

Calcolo delle probabilità e statistica elementare

Dispense per il Corso di Laurea in Scienze Ambientali

Andrea Carpignani (A.A. 2006/2007)

Mathematiker sind wie Franzosen: Wenn du ihnen etwas sagst, übersetzen sie es gleich in ihre eigene Sprache und sofort bedeutet es etwas ganz anderes.

(J.W. Goethe)

INDICE

INTRODUZIONE	ii
1. GLI SPAZI PROBABILIZZATI	
1.1 Esperimenti aleatori, ripetibilità	1
1.2 La tribù degli eventi, misurazione del grado di fiducia	2
1.3 Proprietà generali degli spazi probabilizzati	4
1.4 Misure di probabilità su uno spazio discreto	5
1.5 Probabilità condizionale, formula di Bayes, indipendenza	6
Esercizi	9
2. LE VARIABILI ALEATORIE	
2.1 Definizione di variabile aleatoria, legge, indipendenza	12
2.2 Variabili aleatorie discrete	13
2.3 La speranza di una variabile aleatoria discreta	17
2.4 Il concetto generale di speranza, variabili aleatorie definite mediante densità	18
2.5 Varianza e covarianza di una variabile aleatoria, la legge dei grandi numeri	20
Esercizi	22
3. IL TEOREMA LIMITE CENTRALE	
3.1 Le leggi normali, del chi–quadro e di Student	24
3.2 Funzioni di ripartizione	26
3.3 Il teorema limite centrale	27
Esercizi	29
4. LA STATISTICA INFERENZIALE	
4.1 Introduzione	31
4.2 Il problema della stima	32
4.3 Stima della media e della varianza per campioni gaussiani	35
4.4 Test d’ipotesi statistiche	37
4.5 Il test di Student	38
4.6 Il test di Fisher–Snedecor	40
4.7 Il test del chi–quadro	41
Esercizi	46
TAVOLE NUMERICHE	
I quantili della legge $\mathcal{N}(0, 1)$	51
I quantili delle leggi $t(n)$ di Student	52
I quantili delle leggi $\chi^2(n)$	53

INTRODUZIONE

Una disputa tra giocatori d'azzardo, avvenuta nel 1654, ha dato vita al calcolo delle probabilità, ad opera di due matematici francesi: Blaise Pascal (1623-1662) e Pierre de Fermat (1601-1665). L'interesse di Pascal per la probabilità fu risvegliato dal Cavalier de Méré: spirito vivace, matematico almeno discreto (e, al tempo stesso, accanito giocatore d'azzardo). Egli si lamentò che la matematica lo faceva perdere al gioco, perché aveva calcolato per una combinazione ai dadi una probabilità maggiore di $1/2$, aveva scommesso a lungo su tale combinazione, ma invece di vincere perdeva. Un altro problema posto dal Cavalier de Méré fu quello della ripartizione della posta. Due giocatori impegnati in una partita devono interromperla; tenendo conto del punteggio raggiunto, come va ripartita la posta? Su questi problemi si sviluppò un'intensa corrispondenza tra Pascal e Fermat, nella quale affiorarono, per la prima volta, i principi fondamentali del calcolo delle probabilità. Sebbene qualche piccolo problema sui giochi d'azzardo fosse stato risolto da Gerolamo Cardano nel suo trattato *De ludo aleae*, e anche da Galileo Galilei, nessuna teoria generale della probabilità era mai stata scritta prima della famosa corrispondenza.

Lo scienziato olandese Christian Huygens, insegnante di Leibniz, apprese di questa corrispondenza e poco dopo (nel 1657) pubblicò il primo libro di probabilità, intitolato *Ratiociniis in ludo aleae*; si trattava di un trattato sui problemi associati al gioco d'azzardo. Proprio grazie a questo suo legame coi giochi d'azzardo e con le scommesse, il calcolo delle probabilità divenne ben presto molto popolare e si sviluppò rapidamente durante tutto il XVII secolo. I maggiori contributi, in quel periodo, furono dati da Jakob Bernoulli (1654-1705) e da Abraham de Moivre (1667-1754).

