per la media assumendo questa volta ignota la deviazione standard.
 - (d) Sempre nell'ipotesi che la deviazione standard sia sconosciuta, determinare un intervallo di fiducia unilatero sinistro di livello del 95% per la varianza.
20. Gli oggetti prodotti da una linea di produzione devono avere dimensioni comprese tra 2.4 mm e 2.6 mm. Esaminandone 20 si trovano una media empirica $\bar{x} = 2.54$ mm e una deviazione standard empirica di $s = 0.05$ mm. Stimare la frazione di oggetti le cui dimensioni saranno al di fuori delle dimensioni accettabili.
 21. La misura della concentrazione di mercurio di un campione di pesci ha dato i seguenti valori:

11.2	12.2	10.4	11.8	12.6
10.2	11.1	12.3	12.4	10.8

Trovare intervalli di fiducia unilateri destri al 90%, 95% e 99%. Ripetere il problema sapendo che la deviazione standard è nota e pari a 0.8.

22. Un campione di 20 sigarette ha un contenuto medio di 1.2 mg di nicotina, con una deviazione standard empirica di 0.2 mg. Determinare intervalli di fiducia bilateri al 90%, 95% e 99% per il contenuto di nicotina. Trovare poi un valore c tale che, con fiducia del 99% il contenuto di nicotina sia maggiore di c .
23. Si misura il diametro di un campione di rondelle, ottenendo i seguenti dati espressi in millimetri:

6.68	6.76	6.78	6.76	6.74	6.64
6.81	6.74	6.70	6.66	6.67	6.66

Trovare un intervallo di fiducia al 95% per la media e al 99% per la varianza del diametro della rondella.

24. Si considerino i dati x_1, \dots, x_n raccolti da un campione statistico X_1, \dots, X_n definito su un opportuno modello statistico $(\Omega, \mathcal{A}, (P^\theta)_{\theta \in \Theta})$ con legge di densità:

$$f_\theta(x) = \begin{cases} c_\theta(\theta - |x|) & \text{per } x \in [-2, 2], \\ 0 & \text{altrimenti,} \end{cases}$$

dove sia $\theta \geq 2$.

- (a) Determinare il valore di c_θ che rende questa funzione una densità di probabilità.
- (b) Trovare uno stimatore puntuale T del parametro θ .
- (c) Nel caso di n grande si indichi un intervallo di fiducia approssimato al 95% per il parametro θ .
25. Si supponga che i tempi di attesa alla fermata di un autobus abbiano una distribuzione esponenziale. Nel corso di una giornata si raccoglie una serie di 20 misure del tempo di attesa, ottenendo un tempo medio di $\bar{x} = 13$ min. Trovare un opportuno intervallo di fiducia per il parametro della legge esponenziale ad un livello di fiducia del 95%.

2.2 I test d'ipotesi statistiche

2.2.1 Introduzione

Un tipico problema di statistica consiste nello stabilire se il parametro sconosciuto $\psi(\theta)$ sia di un certo tipo oppure no. In effetti, succede molto spesso in numerosi campi di applicazione (come la medicina, la scienza, la tecnologia, l'industria) di dover trovare una risposta del tipo “sì” o “no” ad una domanda. In tutti questo genere di situazioni, anziché stimare direttamente il parametro incognito, si formula una qualche ipotesi circa il parametro sconosciuto e si utilizza il campione statistico di dati raccolti per verificare questa ipotesi. Per chiarire questo concetto, supponiamo che una casa farmaceutica produca un determinato farmaco, vendendolo in confezioni da 30 pastiglie, e che sia dichiarato sulla confezione che ogni pastiglia debba contenere 25 mg di principio attivo. Volendo scoprire se quanto dichiarato dalla casa farmaceutica corrisponde al vero, supponiamo di prendere una scatola e di misurare la quantità di principio attivo (mediante apposite strumentazioni di laboratorio), ottenendo una media di 24.0 mg con

una deviazione standard di 0.5 mg. Ci si può domandare, allora, se il valore ottenuto sia *compatibile* con quanto dichiarato dalla casa farmaceutica, oppure se esso sia inferiore. Per formalizzare questo problema, occorrerà suddividere i possibili valori del parametro sconosciuto θ in due classi che siano tra loro disgiunte e la cui riunione contenga tutti i possibili valori di θ . In altri termini, occorrerà scegliere una partizione $\{\Theta_0, \Theta_A\}$ dell'insieme Θ costituito da tutti i possibili valori assunti da θ . Un'ipotesi statistica corrisponderà allora a domandarsi se il parametro θ si trovi nella parte Θ_0 oppure nella parte Θ_A . Tradizionalmente, si usa dire che $\mathbf{H}_0 : \theta \in \Theta_0$ è l'ipotesi, e $\mathbf{H}_A : \theta \in \Theta_A$ l'alternativa. Nell'esempio della ditta farmaceutica, si potrebbe prendere:

$$\mathbf{H}_0 : \mu < 25 \text{ mg}, \quad \mathbf{H}_A : \mu = 25 \text{ mg}.$$

Lo scopo di un test è quello di scegliere tra due possibilità: *respingere oppure no l'ipotesi*. Anche quando si dirà “*si accetta l'ipotesi*” s'intenderà sempre affermare che “non siamo in grado di rifiutare l'ipotesi”. Per questo, sarà sufficiente stabilire qual è l'insieme D formato da tutti i possibili esiti dell'esperimento che conducono al rigetto (o rifiuto) dell'ipotesi. Chiameremo questo insieme la *regione critica* del test e la interpreteremo nel modo seguente: se il risultato dell'esperimento cadrà in D , allora lo sperimentatore s'impegnerà a ritenere che l'ipotesi \mathbf{H}_0 sia falsa, e, come tale, la respingerà. Viceversa, se il risultato dell'esperimento cadrà nell'insieme D^c , che prende anche il nome di *regione d'accettazione* del test, lo sperimentatore s'impegnerà ad accettare l'ipotesi \mathbf{H}_0 . Precisamente, possiamo dare la seguente definizione.

Definizione 2.2.1 Assegnato un modello statistico $(\Omega, \mathcal{A}, (P^\theta)_{\theta \in \Theta})$, sia $\{\Theta_0, \Theta_A\}$ una partizione dell'insieme Θ dei possibili valori del parametro sconosciuto θ . Un *test statistico* per la verifica dell'ipotesi

$$\mathbf{H}_0 : \theta \in \Theta_0 \quad \text{contro} \quad \mathbf{H}_A : \theta \in \Theta_A$$

è semplicemente una partizione (D, D^c) di Ω costituita a partire da un evento D , detto *regione critica* del test, con la regola secondo cui, qualunque sia il risultato ω dell'esperimento, se ω cade in D , allora si *rifiuta* l'ipotesi \mathbf{H}_0 , mentre se ω cade in D^c , si *accetta* l'ipotesi \mathbf{H}_0 .

In generale, assegnato un modello statistico, qualunque sia la scelta della regione critica di un test, se l'ipotesi è vera, c'è una probabilità positiva di avere un'osservazione nella regione critica e quindi di respingere a torto l'ipotesi: quello che si chiama un *errore di prima specie*. D'altra parte, vi è una probabilità positiva di non respingere un'ipotesi falsa, e questo si chiama un *errore di seconda specie*. Tradizionalmente, c'è una certa asimmetria tra l'ipotesi e l'alternativa: come ipotesi si considera sempre il caso peggiore dei due (per esempio, nel caso di un medicinale, l'ipotesi è che questo non sia efficace) e dunque l'errore di prima specie è un errore molto più grave di quello di seconda specie. È come dire: mettere in commercio un farmaco inefficace, e forse dannoso, è peggio che non mettere in commercio un farmaco efficiente.

Dato un test, denotiamo con D la sua regione critica. Si chiama allora la *potenza* del test l'applicazione $\theta \mapsto P^\theta(D)$ di Θ in $[0, 1]$. Quando θ è un elemento di Θ_0 , il

numero $P^\theta(D)$ è la probabilità di respingere a torto l'ipotesi, cioè la probabilità di commettere un errore di prima specie supposto che il vero valore del parametro sia θ . Invece, quando θ appartiene a Θ_A , il numero $P^\theta(D)$ è la probabilità che il test di regione critica D porti davvero all'ipotesi, supponendo che il vero valore del parametro sia θ , dunque la probabilità di commettere un errore di seconda specie è $1 - P^\theta(D)$.

L'estremo superiore dei numeri della forma $P^\theta(D)$, con $\theta \in \Theta_0$ si chiama anche il *livello* del test di regione critica D . È chiaro che il livello del test corrisponde all'estremo superiore di tutte le probabilità di compiere un errore di prima specie. In generale, poiché, come si è detto, l'errore di prima specie è considerato più grave di un errore di seconda specie, si cerca di determinare una regione critica che abbia un valore del livello pari ad un prefissato numero reale α (tipicamente i valori sono $\alpha = 0.1, 0.05, 0.01$).

Esempio 2.2.2 La settimana successiva al suicidio di un famoso matematico napoletano, nella città si sono registrati 12 suicidi, contro una media di 8. Si può dire che vi sia stato un fenomeno d'imitazione?

Se supponiamo che ogni cittadino abbia una probabilità p di suicidarsi, e se supponiamo che il fatto che una persona ceda a questo atto non influenzi il comportamento degli altri, giungiamo a modellizzare il numero X di suicidi con una legge binomiale $\mathcal{B}(n, p)$, dove n è il numero degli abitanti. Poiché è ragionevole supporre che p (la probabilità che un singolo individuo si uccida) sia molto piccola, e n molto grande, si può approssimare questa legge con la legge di Poisson di parametro $\lambda = np$. Arriviamo dunque ad affermare che, in condizioni normali, il numero di suicidi si può modellizzare con una variabile aleatoria X avente legge di Poisson di parametro 8.

Negare l'ipotesi, ossia sostenere che vi è stato un fenomeno d'imitazione, significa dire che ora la variabile aleatoria X segue una legge, sempre di Poisson, ma di parametro λ maggiore di 8. Usiamo dunque come modello statistico un campione (di taglia 1) avente legge di Poisson di parametro θ , con $\theta \in \Theta = \mathbb{R}_+$.

In questo caso, l'ipotesi è $\Theta_0 = \{8\}$ contro l'alternativa $\Theta_1 = (8, +\infty)$. Un modo ragionevole di affrontare questo test è di stabilire di respingere l'ipotesi se il valore di X è troppo grande. Se fissiamo il livello al valore $\alpha = 0.05$, scegliamo come regione critica $D = \{X \geq k\}$ dove k dev'essere tale che sia

$$\sup_{\lambda=8} P^\lambda(X \geq k) = P^8(X \geq k) \leq 0.05.$$

Calcolando numericamente la funzione di ripartizione della legge di Poisson di parametro 8, si trae

$$P^8(X \geq 12) = 0.112, \quad P^8(X \geq 13) = 0.064, \quad P^8(X \geq 14) = 0.034.$$

Poiché 14 è il più piccolo dei numeri k tali che risulti $P^8(X \geq k) \leq 0.05$, l'evento $\{X \geq 14\}$ è la regione critica di un test di livello 0.05.

Dunque l'ipotesi *non viene respinta*, perché il numero 12 non appartiene alla regione critica del test. In effetti, il numero 12 non è sufficientemente grande per stabilire il manifestarsi di un fenomeno sociale rilevante. Se invece si fossero osservati più di 14 suicidi, il dato sarebbe stato da considerarsi significativo (almeno al livello 0.05).

2.2.2 Il test di Student

Una classe importante di test riguarda la media di una popolazione. Supponiamo di osservare un campione X_1, \dots, X_n di variabili aleatorie indipendenti e di voler stabilire se la media μ del campione è eguale oppure no ad una quantità prefissata μ_0 . Si tratta quindi di realizzare un test per l'ipotesi “la media μ coincide con μ_0 ” contro l'alternativa “la media μ è diversa da μ_0 ”.

A questo scopo, consideriamo la media empirica del campione: $\bar{X} = \frac{1}{n}(X_1 + \dots + X_n)$, che è uno stimatore corretto di μ , e cerchiamo di determinare un numero δ maggiore di 0 in modo tale che, se l'ipotesi è vera, allora si abbia

$$P^\mu \{ |\bar{X} - \mu_0| > \delta \} = \alpha.$$

Per un tale valore di δ , l'evento $\{ |\bar{X} - \mu_0| > \delta \}$ sarà la regione critica di un test di livello α . Ora, il calcolo della probabilità di un evento di questo tipo è in generale molto complicato (per non dire impossibile) a meno di non disporre di talune informazioni aggiuntive sul campione. Supponiamo dunque che le variabili aleatorie X_1, \dots, X_n siano gaussiane, oppure che n sia abbastanza grande da poter applicare l'approssimazione normale. Sotto questa ipotesi sappiamo (per il teorema di Cochran) che, se poniamo $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$, la variabile aleatoria

$$T = \frac{\bar{X} - \mu}{S} \sqrt{n}$$

ha legge di Student $t(n-1)$. Di qui, osservato che sotto l'ipotesi si ha $\mu = \mu_0$, si trae:

$$P^\mu (|\bar{X} - \mu_0| > \delta) = P^{\mu_0} (|\bar{X} - \mu| > \delta) = P^{\mu_0} \left(|T| \geq \frac{\delta \sqrt{n}}{S} \right)$$

e, come abbiamo visto nel paragrafo 1.3.3, questa quantità vale α se risulta $\frac{\delta \sqrt{n}}{S} = t_{1-\alpha/2}(n-1)$, ovvero se

$$\delta = \frac{S}{\sqrt{n}} t_{1-\alpha/2}(n-1). \quad (2.12)$$

Per questo valore di δ , dunque, l'evento $\{ |\bar{X} - \mu_0| > \delta \}$ è una regione critica di livello α . La realizzazione del test consiste dunque nel verificare se la media empirica \bar{X} differisce da μ_0 , secondo la probabilità P^{μ_0} , per una quantità maggiore di δ , dove δ è dato appunto dalla (2.12). Oppure, in maniera più semplice, basterà calcolare la statistica $T_0 = \frac{\bar{X} - \mu_0}{S} \sqrt{n}$ e rigettare l'ipotesi se $|T_0|$ risulta più grande di $t_{1-\alpha/2}(n-1)$.

Esempio 2.2.3 L'altezza media degli uomini di un paese era di 170 cm nel 1957. Su $n = 81$ reclute alla visita di leva nel 1967 la media era $\bar{X} = 171$ cm con una varianza $S^2 = 16$ cm². Si può dire che l'altezza media sia cambiata ad un livello $\alpha = 0.05$?

Si tratta di verificare l'ipotesi “ μ coincide con $\mu_0 = 170$ cm” contro l'alternativa “ μ è diversa da $\mu_0 = 170$ cm”. Come abbiamo visto, si tratta di calcolare la statistica

$$|T_0| = \left| \frac{\bar{X} - \mu_0}{S} \right| \sqrt{n}$$

e di confrontarla con il numero $t_{0.975}(99)$. Sostituendo i valori, si ha:

$$\left| \frac{\bar{X} - \mu_0}{S} \sqrt{n} \right| = \left| \frac{171 - 170}{4} \cdot 9 \right| = 2.25,$$

mentre il quantile è $t_{0.975}(99) = 1.98$. Quindi il risultato del test è all'interno della regione critica e l'ipotesi $\mu = \mu_0$ è respinta. Si può dunque affermare che l'altezza media è effettivamente cambiata.

Talvolta, confrontando la media \bar{X} del campione con μ_0 si vuole soprattutto verificare che μ sia più grande di μ_0 (oppure più piccolo). Si considera allora il test per rigettare l'ipotesi “la media μ è inferiore o eguale a μ_0 ” contro l'alternativa “la media μ è superiore a μ_0 ”. Sappiamo che, se μ è il vero valore della media, la quantità

$$T = \frac{\bar{X} - \mu}{S} \sqrt{n}$$

ha legge di Student $t(n-1)$; dunque, se l'ipotesi è vera, ossia $\mu \leq \mu_0$, e poniamo

$$T_0 = \frac{\bar{X} - \mu_0}{S} \sqrt{n} = \frac{\bar{X} - \mu}{S} \sqrt{n} + \frac{\mu - \mu_0}{S} \sqrt{n},$$

allora l'ultimo termine della precedente eguaglianza è negativo e quindi si ha

$$T_0 = T + \frac{\mu - \mu_0}{S} \sqrt{n} \leq T$$

e dunque, per ogni $\mu \leq \mu_0$,

$$P^\mu (T_0 \geq t_{1-\alpha/2}(n-1)) \leq P^\mu (T \geq t_{1-\alpha/2}(n-1)) = \alpha$$

ovvero $\{T_0 \geq t_{1-\alpha/2}(n-1)\}$ è una regione critica di livello α .

I due test introdotti in questo paragrafo si chiamano *test di Student*. Il primo di questi si dice anche un test “bilatero” mentre il secondo un test “unilatero”.