Nel 1812 Pierre de Laplace (1749-1827) introdusse molte nuove idee e tecniche matematiche nel suo libro *Théorie analytique des probabilités*. Prima di Laplace, il calcolo delle probabilità era concentrato soltanto sullo sviluppo di una sorta di analisi matematica dei giochi d'azzardo. Laplace invece impiegò le idee probabilistiche in molti problemi di matematica applicata. La teoria degli errori, la statistica matematica e la meccanica statistica sono solo alcuni esempi di applicazioni della probabilità sviluppatasi nel XIX secolo.

Come spesso accade in matematica, lo sviluppo del calcolo delle probabilità è stato stimolato dalle sue applicazioni e, reciprocamente, ne ha allargato gli orizzonti. La statistica matematica, ad esempio, è una parte molto importante della probabilità applicata; altre applicazioni sono state trovate nell'ambito di discipline quali la genetica, l'economia, la recentissima finanza matematica. Molti autori hanno studiato la teoria della probabilità a partire dal tempo di Laplace: i maggiori contributi sono stati dati, senz'alcun dubbio, da Cebiscev, Markov, von Mises e Kolmogorov.

Una delle maggiori difficoltà, nello sviluppo del calcolo delle probabilità, è stata il fatto di arrivare ad una definizione del concetto di probabilità che fosse sufficientemente precisa da un punto di vista matematico, e, allo stesso tempo, abbastanza duttile per essere applicata ad una vasta classe di fenomeni. La ricerca di una definizione accettabile è durata circa tre secoli ed è stata molto controversa. Il problema è stato risolto, nel XX secolo, dal matematico russo A. Kolmogorov il quale introdusse, in una monografia pubblicata nel 1933, uno schema assiomatico nel quale inquadrare il calcolo delle probabilità, che permise di inserire la probabilità all'interno di una teoria generale, conosciuta sotto il nome di "teoria della misura". Lo sviluppo del calcolo delle probabilità all'interno di questa disciplina ha poi influenzato enormemente quest'ultima arricchendola di nuove idee e tecniche prima sconosciute.

1. Gli spazi probabilizzati

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 fissare un "ventaglio completo di eventualità, a due a due incompatibili", ossia un insieme Ω , i cui elementi rappresentino ipotetici risultati dell'esperimento, con la certezza che, comunque vadano le cose, il risultato effettivo dell'esperimento "cadrà in Ω " (nel senso che sarà rappresentato da uno ed un sol elemento di Ω).

Esempio 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 $\{1, 2, 3, 4, 5, 6\}$.

Esempio 1.2 (*Estrazioni del lotto*) Si supponga che l'esperimento sia costituito dalle estrazioni del lotto che verranno eseguite, la settimana prossima, sulla ruota di Napoli. 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 cinque, 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.

Giova osservare che, nella scelta dell'insieme Ω , c'è sempre una certa dose di *arbitrarietà*. Non bisogna infatti dimenticare che gli elementi di Ω *rappresentano* ipotetici risultati dell'esperimento, secondo un opportuno *codice*: 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 prendere $\Omega = \{0, 1\}$, 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.

La parola *esperimento* potrebbe far credere che per "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 1.3 (*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. Per conoscerlo abbia a disposizione 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 gl'interi compresi tra 1 e $n - 1$ (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 chiaro 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 qualche errore di conteggio).

Tornando al caso generale di un arbitrario esperimento aleatorio, al quale sia stato associato un certo insieme Ω di eventualità, consideriamo ora una qualsiasi parte A di Ω . Si può interpretare A come rappresentazione di un *evento legato al risultato dell'esperimento*: l'evento che si realizza se e solo se tale risultato “cadrà in A ”. (Si può anzi *identificare* questo evento con l'insieme A stesso.)

Esempio 1.4 (*Uscita di una faccia pari*) Nel caso del lancio di un dado (Esempio 1.1), la parte $\{2, 4, 6\}$ dell'insieme $\Omega = \{1, 2, 3, 4, 5, 6\}$ si può interpretare come rappresentante dell'evento indicato dalle parole: “uscita di una faccia pari”.

1.2 La tribù degli eventi, misurazione 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 . In ogni caso sarà però opportuno scegliere \mathcal{A} in modo tale che essa possieda buone doti di stabilità (rispetto alle comuni operazioni insiemistiche). Precisamente, sarà conveniente esigere che il complementare (rispetto a Ω) di un qualsiasi elemento di \mathcal{A} sia ancora un elemento di \mathcal{A} , e che l'unione di una qualsiasi famiglia numerabile di elementi di \mathcal{A} sia ancora un elemento di \mathcal{A} . (Come conseguenza, anche l'intersezione di una qualsiasi famiglia numerabile di elementi di \mathcal{A} è ancora un elemento di \mathcal{A} .)

Una classe \mathcal{A} con queste proprietà si chiama, in termini tecnici, una *tribù* (o σ -*algebra*) su Ω . Inoltre, se \mathcal{A} è una tribù su Ω , la coppia (Ω, \mathcal{A}) si chiama uno *spazio probabilizzabile* (e, nell'ambito di un fissato spazio probabilizzabile (Ω, \mathcal{A}) , l'insieme Ω si chiama l'*insieme delle eventualità*, mentre la tribù \mathcal{A} si chiama la *tribù degli eventi*). Usando questo linguaggio, le considerazioni precedenti si possono così riassumere:

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 uno spazio probabilizzabile (Ω, \mathcal{A}) , si adopera abitualmente una terminologia particolarmente suggestiva: 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 A , si chiama la *negazione* 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* dei due eventi A, B si chiama 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 si chiama 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*, ossia privi di elementi in comune.

Esempio 1.5 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 cadrà 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 $[0, 1]$.

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 per 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 l'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 nella tribù \mathcal{A} . Per convenzione, si può prendere questa funzione a valori in $[0, 1]$, e assumere il valore 1 sull'evento Ω . Sarà anche naturale pretendere che essa sia **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, converrà esigere che essa verifichi la relazione

$$P\left(\bigcup_n A_n\right) = \sum_n P(A_n)$$

per ogni successione $A_0, A_1, \dots, A_n, \dots$ di elementi di \mathcal{A} a due a due incompatibili. (Quest'ultima condizione, chiamata **additività numerabile**, è automaticamente verificata quando la tribù \mathcal{A} sia finita e dunque, in particolare, quando Ω sia finito.) Una funzione P con queste proprietà si chiama una **misura di probabilità** sullo spazio probabilizzabile (Ω, \mathcal{A}) (o, semplicemente, sulla tribù \mathcal{A}). Inoltre 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, è forse utile sottolineare 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 tribù \mathcal{A} , e ciascuna di queste può assegnare ad un evento A una diversa probabilità.

Usando il linguaggio sopra introdotto, possiamo così ulteriormente riassumere le considerazioni precedenti:

Compito preliminare, per un individuo che intenda studiare dal punto di vista probabilistico un esperimento aleatorio, è quello di associargli un opportuno 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 un po' diversa. Costui parte di volta in volta da uno specifico problema legato a un determinato esperimento aleatorio. Egli ha dunque davanti a sé i seguenti compiti preliminari:

- (a) *Fissare un adeguato insieme delle eventualità Ω .*
- (b) *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.*
- (c) *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 (per 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 molto complicati).

Abbiamo già richiamato l'attenzione sul carattere inevitabilmente arbitrario della scelta dello spazio probabilizzabile (Ω, \mathcal{A}) . Ancor meno scontata e automatica è la scelta di P . Quale sarà il modo migliore per compierla? Osserviamo, a questo proposito, che la misura P , da mettere sulla tribù \mathcal{A} , è un oggetto matematico col quale si vuole “fotografare” la distribuzione della propria fiducia tra i diversi eventi legati al risultato dell'esperimento considerato. Converrà dunque scegliere P in modo che questa fotografia risulti “il più fedele possibile”. Ma sarebbe arduo precisare che cosa ciò significhi dal punto di vista matematico.