2.2.3 Il test di Fisher–Snedecor

Nel paragrafo precedente abbiamo costruito un test per la media di una popolazione; occupiamoci ora di costruire un test per la varianza. Supponiamo a questo scopo di osservare un campione X_1, \dots, X_n di variabili aleatorie indipendenti e di voler stabilire se la varianza σ^2 del campione è più piccola, o no, di una certa quantità prefissata σ_0^2 . Si tratta quindi di realizzare un test per l'ipotesi “la varianza σ^2 è inferiore o eguale a σ_0^2 ” contro l'alternativa “la varianza σ^2 è maggiore di σ_0^2 ”.

Consideriamo la varianza empirica $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$, che è uno stimatore corretto di σ^2 , e cerchiamo di determinare un numero δ maggiore di 0 in maniera tale che, se l'ipotesi è vera, ossia $\sigma \leq \sigma_0$, allora si abbia

$$P^\sigma (S^2 > \sigma_0^2(1 + \delta)) = \alpha.$$

Per un tale valore di δ , l'evento $\{S^2 > \sigma_0^2(1 + \delta)\}$ sarà una regione critica di un test di livello α . Se supponiamo che le variabili aleatorie X_1, \dots, X_n siano gaussiane o, comunque, che n sia abbastanza grande da potersi applicare l'approssimazione normale, sappiamo, per il teorema di Cochran, che la variabile aleatoria

$$W = \frac{(n-1)S^2}{\sigma^2}$$

ha legge del chi-quadro $\chi^2(n-1)$. Di qui, osservato che, sotto l'ipotesi, risulta $\sigma^2 \leq \sigma_0^2$, si trae:

$$P^\sigma(S^2 > \sigma_0^2(1 + \delta)) = P^\sigma\left(W > \frac{\sigma_0^2}{\sigma^2}(1 + \delta)(n-1)\right) \leq P^\sigma(W > (1 + \delta)(n-1)).$$

Ora, quest'ultima quantità vale α se risulta $(1 + \delta)(n-1) = \chi_{1-\alpha}^2(n-1)$, ovvero se

$$\delta = \frac{\chi_{1-\alpha}^2(n-1)}{n-1} - 1. \quad (2.13)$$

Per questo valore di δ , dunque, l'evento $\{S^2 > \sigma_0^2(1 + \delta)\}$ è una regione critica di livello α . La realizzazione del test consiste dunque nel verificare che la varianza empirica S^2 sia maggiore della quantità δ data da (2.13). Oppure, in maniera più semplice, rivedendo il conto precedente si vede che basta calcolare la statistica $W_0 = \frac{(n-1)S^2}{\sigma_0^2}$ e rigettare l'ipotesi se il valore trovato risulta maggiore di $\chi_{1-\alpha}^2(n-1)$.

Esempio 2.2.4 Una macchina che riempie i barattoli di caffè funziona correttamente se il peso dei barattoli ha una varianza inferiore o eguale a 15 g^2 . Su un campione di 25 barattoli di caffè, si rileva una varianza empirica di 25 g^2 . Si può dire, ad un livello $\alpha = 0.01$, che vi è un malfunzionamento della macchina?

Si tratta di verificare l'ipotesi “ σ^2 è inferiore o eguale a $\sigma_0^2 = 15 \text{ g}^2$ ” contro l'alternativa “ σ^2 è maggiore di $\sigma_0^2 = 15 \text{ g}^2$ ”. Come abbiamo visto, si tratta di calcolare la statistica

$$W = \frac{(n-1)S^2}{\sigma_0^2}$$

e di confrontarla con il numero $\chi_{0.99}^2(24)$. Sostituendo i valori, si ha

$$W = \frac{24 \cdot 25}{15} = 40,$$

mentre il quantile è $\chi_{0.99}^2(24) = 42.980$. Quindi l'ipotesi è accettata e si può affermare che la macchina ha effettivamente un malfunzionamento al livello $\alpha = 0.01$.

Talvolta, confrontando la varianza S^2 del campione con σ_0^2 si vuole soprattutto sapere se questa coincide oppure no con σ_0^2 . Si considera allora il test per rigettare l'ipotesi “la varianza σ^2 coincide con σ_0^2 ” contro l'alternativa “la varianza σ^2 è diversa da σ_0^2 ”. Sappiamo in questo caso che, sotto l'ipotesi, la variabile aleatoria

$$W = \frac{(n-1)S^2}{\sigma^2} = \frac{(n-1)S^2}{\sigma_0^2}$$

ha legge $\chi^2(n-1)$ e dunque, ripetendo i passaggi fatti per arrivare alla (2.13), si ottiene che

$$\left\{ S^2 < \frac{\chi_{\alpha/2}^2(n-1)\sigma_0^2}{n-1} \right\} \cup \left\{ S^2 > \frac{\chi_{1-\alpha/2}^2(n-1)\sigma_0^2}{n-1} \right\}$$

è una regione critica di livello α .

I due test introdotti in questo paragrafo si chiamano *test di Fisher-Snedecor*. Il primo di essi è un test “unilatero” mentre il secondo è un test “bilatero”.

2.2.4 Il test del chi-quadro

I test che abbiamo incontrato fino a questo momento riguardavano delle quantità numeriche. In questo paragrafo vedremo invece un test che si applica per decidere se un campione segue una certa legge oppure no. Vediamo dapprima la situazione “classica” in cui il campione assume soltanto un numero finito di valori.

Supponiamo a questo scopo di avere un campione X_1, \dots, X_n di taglia n , a valori in un insieme *finito* $\{x_1, \dots, x_m\}$, denotiamo con Θ l'insieme formato da tutti i vettori $\boldsymbol{\theta} = (\theta_1, \dots, \theta_m)$, con $\theta_1 + \dots + \theta_m = 1$, e poniamo, per ciascun indice j compreso tra 1 e m ,

$$P^{\boldsymbol{\theta}}(X_1 = x_j) = \theta_j.$$

Così, la legge del campione è determinata non appena si conosca il vettore $\boldsymbol{\theta}$, che rappresenta proprio la “densità discreta” della suddetta legge secondo $P^{\boldsymbol{\theta}}$. Noi vogliamo stabilire se il campione segue la legge corrispondente ad un certo parametro $\boldsymbol{\theta}_0 = (p_1, \dots, p_m)$ che, senza ledere la generalità, possiamo supporre formato da numeri strettamente positivi. Si tratta dunque di realizzare un test per l'ipotesi “*il campione segue la legge determinata da $\boldsymbol{\theta}_0$* ” contro l'alternativa “*il campione segue una legge differente da quella determinata da $\boldsymbol{\theta}_0$* ”. A questo scopo, per ciascun indice j compreso tra 1 e m , poniamo

$$O_j(\omega) = \#\{i \in \{1, \dots, n\} : X_i(\omega) = x_j\}.$$

Questa variabile aleatoria altro non è che il numero di osservazioni che hanno dato il valore x_j e viene chiamato l'*effettivo empirico* di x_j . Definiamo anche l'*effettivo teorico* di x_j , ponendo $E_j = np_j$. Questa quantità indica il numero di volte nelle quali, in teoria, dovremmo aspettarci di trovare il risultato x_j se la legge del campione fosse veramente quella stabilita da $\boldsymbol{\theta}_0$. Poniamo infine

$$T = \sum_{j=1}^m \frac{(O_j - E_j)^2}{E_j}. \quad (2.14)$$

Ora, nel quadro appena descritto, il *teorema di Pearson* afferma che la statistica T (detta, appunto, la *statistica di Pearson*) ha “approssimativamente” legge $\chi^2(m-1)$, purché n sia abbastanza grande. Applicando questo risultato, si vede subito che l'evento $\{T > \chi_{1-\alpha}^2(m-1)\}$ è una regione critica di livello α . La realizzazione del test consiste, dunque, nel calcolare la statistica T , e nel rigettare l'ipotesi se questa è maggiore del numero $\chi_{1-\alpha}^2(m-1)$. Questo che abbiamo appena descritto è il classico *test*

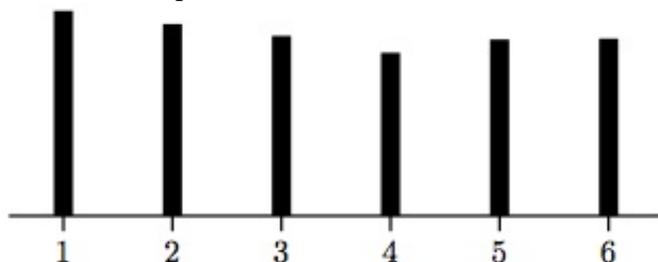
del chi-quadro.

Come nel caso dell'approssimazione normale, non entriamo nel dettaglio della questione su quanto debba essere grande n perché l'approssimazione possa applicarsi. Tradizionalmente, l'approssimazione si considera valida se n è sufficientemente grande in modo da avere $E_j = np_j \geq 5$ per ciascun indice j .

Esempio 2.2.5 Un dado viene lanciato 2400 volte con i seguenti risultati:

1	2	3	4	5	6
450	421	395	358	387	389

L'andamento degli effettivi empirici è riportato qui sotto. Si può affermare, ad un livello $\alpha = 0.05$, che il dado è equilibrato?



Effettivamente, il risultato 1 è apparso un numero di volte sensibilmente maggiore degli altri. In questo caso gli effettivi teorici sono $E_1 = E_2 = \dots = E_6 = 2400/6 = 400$, che è un numero largamente superiore a 5. Possiamo dunque applicare tranquillamente il test del chi-quadro.

x_j	O_j	E_j	$O_j - E_j$	$(O_j - E_j)^2/E_j$
1	450	400	50	6.25
2	421	400	21	1.10
3	395	400	-5	0.06
4	358	400	-42	4.41
5	387	400	-13	0.42
6	389	400	-11	0.30

Si tratta di calcolare la statistica T e di confrontarla con il numero $\chi_{0.95}^2(5)$. Sostituendo i valori, si ha $T = 12.54$ mentre il quantile è $\chi_{0.95}^2(5) = 11.07$. Poiché la statistica produce un valore maggiore del quantile, l'ipotesi che il dado sia equilibrato è rigettata.

Senza grossi sforzi, il test del chi-quadro può essere adattato al caso in cui il campione assume una quantità numerabile o addirittura continua di valori. Per riconoscerlo, consideriamo un campione X_1, \dots, X_n di variabili aleatorie, che supporremo avere valori reali, senza cioè nessuna restrizione. Denotiamo con F_θ la funzione di ripartizione del campione secondo P_θ . Poniamo cioè:

$$F_\theta(t) = P^\theta(X_1 \leq t) \quad \text{per ogni } t \in \mathbb{R}.$$

Fissata allora una qualsiasi funzione di ripartizione F , vogliamo costruire un test per l'ipotesi "il campione ammette F come propria funzione di ripartizione" contro l'alternativa "il campione non ammette F come propria funzione di ripartizione". A questo scopo,

scegliamo una suddivisione x_1, x_2, \dots, x_{m-1} della retta reale, con $x_1 < x_2 < \dots < x_{m-1}$, e poniamo:

$$I_1 = (-\infty, x_1], \quad I_2 = (x_1, x_2], \quad \dots, \quad I_{m-1} = (x_{m-2}, x_{m-1}], \quad I_m = (x_{m-1}, +\infty).$$

Possiamo così definire, a partire da X_1, \dots, X_n , n variabili aleatorie discrete Y_1, \dots, Y_n , a valori nell'insieme finito $\{1, 2, \dots, m\}$ nel modo seguente:

$$Y_i = k \quad \text{se e solo se} \quad X_i \in I_k.$$

Risulta allora, evidentemente, $P^\theta(Y_1 = k) = P^\theta(X_1 \in I_k) = F_\theta(x_k) - F_\theta(x_{k-1})$. Possiamo dunque applicare il test del chi-quadro alle variabili aleatorie Y_1, \dots, Y_n per verificare l'ipotesi che la legge sia determinata dal parametro $\theta_0 = (p_1, \dots, p_m)$, ove $p_k = F(x_k) - F(x_{k-1})$.

Notiamo che, nella scelta degli m numeri reali x_1, \dots, x_{m-1} c'è una vasta dose di arbitrarietà. Essi, comunque, dovranno essere scelti "abbastanza vicini" tra di loro; altrimenti si potrebbe correre il rischio di non distinguere tra leggi poco diverse tra loro. Tuttavia, se la suddivisione è troppo piccola, potrebbe capitare che qualche p_k sia piccolo, e dunque che risulti $E_k = np_k < 5$, violando così la tradizionale condizione di attendibilità del test. In genere, dunque, bisogna ricorrere ad un compromesso, da valutare caso per caso.

Esempio 2.2.6 Nella tabella sottostante sono riportati 66 numeri. Si può affermare, ad un livello $\alpha = 0.05$, che si tratta di un campione estratto da una legge gaussiana $\mathcal{N}(0, 1)$?

-0.83	0.30	-1.22	-0.91	0.28	-1.76	0.81	0.20	-0.07	0.71	0.44
-0.11	-1.63	-1.66	0.36	-0.55	-1.80	0.78	2.47	0.24	-1.27	-0.31
-0.76	-1.08	-1.56	-2.85	-0.77	0.05	1.01	-0.96	0.51	-1.39	-0.42
-0.42	1.18	-0.64	-0.59	-0.60	1.69	1.15	1.32	0.79	-0.49	-0.77
0.84	0.71	-1.06	0.07	0.34	0.20	-1.88	-0.86	-0.86	-0.46	0.29
-0.66	-1.50	1.87	-0.37	1.43	1.10	0.50	-0.67	0.08	-0.77	0.05

Si tratta dunque di scegliere la suddivisione x_1, \dots, x_{m-1} in modo opportuno. Poiché siamo interessati a stabilire se si tratta di un campione gaussiano, la funzione di ripartizione è Φ , e i numeri p_k saranno:

$$p_1 = \Phi(x_1), \quad p_m = 1 - \Phi(x_{m-1}),$$

$$p_k = \Phi(x_k) - \Phi(x_{k-1}), \quad k = 2, 3, \dots, m-1.$$

Per semplificare il conto, supponiamo che sia $p_k = 1/m$ per ciascun indice k . In questo modo, come induttivamente subito si riconosce, si ha:

$$x_k = \phi_{k/m}, \quad k = 1, \dots, m-1.$$

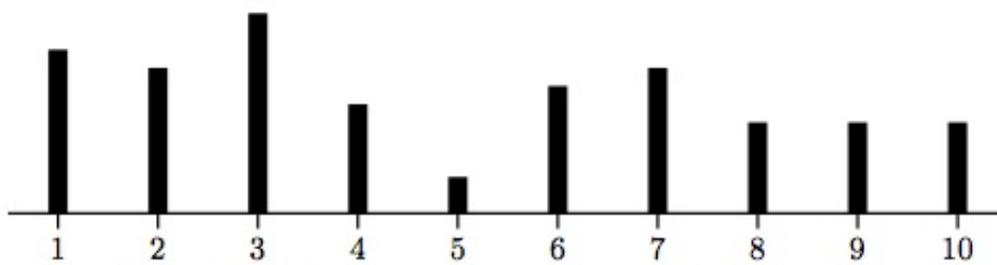
Al solito, affinché il test del chi-quadro si possa applicare dev'essere $np_k \geq 5$, ovvero, in questo caso, $66/m \geq 5$, o ciò ch'è lo stesso, $m \leq 66/5 = 13.2$. Come si vede, il numero m delle suddivisioni non può essere troppo grande. Prendiamo per semplicità $m = 10$ e calcoliamo, per mezzo delle tavole, i quantili $\phi_{k/10}$, con $k = 1, 2, \dots, 9$ (ricordiamo che, in virtù della relazione evidente $\Phi(-x) = 1 - \Phi(x)$, risulta $\phi_\alpha = -\phi_{1-\alpha}$).

$\phi_{0.1}$	$\phi_{0.2}$	$\phi_{0.3}$	$\phi_{0.4}$	$\phi_{0.5}$	$\phi_{0.6}$	$\phi_{0.7}$	$\phi_{0.8}$	$\phi_{0.9}$
-1.28	-0.84	-0.52	-0.25	0.00	0.25	0.52	0.84	1.28

Occorre ora fare la ripartizione in classi: la prima classe è composta dalle osservazioni che si trovano nell'intervallo $(-\infty, -1.28]$ (che sono 9); la seconda è quella formata dalle osservazioni che si trovano nell'intervallo $(-1.28, -0.84]$ (che sono 8), e così via fino all'ultima classe, formata dalle osservazioni che si trovano nell'intervallo $(1.28, +\infty)$. Alla fine, si ottiene la seguente tabella:

k	O_k	E_k	$O_k - E_k$	$(O_k - E_k)^2/E_k$
1	9	6.6	2.4	0.87
2	8	6.6	1.4	0.30
3	11	6.6	4.4	2.93
4	6	6.6	-0.6	0.03
5	2	6.6	-4.6	3.21
6	7	6.6	0.4	0.02
7	8	6.6	1.4	0.30
8	5	6.6	-1.6	0.39
9	5	6.6	-1.6	0.39
10	5	6.6	-1.6	0.39

Gli effettivi empirici sono meglio visualizzati nella figura sottostante.