Tuttavia, nella scelta di P , possono essere di grande aiuto i risultati teorici ottenuti dal calcolo delle probabilità nello studio del Problema 1. Questi risultati forniscono infatti, tra l'altro, 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, più o meno “naturali”.

Resta ad ogni modo il fatto 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: o, perlomeno, non ha lo stesso senso che chiedersi se siano giusti o sbagliati determinati calcoli eseguiti nell'ambito di un particolare modello scelto.

1.3 Proprietà generali degli spazi probabilizzati

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, la relazione (1.1) ha alcune conseguenze importanti. Per esempio, per $B = \Omega$, la (1.1) si può riscrivere nella forma $P(A) + P(A^c) = P(\Omega) = 1$, dalla quale si deduce l’importantissima relazione:

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

Questa relazione permette dunque di calcolare la probabilità della negazione di A , conoscendo la probabilità di A . Invece, quando 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 reale non negativo, se ne deduce $P(B) \geq P(A)$. Questa importante proprietà si chiama anche l’*isotonia* della probabilità. Da questa segue, in 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 la relazione $P(B \setminus A) = P(B) - P(A)$.

Sia adesso $A_0, A_1, \dots, A_n, \dots$ una successione di eventi. Dalla ben nota relazione di De Morgan

$$\left(\bigcup_n A_n \right)^c = \bigcap_n A_n^c,$$

e da (1.2), si trae:

$$P\left(\bigcup_n A_n \right) = 1 - P\left(\bigcap_n A_n^c \right).$$

Questa relazione 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.

Le proprietà viste fino a questo momento sono conseguenze della sola proprietà di additività della probabilità. Le due che seguono si dimostrano invece a partire dalla proprietà di additività numerabile.

- Se $A_0, A_1, \dots, A_n, \dots$ è una successione crescente di eventi (cioè se, per ciascun indice n , si ha $A_n \subset A_{n+1}$), e se A denota la riunione degli A_n , si ha $\lim_n P(A_n) = P(A)$.
- Se $A_0, A_1, \dots, A_n, \dots$ è successione decrescente di eventi (cioè se, per ciascun indice n , si ha $A_{n+1} \subset A_n$), e se A denota l’intersezione degli A_n , si ha $\lim_n P(A_n) = P(A)$.

1.4 Misure di probabilità su uno spazio discreto

Consideriamo uno spazio probabilizzabile (Ω, \mathcal{A}) tale che l’insieme Ω sia numerabile, e che la tribù \mathcal{A} sia costituita da tutte le parti di Ω . Un tale spazio si chiama *discreto*. Per costruire su di esso una misura di probabilità si può così procedere. Si scelga una qualsiasi funzione positiva f , definita su Ω 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

$$P(A) = \sum_{\omega \in A} f(\omega). \quad (1.4)$$

È immediato verificare che P è una misura di probabilità. La chiameremo la misura di probabilità **definita dalla densità discreta** f . Inversamente, assegnata una qualsiasi misura di probabilità P su (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 :

$$f(\omega) = P(\{\omega\}) \quad \text{per ogni elemento } \omega \text{ di } \Omega. \quad (1.5)$$

La chiameremo la **densità discreta di** P . Dunque: *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 (1.4) e (1.5)).*

Supponiamo, in particolare, che l'insieme Ω sia *finito*, e precisamente che sia costituito da n elementi. In questo caso, tra le varie densità discrete di probabilità su Ω , ne esiste una e una sola che sia costante: precisamente quella che ha come valore costante $1/n$. La corrispondente misura di probabilità si chiama la **ripartizione uniforme** su Ω . Il suo valore $P(A)$ sulla generica parte A di Ω coincide con il rapporto tra il numero di elementi di A e il numero totale di elementi di Ω :

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

In modo equivalente, si può dire che *la ripartizione uniforme sull'insieme finito Ω è l'unica misura di probabilità (nella tribù di tutte le parti di Ω) che attribuisca la stessa probabilità a tutti i **singoletti**, cioè gli eventi costituiti da un sol elemento*. Sarà questa la misura di probabilità da scegliere sullo spazio discreto (Ω, \mathcal{A}) qualora si giudichi sensato, per ragioni di simmetria suggerite dalla particolare natura del problema, trattare in modo imparziale i diversi singoletti.

Esempio 1.6 Riprendiamo l'esempio del lancio di un dado (Esempio 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, ciascuna delle facce sia *equiprobabile*.

1.5 Probabilità condizionale, formula di Bayes, indipendenza

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} :

$$P_H(A) = \frac{P(A \cap H)}{P(H)}. \quad (1.6)$$

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).

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").

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 riconosce allora immediatamente che gli eventi della forma $A \cap H$, con $H \in \mathcal{H}$, sono a due a due incompatibili, e si ha:

$$P(A) = \sum_{H \in \mathcal{H}} P(A \cap H). \quad (1.7)$$

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:

$$P(A) = \sum_{H \in \mathcal{H}} P(H)P(A|H). \quad (1.8)$$

Questa formula è detta talvolta **formula della disintegrazione**. 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 $P(H)$.

Esempio 1.7 Una popolazione è composta al 40% da fumatori e per il 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?

Per risolvere il problema, 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 rappresenti l'evento che si realizza se e soltanto se si è scelto un individuo fumatore, il secondo dei quali rappresenti l'evento che si realizza se e soltanto se si è scelto un individuo non fumatore e il terzo dei quali che rappresenti 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 incompatibili e la loro riunione coincide con Ω . È quindi possibile calcolare la probabilità di A , utilizzando la formula della disintegrazione di A rispetto ad $\mathcal{H} = \{H, K\}$. Si ha così:

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

Dalla formula della disintegrazione si ricava immediatamente, per ogni evento non trascurabile A e ogni elemento di K di \mathcal{H} , la seguente **formula di Bayes**:

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

Il contenuto intuitivo di quest'ultima eguaglianza è evidente: se (Ω, \mathcal{A}, P) denota lo spazio probabilizzato che un certo individuo ha deciso di associare ad un esperimento aleatorio, e se \mathcal{H} denota l'insieme di tutte le possibili cause che spiegano il realizzarsi dell'esperimento, l'informazione "A si è realizzato" permette di affermare che uno (ed uno soltanto) degli eventi

appartenenti ad \mathcal{H} dev'essersi realizzato, cioè una delle cause deve aver agito. Una volta osservato che l'evento A si è realizzato, l'individuo si può allora domandare quale sia la probabilità che una particolare causa K del risultato dell'esperimento si sia realizzata; la risposta a questo problema viene data appunto dalla formula di Bayes.

Esempio 1.8 Riprendiamo l'Esempio 1.7 e calcoliamo la probabilità che una persona affetta dalla malattia sia un fumatore. Con le notazioni già introdotte, basterà calcolare la probabilità $P(H | A)$. Utilizziamo a questo scopo la formula di Bayes:

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

Esempio 1.9 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?

Allo scopo di 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 l' i -esimo cassetto; 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:

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

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.

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

$$P(A \cap H) = P(A)P(H). \tag{1.9}$$

Notiamo che 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

$$P(A | H) = P(A). \tag{1.10}$$

Il contenuto intuitivo di quest'ultima eguaglianza è evidente: 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 questo: 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).

Esempio 1.10 (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.

(a) 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 dia come risultato il numero ω_1 e il secondo lancio dia come risultato il numero ω_2 . Come *tribù degli eventi* si potrà tranquillamente scegliere la tribù di tutte le parti di Ω .

(b) Come *misura di probabilità* da mettere sulla tribù degli eventi (cioè sull'insieme delle parti di Ω) sarà naturale scegliere la ripartizione uniforme P . 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:

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

Nello stesso modo, si riconosce 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

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

Esercizi

1.1 Da un mazzo composto da quaranta carte se ne estraggono otto in blocco.

- a) Costruire uno spazio probabilizzato (Ω, \mathcal{A}, P) adeguato a descrivere questa situazione.
- b) Calcolare la probabilità che escano cinque assi e tre figure.
- c) Calcolare la probabilità che escano tutte carte rosse.