Si tratta adesso di calcolare la statistica T e di confrontarla con il numero $\chi_{0.95}^2(9)$. Sostituendo i valori, si ha $T = 8.85$ mentre il quantile è $\chi_{0.95}^2(9) = 16.92$. Poiché la statistica produce un valore minore del quantile, l'ipotesi che i numeri seguano una legge normale $\mathcal{N}(0, 1)$ non è respinta.

È bene tener presente che questo genere di test, fatto per una legge continua, se da una parte è di semplice esecuzione, dall'altra parte è di scarsa potenza: esso porta cioè al rigetto dell'ipotesi solo se il discostamento dalla legge teorica è notevole, oppure se la taglia del campione è grande.

Sarebbe molto utile riuscire ad adattare il test del chi-quadro per studiare se le osservazioni seguono una legge appartenente ad una data *famiglia* di leggi (Poisson, binomiali, normali, e via dicendo), invece che ad una singola legge. L'idea naturale che potrebbe venire in mente è quella di scegliere uno stimatore corretto per il parametro sconosciuto, e quindi fare il test del chi-quadro alla legge teorica ottenuta con questo parametro sconosciuto. In realtà questo modo di procedere (molto utilizzato nella pratica) non è molto corretto. In effetti, esistono dei risultati teorici che garantiscono che, se gli stimatori sono scelti in maniera opportuna, allora la statistica T ha ancora una legge che

converge ad una legge χ^2 , ma con un numero di gradi di libertà diverso. Si tratta di un risultato molto difficile da dimostrare, soprattutto perché gli stimatori opportuni non sono facili da determinare; ad esempio, nel caso dell'adattamento ad una legge normale, i due stimatori classici \bar{X} e S^2 non vanno bene (anche se nella pratica vengono utilizzati lo stesso). Ad ogni modo cerchiamo di darne un'idea grossolana.

La situazione si presenta nel modo seguente: si vuole stabilire se un campione X_1, \dots, X_n segua una legge appartenente ad una famiglia $\mathcal{Q}(\theta_1, \dots, \theta_r)$ dipendente dagli r parametri $\theta_1, \dots, \theta_r$. Per prima cosa, è necessario stimare i parametri $\theta_1, \dots, \theta_r$ con i loro stimatori di massima verosimiglianza, che indichiamo con $\hat{\theta}_1, \dots, \hat{\theta}_r$. Dunque possiamo calcolare gli effettivi teorici a partire dalle probabilità p_1, \dots, p_m , calcolate attraverso la legge $\mathcal{Q}(\hat{\theta}_1, \dots, \hat{\theta}_r)$. Allora, se l'ipotesi è vera, cioè se il campione segue una legge appartenente alla famiglia $\mathcal{Q}(\theta_1, \dots, \theta_r)$, la statistica T ha legge $\chi^2(m - r - 1)$ (cioè si devono togliere tanti gradi di libertà quanti sono i parametri stimati).

Vediamone un semplice esempio riguardante la legge di Poisson.

Esempio 2.2.7 In un parco nazionale inglese è stata effettuata un'indagine per studiare la distribuzione del numero di tane di volpe. Sono stati ispezionati a questo scopo 95 ettari di bosco, e sono state rilevate le seguenti tane:

n. di tane = x_j	0	1	2	3	4
n. di ettari = O_j	19	30	20	14	12

Se si suppone che le volpi scelgano il luogo dove costruire la propria tana "a caso", come già sappiamo sarà naturale pretendere che il numero di tane abbia legge di Poisson. In altri termini: se X denota il numero di tane presenti in un ettaro, la nostra ipotesi è " X ha legge $\mathcal{P}(\lambda)$ ". D'altra parte, poiché il parametro λ non è noto, esso dev'essere stimato utilizzando il suo stimatore di massima verosimiglianza. Nel nostro caso esso coincide con la media empirica; si ha, dunque:

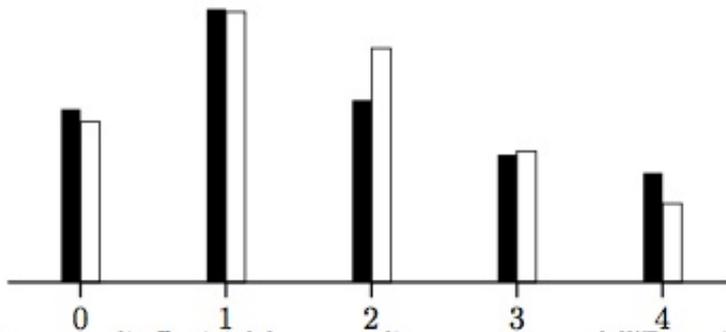
$$\bar{X} = \frac{0 \cdot 19 + 1 \cdot 30 + 2 \cdot 20 + 3 \cdot 14 + 4 \cdot 12}{95} = 1.68.$$

Poniamo dunque $\lambda = 1.68$ e calcoliamo gli effettivi teorici.

$$\begin{aligned} P^\lambda(X = 0) &= e^{-\lambda} = 0.19, & E_0 &= 95 \cdot 0.19 = 17.63 \\ P^\lambda(X = 1) &= \lambda e^{-\lambda} = 0.31, & E_1 &= 95 \cdot 0.31 = 29.69 \\ P^\lambda(X = 2) &= \frac{\lambda^2}{2} e^{-\lambda} = 0.26, & E_2 &= 95 \cdot 0.26 = 25.01 \\ P^\lambda(X = 3) &= \frac{\lambda^3}{3!} e^{-\lambda} = 0.15, & E_3 &= 95 \cdot 0.15 = 14.4 \\ P^\lambda(X \geq 4) &= 1 - P(X < 4) = 0.09, & E_4 &= 95 \cdot 0.09 = 8.63. \end{aligned}$$

Si tratta di calcolare la statistica T e di confrontarla con il quantile della legge del chi-quadro. Poiché gli effettivi teorici sono stati costruiti *a partire da un parametro stimato*, la statistica T non dovrà essere confrontata con il quantile $\chi_{0.95}^2(4)$. Essa dovrà invece essere confrontata con il quantile $\chi_{0.95}^2(3) = 7.815$. Andando a sostituire i valori,

si ottiene $T = 2.43$. Poiché la statistica produce un valore minore del quantile, l'ipotesi che il numero di tane per ettaro segua una legge di Poisson di parametro $\lambda = 1.68$ non può essere rigettata. Nella figura sottostante riportiamo il confronto fra gli effettivi empirici (in nero) e gli effettivi teorici (in bianco) del numero di tane per ettaro.



Concludiamo questo paragrafo analizzando un esempio riguardante le leggi normali.

Esempio 2.2.8 La tabella seguente riporta risultati di cento misurazioni in cm della lunghezza di una matita.

13.79	13.56	14.05	14.05	13.44	12.96	13.22	12.44	14.27	13.08	13.41	13.17
13.18	13.36	13.32	13.41	13.75	13.79	13.80	13.31	13.18	13.15	12.32	14.01
12.88	13.04	13.05	13.66	12.45	13.81	13.31	13.01	13.80	13.11	13.46	13.11
13.21	13.42	12.96	14.16	14.07	13.44	13.40	13.69	13.41	13.52	13.32	13.72
13.69	14.12	13.32	13.49	14.15	13.44	12.76	13.33	12.40	13.70	12.52	13.04
13.65	14.01	13.12	14.23	13.73	13.39	13.12	13.20	13.09	14.33	12.69	13.67
12.59	13.63	12.65	13.13	13.76	12.83	12.95	14.13	13.12	12.11	13.93	12.28
14.81	14.12	12.97	12.41	13.76	12.95	13.61	14.25	13.36	12.91	13.55	14.36
13.52	13.30	13.83	13.05								

Vogliamo stabilire se, ad un livello $\alpha = 0.05$, il campione risulta estratto da una legge normale. A questo scopo, poiché i parametri μ e σ^2 non sono noti, occorrerà stimarli entrambi attraverso i loro stimatori di massima verosimiglianza. Nel nostro caso lo stimatore della media coincide con la media empirica, mentre lo stimatore per la varianza coincide con lo stimatore (non corretto) $\hat{\Sigma}^2 = \frac{n-1}{n}S^2$, ove S^2 è la varianza empirica (si veda l'esempio 2.1.22). Si ha dunque:

$$\hat{\mu} = 13.39, \quad \hat{\sigma}^2 = 0.27.$$

Si tratta ora di vedere se le misurazioni risultano estratte da una legge $\mathcal{N}(\hat{\mu}, \hat{\sigma}^2)$. A questo scopo, è consigliabile ridurci ad una legge normale ridotta $\mathcal{N}(0, 1)$, trasformando ogni misurazione x nel modo seguente: $x \mapsto x - \frac{\hat{\mu}}{\hat{\sigma}}$. Scegliamo poi una suddivisione x_1, \dots, x_{m-1} in modo che sia soddisfatta la solita regola d'applicabilità del test del chi-quadro. Poniamo al solito

$$\begin{aligned} p_1 &= \Phi(x_1), & p_m &= 1 - \Phi(x_{m-1}), \\ p_k &= \Phi(x_k) - \Phi(x_{k-1}) & (k &= 2, 3, \dots, m-1). \end{aligned}$$

Se scegliamo, come nell'esempio 2.2.6, $p_k = 1/m$ per ciascun indice k , si riconosce subito che è $x_k = \phi_{k/m}$ per ogni k , e dunque, se vogliamo che sia $np_k \geq 5$, ovvero $100/m \geq 5$, dev'essere $m \leq 100/5 = 20$. Prendiamo, per esempio, $m = 10$ e calcoliamo, per mezzo delle tavole, i quantili $\phi_{k/10}$, con $k = 1, 2, \dots, 9$.

$\phi_{0.1}$	$\phi_{0.2}$	$\phi_{0.3}$	$\phi_{0.4}$	$\phi_{0.5}$	$\phi_{0.6}$	$\phi_{0.7}$	$\phi_{0.8}$	$\phi_{0.9}$
-1.28	-0.84	-0.52	-0.25	0.00	0.25	0.52	0.84	1.28

Occorre adesso suddividere le misurazioni in classi. Il risultato della suddivisione è riassunto nella tabella seguente:

k	O_k	E_k	$O_k - E_k$	$(O_k - E_k)^2/E_k$
1	11	10	1	0.1
2	6	10	-4	1.6
3	12	10	2	0.4
4	11	10	1	0.1
5	10	10	0	0
6	12	10	2	0.4
7	6	10	-4	1.6
8	15	10	5	2.5
9	5	10	5	2.5
10	12	10	2	0.4

Per concludere si tratta dunque di calcolare la statistica T (vedi (2.14)) e di confrontarla con il quantile $\chi_{0.95}^2(7) = 14.067$. (In effetti, il numero totale n delle classi è 10, ed a questo devono essere sottratti i due parametri ed un ulteriore grado di libertà dettato dalla teoria, ottenendo così $10 - 2 - 1 = 7$.) Andando a sostituire i valori, si nota che la statistica $T = 9.6$ produce un valore inferiore al quantile, per cui l'ipotesi che le misurazioni siano estratte da una legge normale $\mathcal{N}(13.39, 0.27)$ non può essere rigettata.

Esercizi del §2.2

1. Una fabbrica produce chiodi di metallo di peso medio 25g. Poiché la produzione ha un costo troppo elevato, si decide di cambiare il processo di lavorazione dei chiodi. Per capire se il nuovo processo di lavorazione ha portato variazioni sul peso dei chiodi, si estrae un campione di 30 chiodi e se ne misura un peso medio $\bar{X} = 21$ g e una varianza $S^2 = 16$ g². Ad un livello $\alpha = 0.05$, si può dire che la media è cambiata?
2. In base all'esperienza degli anni precedenti, risulta che gli studenti universitari di un certo corso di laurea riportano, nell'esame di matematica, una votazione media di 23. Se un gruppo di 50 studenti dell'anno in corso riporta una valutazione media di 25, con una varianza di 16, si può accettare l'ipotesi che gli studenti non differiscono da quelli degli anni precedenti ad un livello $\alpha = 0.01$?
3. Una pasticceria confeziona pacchetti di biscotti con peso netto dichiarato di 350 grammi. Poiché il peso viene determinato automaticamente, un certo giorno, per controllare che non vi siano state variazioni significative, vengono scelte a caso e pesate 20 confezioni

che risultano avere un peso medio di 340 grammi con uno scarto quadratico medio di 15 grammi. Si può affermare, ad un livello $\alpha = 0.05$, che il peso medio è cambiato?

4. Il proprietario di una ditta afferma che il numero di suoi prodotti venduti giornalmente è stato di 1500 unità. Un impiegato della ditta vuole verificare che non ci sia stato un calo nelle vendite: egli considera un campione casuale di 36 giorni e osserva che in media sono state vendute 1450 unità, con uno scarto quadratico medio di 120 unità. Ad un livello $\alpha = 0.01$, si può concludere che il numero di vendite è calato?
5. Il responsabile di una compagnia di trasporti ritiene che il carico medio consegnato sia 450 t (tonnellate). Il responsabile di magazzino contesta l'affermazione e registra un campione casuale di 25 trasporti e trova che il carico medio corrisponde a 446 t, con uno scarto quadratico medio di 0.25 t. Ad un livello $\alpha = 0.05$ l'affermazione del responsabile può essere rigettata?
6. Una macchina dovrebbe fabbricare chiodi di lunghezza media di 10 cm con una varianza di 0.25 cm^2 . Per verificare che questi parametri non siano stati alterati dall'usura, viene esaminato un campione di 28 chiodi la cui lunghezza media risulta essere 9,89 cm con una varianza di 0.35 cm^2 . Ad un livello $\alpha = 0.01$ si può dire che i due parametri sono cambiati?
7. Uno strumento per la misurazione della quota di un aereo presenta una precisione misurata da una varianza pari a 0.1. Dopo un guasto e relative riparazioni, lo strumento viene reinserito nell'aereo, ma il suo funzionamento è sospetto. Vengono così eseguite 23 misurazioni che danno luogo ad una varianza pari a 0.16. Ad un livello $\alpha = 0.05$, si può dire che lo strumento funziona correttamente oppure no?
8. Un distributore di caffè è tarato in maniera tale da fornire 25 cm^3 di caffè con una varianza di 6 cm^6 . L'addetto alla manutenzione non è sicuro della corretta taratura del distributore e decide sottoporlo a test. Su un campione di 25 tazze di caffè, egli rileva una varianza di 10 cm^6 . Può egli affermare, ad un livello $\alpha = 0.1$, che il distributore è tarato male?
9. Un negozio di pasta fresca produce ravioli con una macchina che ha uno scarto quadratico medio di 0.5 g. Poiché la macchina consuma troppo, il negoziante decide di sostituirla con una più nuova e più tecnologica, e vuole vedere se la nuova macchina è per giunta più efficiente. A questo scopo, egli estrae un campione di 28 ravioli e osserva che essi hanno uno scarto quadratico medio di 0.25 g. Si può dire, ad un livello $\alpha = 0.05$, che la nuova macchina è più efficiente?
10. In cento pagine dattiloscritte da una segretaria, sono stati contrassegnati i seguenti numeri di errori per pagina:

n. di errori:	0	1	2	3	4	5	6
n. di pagine:	36	40	19	2	0	2	1

Questi risultati, ad un livello $\alpha = 0.05$, giustificano il dubbio che gli errori commessi abbiano una legge di Poisson?

11. Nella tabella sottostante sono riportati i valori della velocità del vento al suolo, espressa in nodi, registrati a Livorno lo scorso novembre.

19	17	19	12	16	16
15	23	23	23	18	11
13	10	12	22	21	18
26	28	14	14	16	15
15	18	23	22	21	15

Verificare che, ad un livello $\alpha = 0.01$, essi seguono una legge uniforme.

12. Durante un certo periodo, un apparecchiatura sottoposta a controllo ha prodotto lotti di 60 pezzi ciascuno; in 100 lotti è stata registrata la seguente distribuzione di pezzi difettosi:

pezzi difettosi:	0	1	2	3	4	5	6
n. di lotti:	11	32	26	14	12	4	1

Verificare, al livello di significatività $\alpha = 0.01$, se è possibile adattare a questa distribuzione empirica una legge di Poisson, stimandone il parametro.

13. Il numero di passeggeri di un autobus di linea è stato, durante la scorsa settimana lavorativa, il seguente:

lun.	mar.	mer.	gio.	ven.
53	24	32	44	39

Si può affermare, ad un livello $\alpha = 0.025$, che il numero di passeggeri al giorno segue una legge uniforme?

14. In 100 periodi di tempo di un minuto sono stati conteggiati i seguenti raggi cosmici:

conteggio:	1	2	3	4	5	6	7	8 o più
frequenza:	17	29	20	16	8	1	2	0

Verificare se questa distribuzione empirica segue una legge di Poisson ad un livello $\alpha = 0.01$.

15. Uno studente di fisica pisano misura la lunghezza di ottanta chiodini con un calibro ventesimale. I risultati che ha trovato, espressi in millimetri, sono riassunti nella tabella seguente:

20.35	20.85	21.90	20.05	19.40	20.50
21.40	19.75	20.50	19.90	21.05	20.65
20.15	19.90	21.15	19.95	18.60	20.25
21.65	19.70	20.60	20.95	20.90	20.10
20.90	21.10	21.15	19.95	19.90	19.10
21.35	19.90	21.25	21.10	22.55	21.05
21.40	19.65	19.85	21.15	19.30	19.05
20.75	19.90	20.05	20.50	20.50	20.55
20.20	19.60	20.25	21.25	20.30	20.10
21.55	20.15	20.55	21.50	21.05	20.55
20.60	20.75	21.60	20.70	21.40	20.60
20.50	20.10	21.40	19.50	19.35	20.85
20.95	20.55	21.60	22.05	21.40	21.55
20.95	19.55				

Si può affermare, ad un livello $\alpha = 0.1$, che i chiodini sono estratti da una legge gaussiana?

Capitolo 3

Cenni di statistica descrittiva

3.1 Rilevazione ed elaborazione dei dati

3.1.1 Introduzione

La *statistica descrittiva* è la disciplina che studia le metodologie di cui si serve uno sperimentatore per raccogliere, rappresentare ed elaborare i dati osservati ai fini dell'analisi di un certo fenomeno. Il problema è che, quando si raccoglie questo genere di informazioni, si ha a che fare con una mole notevole di dati grezzi. Di conseguenza, il primo problema che ci si trova ad affrontare è quello di sintetizzare la massa di dati grezzi in pochi numeri o indicatori particolarmente significativi, utilizzando metodiche grafiche o numeriche, che siano in grado di descrivere la massa di dati, senza alterarne il senso complessivo.

La statistica descrittiva analizza in particolare le caratteristiche di una *popolazione* (di individui, di animali, di grandezze numeriche), i cui elementi si chiamano *unità statistiche* o *modalità*. Le caratteristiche che si studiano possono essere di tipo qualitativo, oppure numerico, e devono essere comuni a tutti gli elementi della popolazione.

La tabella sottostante mostra alcuni esempi di popolazioni.

popolazione	unità statistica	caratteristica	qualità/numero
nati a Pisa nel 2017	bambino	sexso	qualità
studenti di ingegneria	persona	altezza, età, ecc.	numero
giorni dell'anno	giorno	temperatura	numero

In generale si parla di *caratteri* o *attributi* che sono presenti, eventualmente in un certo grado, o assenti negli elementi della popolazione. Noi ci occuperemo solo di caratteristiche di tipo numerico, per le quali si usa il termine di *variabili* (statistiche, non aleatorie). Pertanto le popolazioni oggetto di studio sono costituite da un insieme di numeri (le variabili) che costituiscono le misurazioni della caratteristica comune agli elementi della popolazione in oggetto.

La statistica descrittiva si articola in tre processi fondamentali: la rilevazione, la rappresentazione e l'elaborazione dei dati.

La *rilevazione dei dati* consiste nell'acquisire le informazioni sul fenomeno collettivo. Schematicamente consiste nelle fasi seguenti:

- descrizione del fenomeno oggetto dell'indagine;
- individuazione della popolazione e delle unità statistiche che la compongono;
- determinazione dei caratteri, cioè degli aspetti del fenomeno che si vogliono rilevare;
- raccolta dei dati;
- spoglio dei dati, ossia conteggio, ordinamento e classificazione.

Con la *rappresentazione dei dati* si descrivono mediante grafici o tabelle le caratteristiche dei dati rilevati.

L'*elaborazione dei dati* è l'insieme delle metodologie, per mezzo delle quali si ottengono indici di sintesi sui dati rilevati e si studiano relazioni statistiche tra gli indici stessi.

3.1.2 Ordinamento e frequenze

I dati grezzi di una variabile Z raccolti nella fase di rilevazione, ad esempio

$$z_1, z_2, \dots, z_r,$$

sono generalmente di difficile interpretazione; una prima operazione utile consiste nell'elencare i dati stessi secondo grandezza, ad esempio in ordine crescente:

$$y_1 \leq y_2 \leq \dots \leq y_r.$$

Il *rango* di un insieme di dati z_1, z_2, \dots, z_r è il campo di variazione dei dati, ossia la differenza $y_r - y_1$ tra il più grande e il più piccolo.

I dati numerici raccolti potranno essere in parte, o anche tutti, coincidenti: quindi, indicando con x_1, x_2, \dots, x_n , ove $n \leq r$, i valori effettivamente distinti della variabile Z , si ha

$$y_j \in \{x_1, x_2, \dots, x_n\} \quad \text{per ogni } j = 1, 2, \dots, r.$$

Se, per ogni $i = 1, 2, \dots, n$, indichiamo con r_i il numero di dati uguali a x_i , si avrà

$$\sum_{i=1}^n r_i = r.$$

I valori r_1, \dots, r_n sono le *frequenze assolute* con cui si presentano i dati x_1, \dots, x_n , mentre i valori

$$f_1 = \frac{r_1}{r}, \quad f_2 = \frac{r_2}{r}, \quad \dots, \quad f_n = \frac{r_n}{r}$$

sono le *frequenze relative*; ovviamente,

$$\sum_{i=1}^n f_i = 1.$$

Se i dati sono ordinati, si può definire la *frequenza cumulata*, riferita alla modalità x_k , come la somma delle frequenze dalla prima modalità x_1 fino a x_k . Si parlerà di *frequenza cumulata assoluta* o di *frequenza cumulata relativa*, a seconda che le frequenze che si sommano da 1 a k siano assolute o relative: dunque, la frequenza cumulata assoluta di x_k è data da

$$R_k = \sum_{i=1}^k r_i, \quad k = 1, 2, \dots, n,$$

mentre la frequenza cumulata relativa di x_k è data da

$$F_k = \sum_{i=1}^k f_i, \quad k = 1, 2, \dots, n.$$

Esempio 3.1.1 Supponiamo che in una classe di 28 ragazzi sia stato proposto il quesito: quale sport preferisci? Se le risposte sono state: il calcio (10 ragazzi), il tennis (4 ragazzi), la pallavolo (6 ragazzi), il nuoto (3 ragazzi), altro (5 ragazzi), possiamo riassumere l'indagine con la tabella delle frequenze assolute r_k e delle frequenze relative f_k riportata qui sotto.

sport	frequenza assoluta	frequenza relativa
calcio	10	0.36
tennis	4	0.14
pallavolo	6	0.21
nuoto	3	0.11
altro	5	0.18

In questo esempio, non potendo ordinare i dati, non si possono calcolare le frequenze cumulate R_k e F_k .

Esempio 3.1.2 Consideriamo la popolazione dell'esercizio precedente, alla quale stavolta viene formulata la domanda: quanti anni hai? Se le risposte sono state: 18 (4 persone), 19 (13 persone), 20 (7 persone), 21 (3 persone), 22 (1 persona), si ottiene la seguente tabella.

età	frequenza assoluta	freq. cum. assoluta	frequenza relativa	freq. cum. relativa
18	4	4	0.14	0.14
19	13	17	0.46	0.61
20	7	24	0.25	0.86
21	3	27	0.11	0.96
22	1	28	0.04	1
totale	28		1	

Se l'insieme di dati da studiare è troppo grande, si può pensare di raggrupparli in classi. Ad esempio, sia X una variabile che assume valori in un intervallo $[a, b]$: una *suddivisione in classi* consiste nel dividere $[a, b]$ in intervalli disgiunti (in genere, ma non necessariamente, di uguale ampiezza):

$$[a_0, a_1), \quad [a_1, a_2), \quad \dots, \quad [a_{m-1}, a_m], \quad \text{con } a_0 = a < a_1 < \dots < a_m = b.$$

I dati vengono raggruppati nelle rispettive classi di appartenenza, calcolando le *frequenze di classe assolute* n_1, \dots, n_m , oppure le *frequenze di classe relative* p_1, \dots, p_m . La frequenza n_k rappresenta il numero di dati appartenenti all'intervallo $[a_{k-1}, a_k)$, mentre la frequenza relativa p_k è pari al rapporto n_k/r , dove r è il numero dei dati osservati. Il numero delle classi deve essere scelto in modo che non siano troppe, nel qual caso in ogni classe ci sarebbero pochissimi dati, ma nemmeno troppo poche: in tal caso, infatti, si avrebbero molti elementi in poche classi e la rappresentazione risultante non sarebbe significativa, in quanto avremmo perso troppa informazione sulla distribuzione reale. Per prassi consolidata, dettata dall'esperienza, un buon compromesso è quello di scegliere un numero di classi m prossimo al valore $1 + \frac{10}{3} \log_{10} r$. I valori delle frequenze (assolute, relative, cumulate, cumulate relative) possono poi essere riportati in corrispondenti tabelle di frequenza.

Esempio 3.1.3 I risultati ottenuti da 74 studenti durante un test (il voto massimo è 250) sono riportati nella tabella sottostante.

65	158	114	183	124	94
76	203	120	145	177	123
81	121	150	90	137	213
25	186	103	105	194	129
36	40	164	55	173	213
103	97	246	200	159	67
144	106	238	218	156	147
73	108	46	230	151	148
184	89	111	206	157	126
64	118	151	236	137	237
84	196	134	205	187	148
149	185	132	160	168	143
155	161				

Raggruppiamo i dati in classi e costruiamo una tabella che riporti le frequenze, le frequenze cumulative e quelle relative. Seguendo la regola suggerita, essendo $1 + \frac{10}{3} \log_{10} r \approx 7.23$, si ricava un numero di classi arrotondato per eccesso pari a 8. Il voto minimo è 25, il voto massimo è 246, e di conseguenza il rango è 221. Arrotondando a 224, che è multiplo di 8, risultano 8 classi di ampiezza 28 e si ottiene la suddivisione qui sotto.

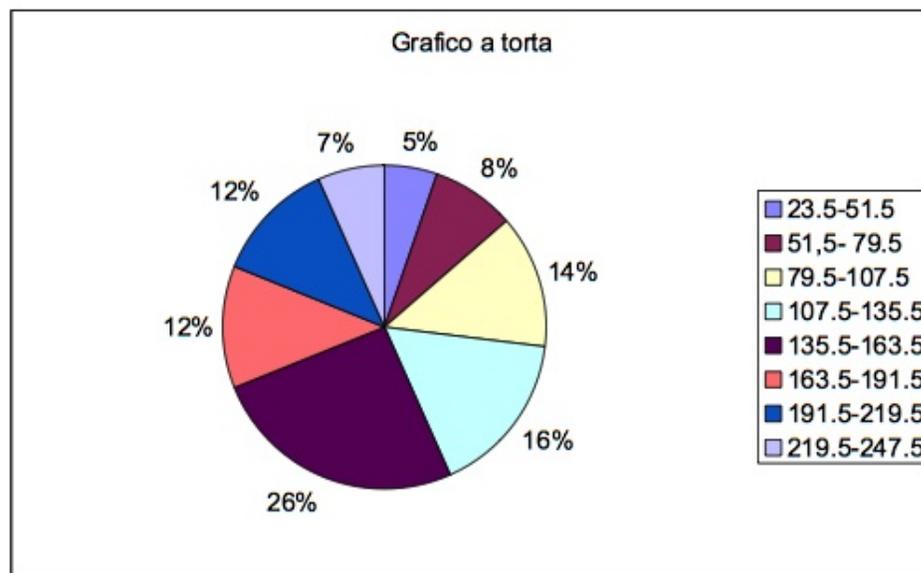
classe	centro di classe	frequenza di classe	frequenza relativa	frequenza cumulata	frequenza cum. relativa
23.5 - 51.5	37.5	4	0.054	4	0.054
51.5 - 79.5	65.5	6	0.081	10	0.135
79.5 - 107.5	93.5	10	0.135	20	0.270
107.5 - 135.5	121.5	12	0.162	32	0.432
135.5 - 163.5	149.5	19	0.257	51	0.689
163.5 - 191.5	177.5	9	0.122	60	0.811
191.5 - 219.5	205.5	9	0.122	69	0.932
219.5 - 247.5	233.5	5	0.067	74	1.000

Le tabelle di frequenza, pur contenendo molte informazioni, non consentono di cogliere a colpo d'occhio eventuali peculiarità significative presenti nei dati. Ciò è invece reso possibile dai diversi metodi di rappresentazione grafica.

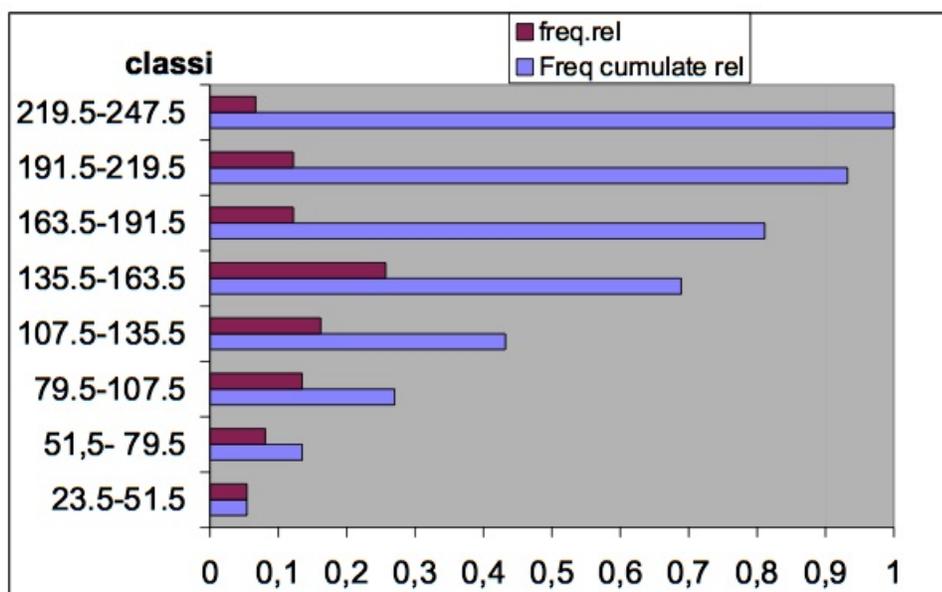
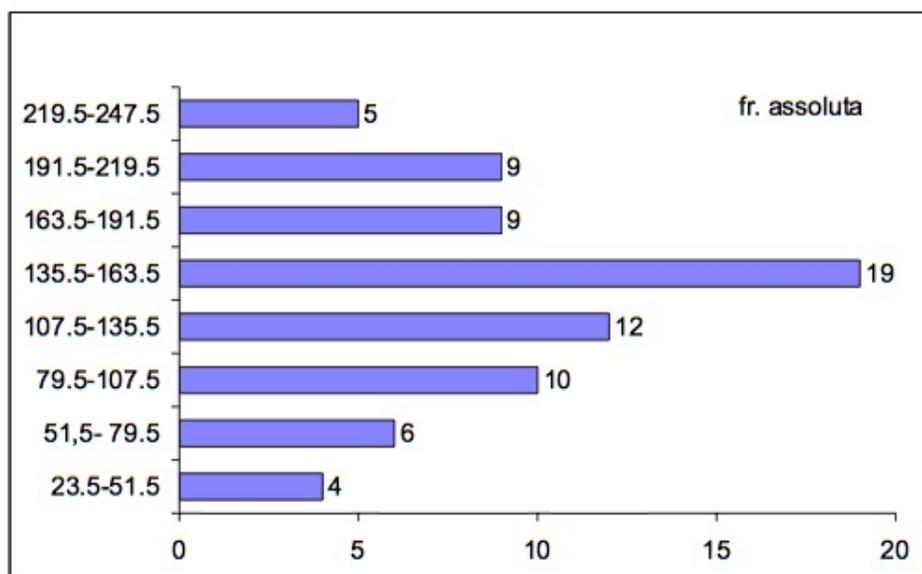
3.1.3 Rappresentazione grafica

Di seguito elenchiamo alcuni dei metodi di rappresentazione grafica più usati.

- **Diagrammi a torta:** si divide un cerchio in settori circolari che rappresentano le categorie considerate; ogni settore ha un'ampiezza proporzionale alla frequenza della corrispondente categoria. Il grafico sottostante si riferisce all'esempio 3.1.3 precedente.