1.2 Da un'urna, contenente cinquanta palline, numerate da 1 a 50, se ne estraggono in sequenza dieci, rimettendo ogni volta la pallina nell'urna.

- a) Costruire uno spazio probabilizzato (Ω, \mathcal{A}, P) adeguato a descrivere questa situazione.
- b) Calcolare la probabilità che escano tre palline con un numero inferiore o eguale a 45.
- c) Calcolare la probabilità che escano al più tre palline con un numero minore di 16.

1.3 Da un'urna, contenente quattro palline bianche e tre nere, si eseguono due estrazioni, rimettendo ogni volta la pallina nell'urna.

- a) Calcolare la probabilità che le due palline estratte siano del medesimo colore.
- b) Calcolare la probabilità che almeno una delle due palline estratte sia nera.

1.4 Da un'urna contenente cento palline, numerate da 1 a 100, se ne estraggono in sequenza quindici, rimettendo ogni volta la pallina nell'urna. Calcolare la probabilità che escano esattamente cinque palline con un numero superiore a 75.

1.5 Si lanciano contemporaneamente quattro monete equilibrate.

- a) Calcolare la probabilità che escano tre teste.
- b) Calcolare la probabilità che escano al più tre teste.

1.6 Da un mazzo composto da quaranta carte, se ne estraggono in sequenza dieci, rimettendo ogni volta la carta nel mazzo.

- a) Calcolare la probabilità che esattamente tre delle carte uscite siano figure.
- b) Calcolare la probabilità che escano tutte le carte di picche.
- c) Calcolare la probabilità che escano cinque carte rosse e cinque carte nere.

1.7 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.

1.8 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.

1.9 Quattro individui si danno appuntamento al Grand Hotel di Parigi. Ci sono però cinque alberghi che portano questo nome. Se le quattro presone 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?

1.10 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?

1.11 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.

1.12 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; altrimenti, cioè se è uscita croce, si estrae una sola pallina.

- a) Calcolare la probabilità che sia uscita esattamente una pallina rossa.
- b) Sapendo che alla prima estrazione è uscita una pallina rossa, qual è la probabilità che non ci sia una seconda estrazione?

1.13 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.

1.14 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 m palline dello stesso colore. Si riestrae di seguito una pallina. Calcolare la probabilità che la prima pallina estratta sia rossa, sapendo che tale è la seconda.

1.15 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.

1.16 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?

1.17 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:

- 1) **sì**, se esce il numero 1 oppure il numero 2;
- 2) **no**, se esce il numero 3 oppure il numero 4;
- 3) **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.

1.18 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.

2. Le variabili aleatorie

2.1 Definizione di variabile aleatoria, legge, indipendenza

Un individuo che compia un esperimento aleatorio è spesso interessato a studiare delle quantità che sono “funzioni” del risultato dell’esperimento; anzi, si può dire 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 protagoniste, sotto il nome di “variabili aleatorie”.

Su uno spazio probabilizzato (Ω, \mathcal{A}, P) , si chiama una **variabile aleatoria** ogni funzione X di Ω in \mathbb{R} che rispetti la condizione seguente: la tribù \mathcal{A} degli eventi sia abbastanza ricca da contenere tutti gli insiemi della forma

$$\{\omega \in \Omega : X(\omega) \in A\} \quad (2.1)$$

per ogni intervallo A di \mathbb{R} . Ora, se chiamiamo tribù **boreliana** di \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 (2.1), con A intervallo, essa contiene anche tutti gli insiemi della forma (2.1), con A boreliano di \mathbb{R} .

Se X è una siffatta funzione, allora, per ogni insieme boreliano A di \mathbb{R} , l’evento (2.1) (che è realizzato da tutte e sole le eventualità ω di Ω tali che $X(\omega)$ appartenga all’intervallo A) si indica brevemente con $\{X \in A\}$ (da leggere: “ X cade in A ”). Inoltre, la probabilità di un siffatto evento, anziché con $P(\{X \in A\})$, si denota più semplicemente con $P\{X \in A\}$ e si chiama anche la “probabilità che la variabile aleatoria X cada in A ”.