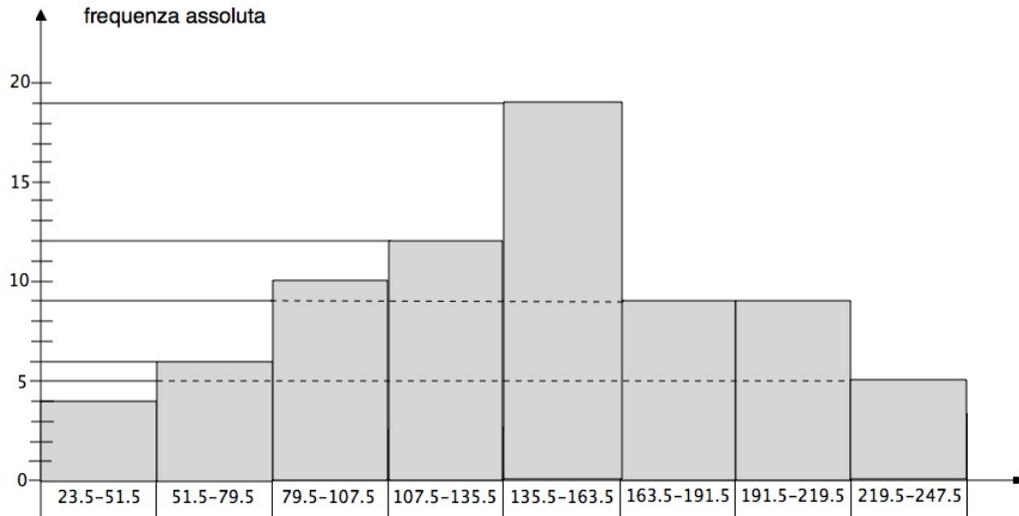
- **Grafici a barre:** ogni raggruppamento è rappresentato da una barra la cui lunghezza è proporzionale alla corrispondente frequenza. Qui sotto, nei due grafici a barre, riportiamo la frequenza assoluta e le frequenze relative, sempre nel caso dell'esempio 3.1.3.



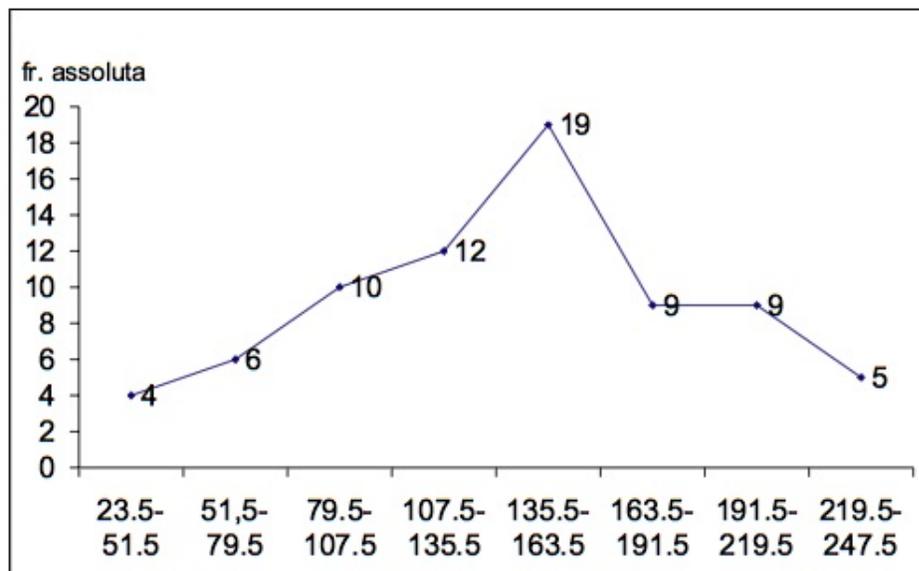
Questi diagrammi sono usati di solito per i fenomeni di tipo qualitativo, nei quali non si possono effettuare misurazioni.

- **Istogrammi:** si divide l'asse delle ascisse in intervalli contigui di ampiezza proporzionale a quella delle corrispondenti classi; su ogni intervallo si riporta un rettangolino di area proporzionale alla frequenza della classe relativa. Si ottiene così un diagramma a scalini. Se si usano le frequenze assolute si parla di *istogramma delle frequenze assolute* e l'area totale dei rettangolini è pari al numero totale di osservazioni. Nel caso delle frequenze relative si parla di *istogramma delle*

frequenze relative e l'area totale dei rettangolini è pari a 1. La figura sottostante mostra l'istogramma delle frequenze assolute riferite, ancora una volta, ai dati dell'esempio 3.1.3.



- **Poligoni di frequenza:** si rappresentano i dati mediante una spezzata. Ogni classe è rappresentata dal suo valore centrale, riportando in corrispondenza un punto di ordinata uguale alla frequenza della classe. Tali punti vengono poi uniti mediante segmenti. In modo analogo agli istogrammi, si possono costruire i poligoni di frequenza (assoluta oppure relativa). Nella figura sottostante, i dati del solito esempio 3.1.3 sono espressi mediante il poligono delle frequenze assolute.



Se gli istogrammi o i poligoni di frequenza si riferiscono alle frequenze cumulate, il diagramma sarà a forma di scalinata crescente o di spezzata crescente.

Si noti che tutte le rappresentazioni dei dati qui illustrate hanno la comune caratteristica di raggruppare i dati in modo più comprensibile, ma di causare in ogni caso una perdita di informazioni. La cura di chi elabora i dati deve essere quella di fare in modo che si perdano soltanto dati inessenziali ai fini della statistica.

3.1.4 Misure descrittive

Allo scopo di presentare in forma chiara e sintetica le principali informazioni rilevabili dai dati, occorre spesso riassumere mediante opportune misure o indici numerici le rilevazioni effettuate. Le misure impiegate più di frequente riguardano principalmente due aspetti:

- misure di posizione (o di tendenza centrale);
- misure di dispersione (o di variazione).

Misure di tendenza centrale

Dato un insieme di dati numerici $Z = \{z_1, \dots, z_r\}$, ne considereremo la *media aritmetica*, la *media geometrica*, la *mediana*, i *quantili* e la *moda*.

Definizione 3.1.4 La *media aritmetica* dei numeri z_1, \dots, z_r è il numero

$$\mathbb{E}[Z] = \bar{z} = \frac{1}{r} \sum_{k=1}^r z_k.$$

Utilizzando i valori effettivamente distinti x_1, \dots, x_n , e le rispettive frequenze relative f_1, \dots, f_n , la media aritmetica \bar{z} può esprimersi come media ponderata nel seguente modo:

$$\bar{z} = \sum_{k=1}^n f_k x_k.$$

È immediato verificare che la media aritmetica è lineare rispetto all'insieme dei dati: in altre parole, se z_1, \dots, z_r e w_1, \dots, w_r sono due insiemi di dati di uguale numerosità, con medie aritmetiche \bar{z} e \bar{w} , e se $a, b \in \mathbb{R}$, allora la media aritmetica dell'insieme di dati

$$u_1 = az_1 + bw_1, \quad u_2 = az_2 + bw_2, \quad \dots, \quad u_r = az_r + bw_r$$

è la quantità

$$\bar{u} = a\bar{z} + b\bar{w}.$$

Definizione 3.1.5 La *media geometrica* dei numeri positivi z_1, \dots, z_r è il numero

$$\bar{z}_g = \left(\prod_{k=1}^r z_k \right)^{\frac{1}{r}}.$$

La media geometrica è la più appropriata in svariate situazioni.

Esempio 3.1.6 In un periodo di 8 anni, una banca ha praticato sui depositi il seguente tasso di interesse composto: 1.1% per i primi 2 anni, 1.9% per i successivi 3 anni, 1.5% per i seguenti 2 anni, e 1.4% l'ultimo anno. Qual è stato il tasso medio annuo? Indicando con r_1, \dots, r_8 i tassi applicati negli 8 anni e con r il tasso medio, si ha

$$r_1 = r_2 = 0,011, \quad r_3 = r_4 = r_5 = 0.019, \quad r_6 = r_7 = 0.015, \quad r_8 = 0.014,$$

e dev'essere

$$(1 + r)^8 = \prod_{i=1}^8 (1 + r_i),$$

e quindi $1 + r$ è la media geometrica dei numeri $1 + r_1, \dots, \dots, 1 + r_8$, cioè

$$1 + r = \left[\prod_{i=1}^8 (1 + r_i) \right]^{\frac{1}{8}} = 1.01537.$$

Pertanto $r = 0.01537$: quindi il tasso medio annuo applicato dalla banca è stato l'1.537%.

Esempio 3.1.7 Dato un parallelepipedo i cui lati misurano rispettivamente 8 cm, 5 cm e 25 cm, calcolare la lunghezza ℓ del lato del cubo avente lo stesso volume. Dev'essere ovviamente $\ell^3 = 8 \cdot 5 \cdot 25 = 1000 \text{ cm}^3$ e quindi $\ell = 10 \text{ cm}$, cioè ℓ è la media geometrica delle misure dei lati del parallelepipedo.

Definizione 3.1.8 La *media armonica* dei numeri positivi z_1, \dots, z_r è il numero

$$\bar{z}_a = \frac{r}{\sum_{i=1}^r \frac{1}{z_i}}.$$

Vediamo un caso in cui si fa uso della media armonica.

Esempio 3.1.9 Un'automobile ha percorso un tratto di strada alla velocità costante di 80 km/h all'andata e di 120 km/h al ritorno. Qual è, ai fini del tempo totale di percorrenza, la velocità media sull'intero percorso?

La media che ci dà il valore esatto è in questo caso la media armonica. Infatti, indicando con s la lunghezza del tratto di strada, i tempi di percorrenza, misurati in ore, all'andata e al ritorno sono rispettivamente

$$t_a = \frac{s}{80}, \quad t_r = \frac{s}{120}.$$

Quindi il tempo totale è $t = t_a + t_r$ e la velocità media sull'intero percorso è

$$v = \frac{2s}{t_a + t_r} = \frac{2s}{\frac{s}{80} + \frac{s}{120}} = \frac{2}{\frac{1}{80} + \frac{1}{120}} = 96 \text{ km/h}.$$

Si dimostra in modo non troppo difficile che quando i dati numerici sono tutti positivi e non tutti coincidenti, risulta

$$\bar{z}_a < \bar{z}_g < \bar{z}.$$

Definizione 3.1.10 La *mediana* di un insieme di numeri z_1, z_2, \dots, z_r , ordinati in ordine crescente oppure decrescente, è il valore centrale se r è dispari, ed è la media aritmetica dei due valori centrali se r è pari.

Dunque, ad esempio, la mediana dei 9 numeri

$$3, \quad 85, \quad 94, \quad 97, \quad 99, \quad 100, \quad 103, \quad 105, \quad 106 \quad (3.1)$$

è uguale a 99, mentre aggiungendo il decimo numero 107 la mediana dei 10 numeri diventa 99.5. Invece la media dei 9 numeri è 88; aggiungendo il decimo numero la media diventa 89.9.

Si osservi che la media aritmetica è fortemente influenzata dai valori estremi (in particolare dalla presenza di valori anomali), mentre la mediana non ne risente: se, ad esempio, sostituiamo il primo numero 3 con 84, la mediana non cambia, mentre la media diventa 96 (con 9 numeri) e 98 (con 10 numeri). Pertanto, la mediana è preferibile nei casi in cui ci sono pochi dati sperimentali, oppure la gran parte dei dati sono concentrati verso un estremo.

Definizione 3.1.11 Dato un insieme di dati numerici z_1, \dots, z_r , ordinati in ordine crescente oppure decrescente, si definisce *quantile di ordine p* quell'unico valore alla sinistra del quale vi è una frazione p del totale dei dati.

Dunque, il quantile di ordine p è quel valore q tale che la frequenza relativa cumulata, calcolata fino a q incluso, sia $\geq p$ (mentre era $< p$ nei valori che precedevano q). Ad esempio, il quantile di ordine $7/9$ dell'insieme di dati (3.1) è il numero 103. Si noti che la mediana è il quantile di ordine $1/2$ (ovvero 50%).

Definizione 3.1.12 Dato un insieme di dati numerici z_1, z_2, \dots, z_r , ordinati in ordine crescente oppure decrescente, i tre *quartili* Q_1 , Q_2 e Q_3 sono i valori che dividono l'insieme ordinato dei dati in quattro parti uguali: cioè, alla sinistra di Q_1 si trova il 25% dei dati, alla sinistra di Q_2 (che coincide con la mediana) sta il 50% dei dati, mentre alla sinistra di Q_3 è situato il 75% dei dati.

In modo analogo si possono definire i *decili* e i *centili* (o *percentili*).

Definizione 3.1.13 Si consideri un insieme z_1, z_2, \dots, z_r di dati di una variabile Z , sia x_1, \dots, x_n l'insieme dei valori effettivamente distinti di Z , e siano r_1, \dots, r_n le rispettive frequenze assolute. Si chiama *moda* dell'insieme z_1, z_2, \dots, z_r ogni valore che compare con frequenza massima, cioè ogni valore x_k per il quale risulti $r_k \geq r_j$, $j = 1, 2, \dots, n$.

Quando i dati sono raggruppati in classi si possono individuare una o più *classi modali*, che corrispondono nell'istogramma ad altrettanti massimi.

La moda può risultare utile quando i dati sono divisi in classi che non sono di tipo

numerico (ad esempio, luogo di nascita, professione, ...). D'altro canto, la moda perde molta della sua utilità nel caso che essa non sia unica. Osserviamo che per le distribuzioni di dati unimodali (ossia per le quali la moda è unica) e simmetriche (ossia le frequenze assolute di x_j e x_{n-j} coincidono per ogni $j \leq n/2$), la media aritmetica, la mediana e la moda coincidono, come è facile verificare.

Misure di dispersione

Le misure di tendenza centrale non ci dicono nulla sul modo in cui i dati di una variabile Z sono distribuiti intorno al valore centrale. Infatti due o più insiemi di dati possono avere uno stesso valore centrale e allo stesso tempo essere distribuiti in modo completamente differente intorno ad esso. Per misurare la dispersione dei dati si introducono degli indici di variabilità. In questo senso il rango o campo di variazione (si veda la definizione all'inizio del paragrafo 3.1.2)) costituisce un primo indice di dispersione: esso, però, diventa poco significativo se uno dei dati è anomalo (cioè molto grande o molto piccolo rispetto agli altri).

Anche la media aritmetica \bar{w} delle deviazioni dalla media

$$w_1 = z_1 - \bar{z}, \quad w_2 = z_2 - \bar{z}, \quad \dots, \quad w_r = z_r - \bar{z}$$

non è utile per misurare la dispersione dei dati, in quanto risulta sempre $\bar{w} = 0$.

Si potrebbe utilizzare la media dei valori assoluti delle deviazioni dalla media, detta *deviazione media* e pari alla quantità

$$\frac{1}{r} \sum_{i=1}^r |z_i - \bar{z}|$$

Tuttavia la deviazione media non è facilmente trattabile dal punto di vista matematico: risulta più conveniente considerare la media dei quadrati delle deviazioni dalla media, che si dice *varianza* della variabile Z ed è definita dalla quantità

$$\mathbf{Var}[Z] = \sigma_Z^2 = \frac{1}{r} \sum_{i=1}^r (z_i - \bar{z})^2.$$

Nel caso particolare $z_1 = z_2 = \dots = z_r = z$ risulta $\bar{z} = z$ e quindi $\sigma_Z^2 = 0$. Al contrario, se per almeno due indici i, j si ha $z_i \neq z_j$, allora $\sigma_Z^2 > 0$. Il numeratore dell'espressione che definisce la varianza si chiama *devianza*: dunque la devianza è il numero

$$r \sigma_Z^2 = \sum_{i=1}^r (z_i - \bar{z})^2.$$

La radice quadrata σ_Z della varianza si chiama *scarto quadratico medio* o *deviazione standard*, e rappresenta anch'essa una misura di dispersione dei dati. A differenza della varianza, però, la deviazione standard è espressa nelle stesse unità di misura dei dati. Il rapporto fra la deviazione standard e il valore assoluto della media si chiama *coefficiente di variabilità* ed è sovente espresso in forma percentuale: esso è utile per confrontare

tra di loro le variabilità di due o più variabili.

Elenchiamo nuovamente le proprietà della varianza, già analizzate nel caso delle variabili aleatorie.

- La varianza è la differenza fra la media dei quadrati e il quadrato della media:

$$\sigma_Z^2 = \overline{z^2} - \bar{z}^2.$$

- La varianza σ_W^2 dell'insieme di dati

$$w_1 = z_1 + a, \quad w_2 = z_2 + a, \quad \dots, \quad w_r = z_r + a,$$

ove $a \in \mathbb{R}$, è uguale alla varianza σ_Z^2 dell'insieme di dati z_1, \dots, z_r .

- La varianza σ_U^2 dell'insieme di dati

$$u_1 = bz_1, \quad u_2 = bz_2, \quad \dots, \quad u_r = bz_r,$$

ove $b \in \mathbb{R}$, è uguale a $b^2\sigma_Z^2$.

- In particolare, indicando con \bar{z} e con σ_Z la media aritmetica e la deviazione standard dell'insieme di dati z_1, z_2, \dots, z_r , la varianza dell'insieme di dati

$$\frac{z_1 - \bar{z}}{\sigma_Z}, \quad \frac{z_2 - \bar{z}}{\sigma_Z}, \quad \dots, \quad \frac{z_r - \bar{z}}{\sigma_Z}$$

è uguale a 1. L'operazione di passaggio dai dati z_i ai dati $\frac{z_i - \bar{z}}{\sigma_Z}$, ossia dalla variabile Z alla *variabile normalizzata* $\frac{Z - \bar{z}}{\sigma_Z}$, si dice *standardizzazione*. Con tale operazione, la media aritmetica dei dati standardizzati è nulla e la varianza è unitaria.

3.1.5 Dati bidimensionali

Un caso importante in statistica è quello in cui ad ogni unità della popolazione in esame sono associate due variabili X, Y (ad esempio, peso e statura, oppure età e reddito, e così via). In questo caso l'insieme dei dati sarà costituito da coppie numeriche

$$(x_1, y_1), \quad (x_2, y_2), \quad \dots, \quad (x_n, y_n).$$

In generale, per tale insieme di dati non esisterà una legge funzionale precisa che lega X ed Y ; tuttavia potrà darsi che, al variare dell'indice i , quando il valore x_i è minore della media aritmetica \bar{x} anche y_i risulti prevalentemente minore di \bar{y} e, viceversa, quando x_i è maggiore di \bar{x} anche y_i tenda ad assumere valori maggiori di \bar{y} . In altri casi potrà presentarsi una tendenza di tipo opposto, nel senso che ai valori x_i maggiori di \bar{x} si associno prevalentemente valori y_i minori di \bar{y} , mentre, all'opposto, ai valori x_i minori di \bar{x} si associno prevalentemente valori y_i maggiori di \bar{y} . O, magari, non si manifesterà nessuna delle due tendenze suddette.

Covarianza

Una misura numerica del modo in cui i valori x_i tendono ad associarsi ai valori y_i è costituita dalla *covarianza* di X, Y , definita da

$$\mathbf{Cov}(X, Y) = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}).$$

Fra le due ipotesi prima formulate, la prima corrisponde a una covarianza positiva, mentre la seconda denota una covarianza negativa. Quando si ha $\mathbf{Cov}(X, Y) = 0$ sono assenti entrambe le tendenze suddette e le variabili X, Y si dicono, come sappiamo dalla probabilità, *non correlate*. Indicando con $\mathbf{Var}(X)$ e $\mathbf{Var}(Y)$ le varianze dei due insiemi di dati corrispondenti alle variabili X, Y , e introdotta la variabile $Z = X + Y$, i cui valori costituiscono l'insieme di dati $z_1 = x_1 + y_1, \dots, z_n = x_n + y_n$, si può verificare che risulta

$$\mathbf{Var}(Z) = \mathbf{Var}(X + Y) = \mathbf{Var}(X) + \mathbf{Var}(Y) + 2 \mathbf{Cov}(X, Y).$$

Analogamente, posto $U = X - Y$, si ha

$$\mathbf{Var}(U) = \mathbf{Var}(X - Y) = \mathbf{Var}(X) + \mathbf{Var}(Y) - 2 \mathbf{Cov}(X, Y).$$

La covarianza soddisfa le seguenti proprietà:

1. $\mathbf{Cov}(X, Y) = \frac{1}{n} \sum_{i=1}^n x_i y_i - \bar{x} \bar{y} = \overline{xy} - \bar{x} \bar{y}$;
2. $\mathbf{Cov}(X, X) = \mathbf{Var}(X) = \frac{1}{n} \sum_{i=1}^n x_i^2 - \bar{x}^2$;
3. $\mathbf{Cov}(aX + b, cY + d) = ac \mathbf{Cov}(X, Y)$ per ogni $a, b, c, d \in \mathbb{R}$.

Coefficiente di correlazione

Dalla terza delle tre proprietà della covarianza sopra scritte si ottiene in particolare

$$\mathbf{Cov}\left(\frac{X}{\sigma_X}, \frac{Y}{\sigma_Y}\right) = \frac{\mathbf{Cov}(X, Y)}{\sigma_X \sigma_Y} =: \rho. \quad (3.2)$$

Il numero ρ definito dalla (3.2) è una *covarianza normalizzata* ed è adimensionale, ossia non dipende dalle unità di misura utilizzate per calcolare i valori di X e Y ; esso si chiama *coefficiente di correlazione* di X e Y .

Posto $X' = aX + b$, $Y' = cY + d$, con $ac > 0$, si può verificare che il coefficiente di correlazione di X', Y' coincide con quello di X, Y . Inoltre, qualunque sia la coppia X, Y , per il coefficiente di correlazione ρ vale la proprietà

$$-1 \leq \rho \leq 1.$$

Infatti, per la disuguaglianza di Cauchy-Schwarz,

$$\begin{aligned} |\mathbf{Cov}(X, Y)| &= \frac{1}{n} \left| \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) \right| \leq \\ &\leq \frac{1}{n} \left[\sum_{i=1}^n (x_i - \bar{x})^2 \right]^{\frac{1}{2}} \left[\sum_{i=1}^n (y_i - \bar{y})^2 \right]^{\frac{1}{2}} = \sigma_X \sigma_Y. \end{aligned}$$

Infine, per il coefficiente di correlazione si può verificare il seguente risultato:

$$|\rho| = 1 \iff Y = aX + b, \quad \text{ove } a, b \in \mathbb{R}. \quad (3.3)$$

Infatti, se $Y = aX + b$ per certi $a, b \in \mathbb{R}$, allora per le proprietà 2 e 3 della covarianza

$$\mathbf{Cov}(X, Y) = a \mathbf{Cov}(X, X) = a \mathbf{Var}(X);$$

inoltre, $\mathbf{Var}(Y) = a^2 \mathbf{Var}(X)$ e pertanto $\sigma_Y = |a|\sigma_X$. Ne segue

$$\rho = \frac{\mathbf{Cov}(X, Y)}{\sigma_X \sigma_Y} = \frac{a\sigma_X^2}{|a|\sigma_X^2} = \frac{a}{|a|} = \begin{cases} 1 & \text{se } a > 0, \\ -1 & \text{se } a < 0. \end{cases}$$

Viceversa, sia $\rho = 1$: allora

$$\mathbf{Var}\left(\frac{X}{\sigma_X} - \frac{Y}{\sigma_Y}\right) = 2(1 - \rho) = 0,$$

per cui tutti i valori della variabile $\frac{X}{\sigma_X} - \frac{Y}{\sigma_Y}$ sono uguali fra loro: perciò esiste $c \in \mathbb{R}$ per cui

$$\frac{x_i}{\sigma_X} - \frac{y_i}{\sigma_Y} = c, \quad i = 1, 2, \dots, n.$$

In altre parole, tutte le coppie (x_i, y_i) appartengono alla retta del piano di equazione

$$\frac{x}{\sigma_X} - \frac{y}{\sigma_Y} = c.$$

Dunque fra le variabili X e Y vi è una relazione lineare della forma $Y = aX + b$, con $a = \frac{\sigma_Y}{\sigma_X}$ e $b = -c\sigma_Y$. Similmente, se $\rho = -1$ si trova

$$\mathbf{Var}\left(\frac{X}{\sigma_X} + \frac{Y}{\sigma_Y}\right) = 2(1 + \rho) = 0,$$

e quindi, analogamente, si ottiene una relazione lineare della forma $Y = aX + b$, con $a = \frac{\sigma_Y}{\sigma_X}$ e $b = c\sigma_Y$.

Come mostrato dalla (3.3), il coefficiente di correlazione esprime una misura della dipendenza lineare che sussiste tra le variabili X e Y . Possiamo disegnare il *grafico di dispersione*, ossia rappresentare nel piano cartesiano tutte le coppie di valori (x_i, y_i) : si nota allora che, quando la “nuvola” costituita dai dati (x_i, y_i) è molto addensata intorno a una retta, il coefficiente di correlazione ρ avrà un valore vicino a $+1$ o -1 , a seconda che il coefficiente angolare della retta sia positivo o negativo. Se invece la nuvola di punti è abbastanza “rotonda”, il valore di ρ sarà vicino a 0. Osserviamo che se tra X ed Y c'è un legame *non lineare*, può benissimo risultare $\rho = 0$. Un esempio molto semplice è rappresentato dal seguente insieme di dati bidimensionali

$$(-2, 4), \quad (-1, 1), \quad (0, 0), \quad (1, 1), \quad (2, 4) :$$

essi soddisfano la relazione $Y = X^2$ e risulta, come si può verificare,

$$\bar{x} = 0, \quad \bar{y} = 2, \quad \sigma_X = \sqrt{2}, \quad \sigma_Y = \sqrt{\frac{14}{5}}, \quad \mathbf{Cov}(X, Y) = \rho = 0.$$

3.1.6 Rette di regressione

In molte applicazioni, tra le variabili X e Y può sussistere un legame lineare $Y = aX + b$, ma, a causa di errori di misura nella rilevazione dei dati, non si possono determinare a e b ; oppure la dipendenza fra X e Y non è esattamente lineare, ma si ritiene che il legame statistico che intercorre tra di esse possa essere approssimato con una opportuna funzione lineare del tipo $Y = aX + b$.

Il metodo che si utilizza per scegliere, tra le infinite rette, quella che meglio approssima la distribuzione di dati bidimensionali risale a Gauss e Legendre ed è noto come metodo dei minimi quadrati.

La logica dietro a tale metodo è la seguente: se i punti (x_i, y_i) appartenessero tutti ad una retta di equazione $y = ax + b$, risulterebbe $(y_i - ax_i - b)^2 = 0$ per ciascun indice i . Se una tale retta non esiste, ci si accontenta di determinare la retta che rende minima la somma dei quadrati, ovvero si determina la coppia (a, b) per la quale risulta minima la quantità

$$f(a, b) = \sum_{i=1}^n (y_i - ax_i - b)^2.$$

Tale retta si chiama *retta di regressione* della variabile Y sulla variabile X . Il modo per trovarla è abbastanza facile: si annulla il sistema delle due derivate parziali

$$\begin{cases} \frac{\partial f}{\partial a}(a, b) = -2 \sum_{i=1}^n x_i (y_i - ax_i - b) = 0 \\ \frac{\partial f}{\partial b}(a, b) = -2 \sum_{i=1}^n (y_i - ax_i - b) = 0, \end{cases}$$

il quale ammette un'unica soluzione. Infatti il sistema si riscrive come

$$\begin{cases} \sum_{i=1}^n x_i y_i - a \sum_{i=1}^n x_i^2 - b \sum_{i=1}^n x_i = 0 \\ \sum_{i=1}^n y_i - a \sum_{i=1}^n x_i - nb = 0, \end{cases}$$

ovvero, con evidente significato dei simboli,

$$\begin{cases} n \overline{xy} - a n \overline{x^2} - nb \overline{x} = 0 \\ b = \overline{y} - a \overline{x}; \end{cases}$$

inserendo il valore di b nella prima equazione si ottiene facilmente l'unica soluzione

$$\begin{cases} a = \frac{\overline{xy} - \overline{x} \overline{y}}{\overline{x^2} - \overline{x}^2} = \frac{\mathbf{Cov}(X, Y)}{\sigma_X^2} = \rho \frac{\sigma_Y}{\sigma_X} \\ b = \overline{y} - \rho \frac{\sigma_Y}{\sigma_X} \overline{x}. \end{cases}$$

A questi valori corrisponde la retta di regressione di equazione

$$y = \bar{y} + \rho \frac{\sigma_Y}{\sigma_X} (x - \bar{x}),$$

che si può anche scrivere nella forma simmetrica

$$\frac{y - \bar{y}}{\sigma_Y} = \rho \frac{x - \bar{x}}{\sigma_X}. \quad (3.4)$$

In modo del tutto analogo, l'equazione della retta di regressione di X su Y è

$$x = \bar{x} + \rho \frac{\sigma_X}{\sigma_Y} (y - \bar{y}),$$

ovvero

$$\frac{x - \bar{x}}{\sigma_X} = \rho \frac{y - \bar{y}}{\sigma_Y}.$$

Osserviamo che le rette di regressione contengono il punto (\bar{x}, \bar{y}) .

Rivediamo il significato del coefficiente di correlazione ρ , calcolando la varianza della differenza tra la variabile normalizzata $\frac{Y - \bar{y}}{\sigma_Y}$ e la variabile $\rho \frac{X - \bar{x}}{\sigma_X}$, differenza che abbiamo stimato sopra tramite la regressione lineare. Si ha

$$\begin{aligned} \mathbf{Var} \left(\frac{Y - \bar{y}}{\sigma_Y} - \rho \frac{X - \bar{x}}{\sigma_X} \right) &= \mathbf{Var} \left(\frac{Y}{\sigma_Y} - \rho \frac{X}{\sigma_X} \right) = \\ &= \mathbf{Var} \left(\frac{Y}{\sigma_Y} \right) + \rho^2 \mathbf{Var} \left(\frac{X}{\sigma_X} \right) - 2\rho \mathbf{Cov} \left(\frac{Y}{\sigma_Y}, \frac{X}{\sigma_X} \right) = \\ &= 1 + \rho^2 - 2\rho^2 = 1 - \rho^2. \end{aligned}$$

Pertanto si ha

$$\mathbf{Var} \left(\frac{Y - \bar{y}}{\sigma_Y} - \rho \frac{X - \bar{x}}{\sigma_X} \right) = 0 \quad \iff \quad \rho = \pm 1.$$

In sostanza, tanto più i valori delle variabili X e Y sono vicini alla retta di regressione, tanto più $|\rho|$ è vicino a 1, ossia tanto più la correlazione fra le due variabili è grande.

Esempio 3.1.14 La tabella sottostante riporta i dati bidimensionali delle variabili X e Y , che rappresentano rispettivamente la statura (in cm), con valori elencati in ordine crescente, e il peso (in kg) dei 28 ragazzi considerati nell'esempio 3.1.2.

statura	peso	statura	peso	statura	peso	statura	peso
X	Y	X	Y	X	Y	X	Y
158.0	45.0	161.8	49.7	165.5	51.5	167.8	53.4
159.0	50.8	162.0	50.5	165.8	51.8	168.0	54.0
159.5	49.0	163.0	51.0	166.0	52.0	170.0	54.8
160.0	49.3	163.4	51.0	166.5	52.3	171.4	55.2
160.7	50.0	163.7	51.5	166.8	53.0	172.6	55.5
161.0	50.0	164.0	51.8	167.0	53.3	173.0	57.2
161.5	50.2	165.0	51.3	167.4	53.5	177.5	56.5

Risultano da questi dati: le medie aritmetiche

$$\bar{x} = 165.3, \quad \bar{y} = 51.97,$$

le deviazioni standard

$$\sigma_X = 4.60, \quad \sigma_Y = 2.52,$$

la covarianza

$$\mathbf{Cov}(X, Y) = 10.81,$$

il coefficiente di correlazione

$$\rho = 0.93,$$

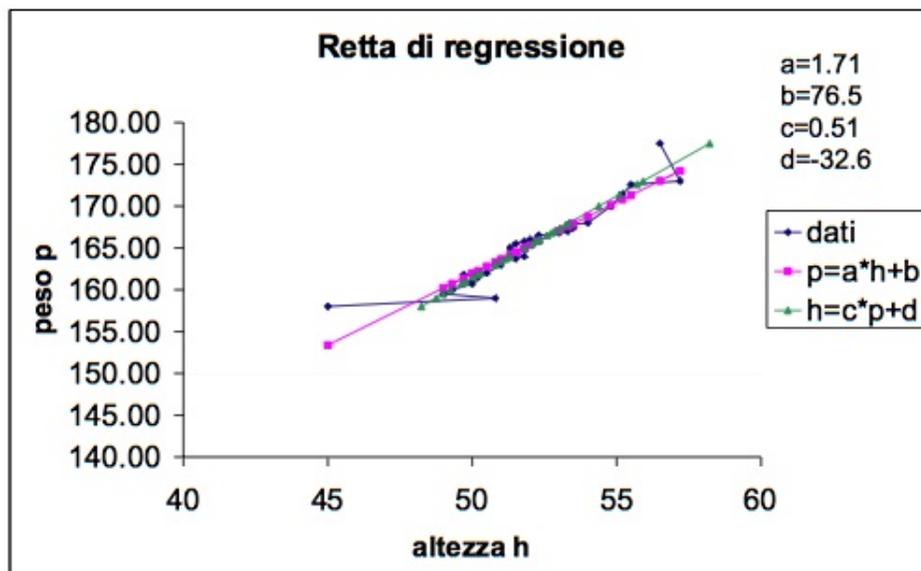
l'equazione della retta di regressione di Y su X

$$y = 0.51x - 32.34,$$

e infine l'equazione della retta di regressione di X su Y

$$x = 1.71y + 76.5.$$

Il valore $\rho = 0.93$, prossimo a 1, indica una forte correlazione lineare tra X e Y , come mostrato dal grafico della figura sottostante.



Esercizi del §3.1

1. Da un collettivo di 20 individui si è rilevata la seguente distribuzione relativa ai caratteri “età”, “sesso”, “numero di automobili possedute”:

unità	età	sexo	n. auto	unità	età	sexo	n. auto
1	35	M	1	11	33	M	2
2	37	M	2	12	46	F	4
3	59	F	1	13	41	F	3
4	54	M	0	14	53	M	1
5	44	F	2	15	38	F	1
6	38	M	1	16	55	M	1
7	62	F	1	17	50	M	3
8	71	F	0	18	63	M	0
9	56	M	3	19	35	F	1
10	60	M	2	20	51	M	2

- Si costruiscano le distribuzioni di frequenza per i caratteri “sexo” e “n. auto”.
- Si consideri il carattere “età” suddiviso nelle classi $[30, 39]$, $[40, 49]$, $[50, 59]$, $[60, 70]$ e $[70, 80]$, e si costruiscano le corrispondenti distribuzioni di frequenza assolute, assolute cumulate, relative e relative cumulate.
- Si rappresentino, mediante i grafici ritenuti più idonei, le distribuzioni di frequenza del sesso, del numero di automobili e dell’età suddivisa in classi.
- Si calcoli il coefficiente di correlazione fra i caratteri “età” e “n. auto”, lo stesso coefficiente limitatamente alla popolazione degli uomini e lo stesso coefficiente limitatamente alla popolazione delle donne.

2. La seguente tabella riporta le votazioni ottenute da una classe di 57 studenti alla fine di un corso universitario:

voto	18	19	20	21	22	23	24	25	26	27	28	29	30	30 L.
n. stud.	7	2	5	1	3	2	12	1	8	4	6	1	4	2

- Calcolare la distribuzione delle frequenze cumulate relative del carattere “voto”.
- Calcolare la distribuzione delle frequenze cumulate relative avendo suddiviso il carattere “voto” nelle seguenti classi: $[18, 22]$, $[23, 24]$, $[25, 26]$, $[27, 28]$, $[29, 30L.]$.
- Disegnare i grafici della distribuzione di frequenza relativa.
- Quanti sono gli studenti che hanno ottenuto un voto inferiore o uguale a 26?
- Quanti sono gli studenti che hanno ottenuto un voto non superiore a 24?

3. Per 6 pazienti sono noti i valori dell’emoglobina registrati prima e dopo una chemioterapia:

prima	13.0	12.8	11.0	13.2	12.5	11.9
dopo	9.4	11.5	11.5	13.1	10.2	12.0

- Si calcoli la riduzione media di emoglobina.
- Che relazione c’è fra questa e le medie dei valori “prima” e “dopo”?

4. In un centro dietologico vengono seguite 8 donne in gravidanza (entro il quarto mese), le quali pesano rispettivamente, in kg: 64.3, 65.2, 70.0, 54.5, 58.8, 81.5, 61.0, 62.0. Qual è la media? Qual è la mediana? I dati suggeriscono una forte asimmetria della distribuzione del peso?
5. Per i primi 15 giorni di un mese viene rilevato il ritardo, espresso in minuti, accumulato da un determinato treno rispetto all'orario previsto di arrivo. Di seguito sono riportati i dati rilevati, che presentano segno negativo nel caso di anticipo sull'orario di arrivo:

giorno	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
ritardo	10	15	-2	0	50	20	0	9	-5	8	10	20	6	10	-10

- Calcolare la moda, la mediana, la media aritmetica, la varianza e lo scarto quadratico medio della distribuzione dei ritardi.
 - Dopo aver raggruppato le osservazioni relative ai primi 15 giorni del mese nelle classi $[-10, 0]$, $]0, 10]$, $]10, 60]$, calcolare media e varianza in questa nuova situazione.
6. Viene rilevato il risparmio medio annuo, espresso in migliaia di euro, di 8 famiglie:

famiglia	A	B	C	D	E	F	G	H
risparmio	0.5	5	2.6	0	9.2	3	5.4	6.3

- Determinare la media aritmetica, la varianza e la mediana dei dati.
 - Considerando per i valori della variabile “risparmio” l'intervallo $[0, 10]$, sintetizzare la serie osservata in una distribuzione con quattro classi di frequenza, tutte della stessa ampiezza e con estremo superiore incluso, e determinare i valori assunti in questo caso dalla media e dalla varianza delle 8 osservazioni.
 - Supponendo che si rendano disponibili le informazioni per altre 2 famiglie, rispetto alle quali il carattere “risparmio” presenta media pari a 5 e devianza pari a 2, determinare la media aritmetica e la devianza del carattere per il complesso delle $8 + 2 = 10$ famiglie.
7. La distribuzione di 40 individui secondo il numero di battiti cardiaci al minuto (variabile X) è la seguente:

intervalli	44-54	55-58	59-62	62-66
freq. assoluta	8	10	14	8

- Determinare media aritmetica e varianza della distribuzione.
 - Sapendo che tra i 40 individui vi sono 10 sportivi e che per questi si registrano mediamente 51 battiti al minuto, con varianza pari a 16.1, determinare media aritmetica e varianza del carattere X per i rimanenti 30 individui.
8. Ai 1000 abitanti di un piccolo comune viene chiesto di esprimere un giudizio su un nuovo servizio comunale, usando una scala da 0 a 4 (0 = pessimo, 4 = ottimo). Le risposte ottenute sono riassunte nella tabella che segue.

giudizio	0	1	2	3	4	5
freq. assoluta	251	260	80	0	154	255

- Rappresentare con un opportuno grafico le risposte.
 - Calcolare media aritmetica, mediana, moda, varianza e deviazione standard dei dati.
9. Vengono intervistati 36 pisani, a cui viene chiesto il numero di vani presente nella propria abitazione. Le 36 risposte ottenute sono le seguenti:

1, 3, 4, 2, 2, 4, 5, 5, 1, 1, 2, 3, 4, 3, 2, 6, 6, 1, 2, 2, 3, 2, 1, 3, 4, 2, 3, 3, 3, 5, 6, 4, 2, 2, 4, 2.

- Calcolare le frequenze assolute, le frequenze relative e le frequenze cumulate relative delle risposte ottenute, rappresentandole graficamente in modo opportuno.
 - Determinare media aritmetica, moda, mediana e varianza delle risposte.
 - La stessa indagine è stata svolta a Livorno, e le risposte fornite da 36 livornesi hanno dato un valor medio pari a 2.5 ed una varianza pari a 3.6. Calcolare il coefficiente di variabilità del numero di vani nelle abitazioni delle due città.
10. Per 300 giorni vengono rilevati i consumi complessivi di energia elettrica presso un piccolo comune montano. I dati ottenuti, espressi in KW, vengono riassunti nella tabella che segue.

classe (KW)	[0, 100)	[100, 200)	[200, 400)	[400, 600)	[600, 1000)
freq. assoluta (giorni)	50	85	65	55	45

- Rappresentare graficamente i consumi osservati.
 - Calcolare la media, la mediana e la deviazione standard dei consumi.
11. Durante l'inverno sono state intervistate 30 persone, a cui è stato chiesto quante volte si sono recate al cinema nell'ultimo mese. Le 30 risposte ottenute sono le seguenti:

1, 0, 4, 2, 2, 4, 5, 0, 1, 1, 2, 3, 4, 3, 2, 2, 2, 3, 0, 1, 3, 4, 0, 0, 3, 3, 5, 6, 4, 2.

- Determinare le frequenze relative e le frequenze cumulate relative delle risposte ottenute.
 - Fornire un rappresentazione grafica delle risposte.
 - Determinare media aritmetica, moda, mediana e varianza delle risposte.
 - La stessa indagine è stata svolta d'estate, e le risposte fornite dalle 30 persone hanno dato un valor medio uguale a 2.5 ed una varianza uguale a 3. Calcolare il coefficiente di variabilità delle due situazioni.
12. Consideriamo le "importazioni" e le "esportazioni" avvenute in un certo anno, espresse in miliardi di dollari, dei paesi partecipanti all'Organizzazione per la Cooperazione e lo Sviluppo Economico (OCSE):

Paese	import	export	Paese	import	export
Austria	26.7	22.4	Irlanda	11.6	12.6
Belgio	68.5	68.6	Islanda	1.1	1.1
Canada	81.3	86.7	Italia	100.0	97.5
Corea del Sud	211.5	218.7	Norvegia	20.3	18.2
Danimarca	22.8	21.2	Olanda	75.4	80.6
Finlandia	15.3	16.3	Portogallo	9.4	7.2
Francia	128.8	119.3	Spagna	34.9	27.1
Germania	189.7	242.4	Svezia	32.5	37.2
Giappone	127.7	210.8	Svizzera	40.9	37.3
Gran Bretagna	126.2	107.0	Turchia	11.1	7.4
Grecia	11.3	5.6	USA	370.0	210.8

- Si costruiscano le distribuzioni di frequenza per i caratteri “import” ed “export”.
- Si rappresentino, mediante i grafici ritenuti più idonei, le distribuzioni di frequenza dei due caratteri.
- Tracciare il grafico di dispersione delle due variabili.
- Costruire la retta di regressione del carattere “import” rispetto al carattere “export”.

13. Consideriamo il peso, in kg, e l’altezza, in cm, di 10 individui:

peso	56	66	84	61	73	90	70	61	75	82
altezza	161	165	186	162	172	191	181	164	179	184

- Costruire il grafico di dispersione per i due caratteri.
- Determinare la retta di regressione che pone il carattere “altezza” in funzione del carattere “peso”.

14. In un’indagine statistica è stato chiesto a 30 madri, occupate come libere professioniste, di indicare il “n. di figli” e il “n. di ore di lavoro casalingo” svolto giornalmente:

madre	figli	ore lavoro casalingo	madre	figli	ore lavoro casalingo
1	1	1	16	2	4
2	1	2	17	2	5
3	1	3	18	3	5
4	1	5	19	3	4
5	2	3	20	3	4
6	2	1	21	4	5
7	3	5	22	5	5
8	3	1	23	5	5
9	4	6	24	4	2
10	4	3	25	4	2
11	5	7	26	2	4
12	5	4	27	1	5
13	1	4	28	2	1
14	1	2	29	3	2
15	1	1	30	6	5

- Costruire il grafico di dispersione per i due caratteri.
- Calcolare la covarianza ed il coefficiente di correlazione dei due caratteri.
- Determinare la retta di regressione che considera il “n. di ore di lavoro casalingo” in funzione del “n. di figli”.

15. Una varietà di frumento è stata saggiata in sei appezzamenti, per verificarne la produttività. Le produzioni ottenute (in tonnellate per ettaro, t/ha) sono state: 6.5, 5.7, 6.4, 6.3, 6.2, 5.8. Valutare media aritmetica, devianza, varianza, deviazione standard e coefficiente di variabilità.

16. Sono state rilevate le altezze di 3000 piante di mais. I dati, suddivisi in classi di frequenza relativi ad intervalli di 10 cm, sono i seguenti:

classi	freq. ass.	classi	freq. ass.
[145, 150[25	[170, 175[594
[150, 155[90	[175, 180[494
[155, 160[224	[180, 185[374
[160, 165[399	[185, 190[176
[165, 170[547	[190, 195[77

- Calcolare le frequenze assolute cumulate, relative e relative cumulate, descrivendole graficamente in modo opportuno.
- Calcolare media aritmetica, mediana e moda dei dati.
- In quale percentile si trovano due individui alti rispettivamente 160 e 190 cm?

17. Un campo di mais è concimato con tre dosi crescenti di azoto, pari a 0, 150 e 300 kg/ha. Le produzioni osservate sono rispettivamente pari a 5, 9 e 12 t/ha. Stabilire la relazione esistente tra dose di concimazione e produzione, il coefficiente di correlazione e l'equazione della retta di regressione.

18. La seguente serie di dati riguarda una casistica di 10 soggetti adulti maschi: si considerano l'età, il valore della FEV1 (Forced Espiratory Volume in 1 secondo) e la pressione diastolica.

età	25	32	28	21	33	33	34	24	41	26
FEV1	2.5	1.8	1.5	2.5	4.5	2.1	3.4	1.2	2.8	3.9
pressione	85	71	92	80	87	83	70	101	90	83

- Calcolare media aritmetica e deviazione standard dei tre caratteri.
- Calcolare il coefficiente di correlazione fra i caratteri “FEV1” e “pressione”.
- Stabilire qual è il carattere più variabile, attraverso il calcolo del coefficiente di variabilità.

Tavole numeriche

Nelle pagine successive compaiono tre tavole numeriche: descriviamo brevemente come si usano. La prima tavola è la tavola della funzione di ripartizione della legge $\mathcal{N}(0, 1)$: essa fornisce, per i quantili $x = \phi_\alpha$ con $0 \leq x \leq 3.2$, il valore della funzione

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt.$$

Ad esempio, $\Phi(1.43) = 0.9236$. Per valori di x superiori a 3.2, si pone $\Phi(x) = 1$; invece per valori negativi di x si utilizza l'uguaglianza $\Phi(-x) = 1 - \Phi(x)$.

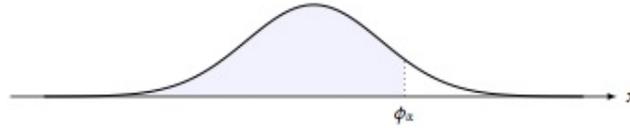
Per ottenere, noto α , il valore del quantile ϕ_α della funzione Φ , cioè il numero x per cui risulta $\Phi(x) = \alpha$, si usa la tavola della funzione di ripartizione a rovescio, ossia si cerca il valore α nella tavola e si ricava il valore di x per il quale si ha $\Phi(x) = \alpha$. Ad esempio, $\phi_{0.95} = 1.65$. In questo modo però si possono ottenere solo i valori di ϕ_α con $\alpha \geq \frac{1}{2}$. Per valori di α inferiori a $\frac{1}{2}$ si usa l'uguaglianza $\phi_\alpha = -\phi_{1-\alpha}$, che è una immediata conseguenza dell'uguaglianza $\Phi(-x) = 1 - \Phi(x)$.

La seconda tavola è la tavola dei quantili della legge di Student $t(n)$: questa tavola permette di trovare direttamente, in funzione del numero n dei gradi di libertà e del numero α con $0 < \alpha < 1$, il valore del quantile $t_\alpha(n)$. Per la legge di Student è stata progettata questa tavola poiché questi sono i quantili che vengono utilizzati nella ricerca degli intervalli di fiducia o della regione critica del test di Student, mentre il valore della funzione di ripartizione è molto meno importante per le applicazioni pratiche. Notiamo tuttavia che questa tavola contiene i quantili per valori di α vicini a 1 (mentre in alcuni test servono i quantili di valori di α vicini a 0): tuttavia, questi valori si recuperano tenendo conto dell'uguaglianza

$$t_\alpha(n) = -t_{1-\alpha}(n).$$

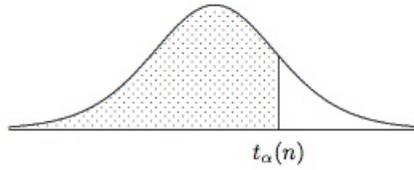
Infine, la terza tavola contiene i quantili $\chi_\alpha^2(n)$ della legge del chi-quadro $\chi^2(n)$ in funzione del numero n dei gradi di libertà e del numero α con $0 < \alpha < 1$: il suo uso è pertanto simile a quello della tavola dei quantili per la legge di Student. C'è però una differenza sostanziale: la densità non è in questo caso una funzione pari, anzi è addirittura nulla per valori negativi di x . Di conseguenza non si possono ricavare i quantili con $\alpha \leq \frac{1}{2}$ da quelli con $\alpha \geq \frac{1}{2}$. Per questo motivo la tavola riporta i quantili $\chi_\alpha^2(n)$ per α vicino a 1 e vicino a 0.

La funzione di ripartizione della legge $\mathcal{N}(0, 1)$



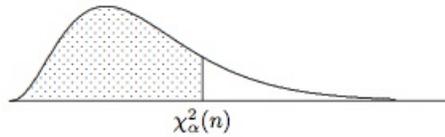
ϕ_x	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998

I quantili delle leggi $t(n)$ di Student



n	$\alpha = 0.95$	$\alpha = 0.975$	$\alpha = 0.99$	$\alpha = 0.995$
1	6.31375	12.7062	31.8206	63.6570
2	2.91999	4.3027	6.6946	9.9248
3	2.35336	3.1824	4.5407	5.8409
4	2.13187	2.7764	3.7470	4.6041
5	2.01505	2.5706	3.3649	4.0322
6	1.94318	2.4469	3.1427	3.7075
7	1.89459	2.3646	3.9980	3.4995
8	1.85955	2.3060	2.8965	3.3554
9	1.83311	2.2622	2.8214	3.2499
10	1.81246	2.2281	2.7638	3.1693
11	1.79589	2.2010	2.7181	3.1058
12	1.78299	2.1788	2.6810	3.0546
13	1.77093	2.1604	2.6503	3.0123
14	1.76131	2.1448	2.6245	2.9769
15	1.75305	2.1315	2.6025	2.9467
16	1.74589	2.1109	2.5835	2.9208
17	1.73961	2.1098	2.5669	2.8982
18	1.73407	2.1009	2.5524	2.8784
19	1.72914	2.0930	2.5395	2.8610
20	1.72473	2.0860	2.5280	2.8453
21	1.72075	2.0796	2.5176	2.8314
22	1.71715	2.0739	2.5083	2.8188
23	1.71388	2.0687	2.4999	2.8073
24	1.71089	2.0639	2.4922	2.7969
25	1.70814	2.0595	2.4851	2.7874
26	1.70562	2.0555	2.4786	2.7787
27	1.70331	2.0518	2.4727	2.7707
28	1.70113	2.0484	2.4671	2.7633
29	1.69914	2.0452	2.4620	2.7564
30	1.69726	2.0423	2.4573	2.7500
40	1.68385	2.0211	2.4233	2.7045
60	1.67065	2.0003	2.3902	2.6604
80	1.66413	1.9901	2.3739	2.6387
120	1.65765	1.9799	2.3578	1.6174
∞	1.64485	1.95996	2.32635	2.57583

I quantili delle leggi $\chi^2(n)$



n	$\alpha = 0.01$	$\alpha = 0.025$	$\alpha = 0.05$	$\alpha = 0.95$	$\alpha = 0.975$	$\alpha = 0.99$
1	0.0002	0.0010	0.0039	3.841	5.024	6.635
2	0.0201	0.0506	0.1026	5.991	7.378	9.210
3	0.1148	0.2158	0.3518	7.815	9.348	11.345
4	0.2971	0.4844	0.7107	9.488	11.143	13.277
5	0.5543	0.8312	1.1455	11.070	12.833	15.086
6	0.8721	1.2373	1.6354	12.592	14.449	16.812
7	1.2390	1.6899	2.1674	14.067	16.013	18.475
8	1.6465	2.1797	2.7326	15.507	17.535	20.090
9	2.0879	2.7004	3.3251	16.919	19.023	21.666
10	2.5582	3.2470	3.9403	18.307	20.483	23.209
11	3.0535	3.8157	4.5748	19.675	21.920	24.725
12	3.5706	4.4038	5.2260	21.026	23.337	26.217
13	4.1069	5.0088	5.8919	22.362	24.736	27.688
14	4.6604	5.6287	6.5706	23.685	26.119	29.141
15	5.2293	6.2621	7.2609	24.996	27.488	30.578
16	5.8122	6.9077	7.9616	26.296	28.845	32.000
17	6.4078	7.5642	8.6718	27.587	30.191	33.409
18	7.0149	8.2307	9.3905	28.869	31.526	34.805
19	7.6327	8.9065	10.1170	30.143	32.852	36.191
20	8.2604	9.5908	10.8508	31.410	34.170	37.566
21	8.8972	10.2829	11.5913	32.671	35.479	38.932
22	9.5425	10.9823	12.3380	33.924	36.781	40.290
23	10.1957	11.6886	13.0905	35.172	38.076	41.638
24	10.8564	12.4012	13.8484	36.415	39.364	42.980
25	11.5240	13.1197	14.6114	37.653	40.647	44.314
26	12.1981	13.8439	15.3792	38.885	41.923	45.642
27	12.8785	14.5734	16.1514	40.113	43.195	46.963
28	13.5647	15.3079	16.9279	41.337	44.461	48.278
29	14.2565	16.0471	17.7084	42.557	45.722	49.588
30	14.9535	16.7908	18.4927	43.773	46.979	50.892

Per valori più grandi di n si usa il fatto che, se X_n è una variabile aleatoria dotata di legge $\chi^2(n)$, allora la variabile aleatoria $\sqrt{2X_n} - \sqrt{2n-1}$ ha approssimativamente legge $\mathcal{N}(0, 1)$. Ovvero: $\chi_{\alpha}^2(n) \approx \frac{1}{2} (\phi_{\alpha} + \sqrt{2n-1})^2$.

Indice analitico

- additività
 - finita, 6
 - numerabile, 6
- alternativa, 120
- approssimazione normale, 85, 87, 122
- assenza di memoria, 40, 66, 67
- attributi di una popolazione, 135
- boreliano, 32, 52, 58
- calcolo
 - combinatorio, 13
 - delle probabilità, 32, 35, 36, 68, 77, 78, 84
- campione, 86, 92, 96, 98, 99, 101, 106, 129
 - gaussiano, 111
 - normale, 100
 - statistico, 95, 100, 107, 109, 111, 129
 - estratto da una popolazione, 95, 98, 108
- campo di variazione, 136, 145
- caratteri di una popolazione, 135
- cardinalità, 13
- casi
 - favorevoli, 11
 - possibili, 1, 11
- centile, 144
- classe
 - di dati, 138
 - modale, 144
- coefficiente
 - di correlazione, 64, 75, 147, 148, 150, 152, 156
 - di variabilità, 145, 154, 156
- combinazione semplice, 16
- componenti di un vettore aleatorio, 52
- convergenza
 - in probabilità, 70, 71
 - puntuale, 50
- correlazione fra variabili, 150
- correzione di continuità, 86, 87
- corse dei cavalli, 1
- costo
 - di uno stimatore, 97
 - medio, 97
 - quadratico, 97
- covarianza, 62, 64, 147, 156
 - negativa, 147
 - normalizzata, 147
 - positiva, 147
- criterio fondamentale
 - per la coincidenza di due misure, 33
- curva di fiducia, 108, 109
- dati bidimensionali, 146
- decile, 144
- densità, 10, 46, 48, 54, 58, 61
 - binomiale, 87
 - condizionale, 59
 - congiunta, 54, 59, 101
 - continua, 101, 106
 - della legge normale ridotta, 82
 - di probabilità, 68, 114
 - di una variabile aleatoria, 46, 58
 - discreta, 10, 36, 52, 101, 106, 125
 - congiunta, 52, 73
 - di probabilità, 10, 36
 - di una legge, 36
 - marginale, 52
 - marginale, 101
 - marginali, 54, 74
- devianza, 145, 153, 156
- deviazione
 - dalla media, 145
 - media, 145
 - standard, 61, 64, 70, 112, 145, 154, 156
 - campionaria, 99
 - empirica, 99
- diagramma
 - a scalini, 140
 - a torta, 139
 - di Venn, 5

- dipendenza
 - lineare, 148
 - non lineare, 149
- disposizione
 - con ripetizione, 14
 - semplice, 13
- distribuzione
 - di dati, 149
 - di frequenza, 152, 155
 - di una variabile aleatoria, 33
- disuguaglianza
 - di Chebyshev, 69, 70
 - di Markov, 69, 70
- effettivo
 - empirico, 125
 - teorico, 125
- elaborazione dei dati, 136
- equiprobabilità, 11
- errore
 - di prima specie, 120
 - di seconda specie, 120
- esperimento aleatorio, 1, 11, 35, 38, 51
- estrazione di una carta da un mazzo, 3
- estrazioni del lotto, 1
- eventi
 - incompatibili, 5, 58
 - indipendenti, 23, 24
- evento, 3, 4, 108
 - contrario, 4
 - quasi certo, 6
 - trascurabile, 6
- eventualità, 1, 4
- famiglia di probabilità, 94
- fattoriale, 14
- fiducia, 108, 109
- formula
 - della disintegrazione, 20, 58
 - di Bayes, 21
- frequenza
 - assoluta, 136, 154
 - cumulata, 137
 - assoluta, 137, 156
 - relativa, 137, 154, 156
 - dei successi, 71
 - di classe, 138
 - assoluta, 138
 - relativa, 138
 - relativa, 136, 154, 156
- funzione
 - Γ , 67
 - d'errore, 81
 - di massima verosimiglianza, 104, 105
 - di ripartizione, 47, 48, 65, 66, 81, 84, 126
 - di verosimiglianza, 101
 - indicatrice, 43, 64
 - logaritmicamente convessa, 67
- gradi di libertà, 79, 80
- grafico a barre, 139
- grafico di dispersione, 148, 155, 156
- indicatrice, 43
- indice
 - di dispersione, 145
 - di variabilità, 145
- indipendenza
 - di eventi, 23, 24
 - di variabili aleatorie, 35
- insieme
 - aleatorio, 108, 109
 - boreliano, 32, 52, 58
 - delle eventualità, 4
 - delle parti di Ω , 4
 - di fiducia, 95, 109
- intervallo
 - di fiducia, 108, 111
 - bilatero, 109
 - unilatero, 109
- ipotesi, 120, 129
 - statistica, 120
- isotonia
 - della probabilità, 9
 - della speranza, 44
- istante di primo successo, 40
- istogramma, 140, 144
 - delle frequenze assolute, 140
 - delle frequenze relative, 141
- lancio
 - di due monete, 24
 - di una moneta, 2
 - di un dado, 1, 11
- legge, 128, 129
 - binomiale, 37, 38, 41, 45, 57, 60, 63, 87

condizionale, 57, 58
 dei grandi numeri, 69, 70, 80
 del chi-quadro, 79, 81, 83, 100, 124, 129
 della media empirica, 98
 della varianza empirica, 100
 di Bernoulli, 37, 45, 62, 71
 di Poisson, 41, 42, 46, 60, 63, 103
 di Student, 80, 81, 83, 100
 di un campione, 97, 108
 di un campione statistico, 95
 di un vettore aleatorio, 52
 di una statistica, 95
 di una variabile aleatoria, 33, 35, 72
 diffusa, 48, 49
 esponenziale, 47, 55, 65, 67, 69, 104, 107
 gamma, 68, 69, 107
 geometrica, 40, 46, 63, 104
 normale, 77, 78, 81, 83, 85, 93, 100, 105
 ridotta, 77, 79, 81, 98
 uniforme, 36, 45, 62, 72, 75
 livello
 di fiducia, 109
 di precisione relativa, 111
 di un insieme di fiducia, 109–114
 di un test, 39, 121–125
 lotus, 44, 50, 55, 56
 media, 43, 51, 61, 69, 153
 aritmetica, 142, 145, 146, 153, 154, 156
 armonica, 143
 campionaria, 98
 di una popolazione, 98, 122
 di una variabile aleatoria, 50, 70
 empirica, 98, 103, 106, 122
 geometrica, 142
 ponderata, 20, 43, 58, 142
 mediana, 144, 153, 154, 156
 metodo
 d'inferenza, 95
 dei minimi quadrati, 149
 dei momenti, 107, 114
 misura
 di dispersione, 142, 145
 di posizione, 142
 di probabilità, 6
 definita da una densità, 10
 di tendenza centrale, 142
 di variazione, 142
 moda, 144, 153, 154, 156
 modalità, 135
 modello statistico, 94–96, 98, 101, 104, 108,
 109, 111, 120
 campionario, 95
 non parametrico, 94
 parametrico, 94
 modularità, 9
 momento
 del prim'ordine, 51, 61
 del second'ordine, 61, 65, 66, 68, 99
 di ordine r , 51, 106
 empirico
 di ordine r , 106
 negazione di un evento, 4
 operatore di speranza, 49
 parametri
 di una legge
 binomiale, 37
 parametro, 37, 40, 42, 87
 di una legge
 binomiale, 76, 87
 di Bernoulli, 37, 40, 63, 71
 di Poisson, 41, 63
 esponenziale, 47, 55, 65, 67, 69
 geometrica, 40
 sconosciuto, 93–96, 98, 101, 103, 104, 108,
 109, 111, 119, 120, 128
 parte
 negativa, 50
 positiva, 50
 percentile, 144, 156
 permutazione, 2, 15
 con ripetizione, 15
 semplice, 15
 poligono di frequenza, 141
 popolazione, 92, 135
 potenza di un test, 120
 precisione relativa, 111
 probabilità
 a posteriori, 20
 condizionale, 19, 20, 23, 57, 58, 66
 di successo, 37
 di un evento, 6, 43
 quantile, 84, 113

- di ordine $1/2$, 144
 - di ordine α , 83, 84
 - di ordine p , 144
- quartile, 144
- raggruppamento, 13
- rango, 136, 145
 - di un insieme di dati, 136
- rappresentazione
 - dei dati, 136
 - grafica, 139
- regione
 - critica, 120–125
 - d'accettazione, 120
- regola
 - a priori, 96
 - d'inferenza, 92
 - di decisione, 96
- regressione lineare, 150
- relazione
 - lineare, 149
 - fra variabili, 148
 - non lineare
 - fra variabili, 148
- retta di regressione, 149, 150, 155, 156
- rilevazione dei dati, 135, 136
- ripartizione uniforme, 11, 12, 24, 33, 37, 51, 53, 56, 60, 64, 89
- rischio
 - di uno stimatore, 97
 - quadratico, 97, 98
- scarto quadratico medio, 61, 145, 153
- schema delle prove indipendenti, 36
- singoletto, 11
- sistema in parallelo, 25
- soglia critica, 108
- sottotribù, 57, 58
- spazio
 - probabilizzabile, 4, 94
 - discreto, 10
 - probabilizzato, 6, 33, 35, 49, 51, 77, 84
- speranza, 43, 51, 61, 62, 66, 68, 72, 79
 - di una variabile aleatoria, 43, 50
 - discreta, 43
 - reale, 49
- standardizzazione, 78, 81, 146
- statistica, 94
 - descrittiva, 92
 - di Pearson, 125
 - di Rayleigh, 116
 - inferenziale, 92
 - integrabile, 94
- statistiche, 108
 - indipendenti, 94, 95
- stima
 - di massima verosimiglianza, 102
 - di un parametro, 96
 - insiemistica, 95
 - puntuale, 95
- stimatore, 95, 96, 101, 111
 - corretto, 97–99, 103, 106, 107, 122, 123, 128
 - dei momenti, 107
 - di massima verosimiglianza, 102–105, 129
 - distorto, 97, 106
 - non distorto, 97
 - ottimale, 97
 - preferibile ad un altro, 97
- stimatori indipendenti, 100
- strategia d'azione, 96
- subadditività, 26
- suddivisione in classi, 138
- svuotamento di un'urna, 2
- taglia, 95, 96, 98–100, 109, 111
 - di un campione, 95, 96, 101, 106, 108
 - statistico, 95
- tempo residuo, 67
- teorema
 - del binomio, 18
 - di Cochran, 112, 122, 124
 - di Pearson, 125
 - limite centrale, 85, 93, 98
- test, 96
 - d'ipotesi, 96
 - del chi-quadro, 125, 127, 128
 - di Fisher–Snedecor, 123
 - bilatero, 125
 - unilatero, 125
 - di Student, 122
 - bilatero, 123
 - unilatero, 123
 - statistico, 120
- testa o croce, 2
- tribù, 4

boreliana, 5, 32, 52
 degli eventi, 4, 24, 53
 delle parti di Ω , 10, 11, 24, 33, 53
 generata da una variabile aleatoria, 32, 59

unità statistica, 135

valor medio, 61, 70

valore

- atteso, 51
- di aspettazione, 51

variabile, 135, 145, 147

- aleatoria, 32, 33, 37, 41, 42, 44, 51, 60, 65, 67, 71, 77, 94, 104
- bernoulliana, 37, 40, 42, 64
- binomiale, 85
- centrata, 61, 77
- continua, 46, 50, 57, 58
- di quadrato integrabile, 61
- diffusa, 47, 83
- discreta, 36, 40, 43, 44, 51, 57, 61, 70
- dotata di densità, 46, 48, 49, 51, 58, 61, 68
- esponenziale, 47
- gaussiana, 77, 81
- integrabile, 43, 50, 61, 98
- normale, 80
- numerica, 33
- positiva, 49, 50, 66
- reale, 46, 48, 49, 70, 83
- standardizzata, 78
- strettamente positiva, 69
- uniformemente ripartita, 51, 64, 65

normalizzata, 146, 150

variabili, 146, 147

- aleatorie
 - bernoulliane, 93
 - equivalenti, 44
 - indipendenti, 35, 38–40, 54, 56, 63, 77, 84, 85, 93
 - non correlate, 62, 63
- non correlate, 147

varianza, 61, 66, 68, 69, 72, 79, 145, 146, 150, 153, 154, 156

- campionaria, 99
- di una popolazione, 98
- di una variabile aleatoria, 61
- empirica, 99, 106, 112, 123

vettore aleatorio, 51

- continuo, 54
- discreto, 52
- dotato di densità, 54