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 .

Notiamo che la legge di X secondo P altri non è che una misura di probabilità sulla tribù boreliana di \mathbb{R} . È importante ricordare che, *se le leggi di due variabili aleatorie coincidono su ogni intervallo di \mathbb{R} , esse sono identiche*. Da questo fatto molto importante 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} .

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

Esempio 2.1 Supponiamo di scommettere sul risultato del lancio di un dado con la regola di guadagnare una moneta se esce 6 e di pagare una moneta per qualsiasi altro risultato. Siamo dunque in presenza di una variabile aleatoria X , definita sullo spazio probabilizzato (Ω, \mathcal{A}, P) costruito nell’Esempio 1.6, nel modo seguente:

$$X(\omega) = \begin{cases} 1 & \text{se } \omega = 6, \\ -1 & \text{se } \omega \neq 6. \end{cases}$$

La legge della variabile aleatoria X appena costruita si calcola facilmente perché la variabile aleatoria assume solo i valori -1 e 1 . Per questo, se I è un qualsiasi intervallo della retta reale contenente il numero 1 , ma non il numero -1 , risulta $\{X \in I\} = \{6\}$ e dunque $P\{X \in I\} = 1/6$.

D'altra parte, se I è un intervallo contenente il numero -1 , ma non il numero 1 , risulta $\{X \in I\} = \{1, 2, 3, 4, 5\}$ e dunque $P\{X \in I\} = 5/6$. Invece, se I è un intervallo che non contiene né il numero 1 né il numero -1 , risulta $\{X \in I\} = \emptyset$ e dunque $P\{X \in I\} = 0$; mentre, se I contiene sia il numero 1 che il numero -1 , si ha $\{X \in I\} = \Omega$ e dunque $P\{X \in I\} = 1$.

Definizione 2.1 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} .

La relazione precedente ammette una generalizzazione al caso di un numero qualsiasi (finito oppure infinito) di variabili aleatorie. Precisamente:

Definizione 2.2 Su uno spazio probabilizzato (Ω, \mathcal{A}, P) , siano X_1, X_2, \dots, X_m un certo numero di variabili aleatorie. Esse si dicono tra loro *indipendenti* se, comunque si scelgano gli intervalli I_1, I_2, \dots, I_m , risulta

$$P\{X_1 \in I_1, X_2 \in I_2, \dots, X_m \in I_m\} = P\{X_1 \in I_1\}P\{X_2 \in I_2\} \cdots P\{X_m \in I_m\}.$$

Analogamente, un numero infinito $X_1, X_2, \dots, X_n, \dots$ di variabili aleatorie sono tra loro *indipendenti*, se per ciascun intero positivo m , lo sono le variabili aleatorie X_1, X_2, \dots, X_m .

Esempio 2.2 Riprendiamo l'Esempio 1.10 e denotiamo con X e con Y le applicazioni che ad ogni coppia $\omega = (\omega_1, \omega_2)$ 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 , esse saranno completamente determinate dagli 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 uno spazio probabilizzato (Ω, \mathcal{A}, P) , di cui spesso ignoreremo la natura, sul quale sono definite delle variabili aleatorie con certe leggi assegnate.

2.2 Variabili aleatorie 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.) In questo caso, la legge di X è individuata, non appena sia determinata la probabilità cui X assume il valore x (con $x \in E$), ovvero non appena si conoscano i numeri

$$f(x) = P\{X = x\}.$$

La funzione f , di E in $[0, 1]$ è una densità discreta di probabilità sulla tribù delle parti di E .

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 un opportuno spazio probabilizzato, che prenda due soli valori: il valore 1 con probabilità p e il valore 0 con probabilità $1 - p$. Questa legge si indica con $\mathcal{B}(1, p)$ e una variabile aleatoria dotata di questa legge si chiama anche una variabile aleatoria **bernoulliana** di parametro p .

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 un opportuno spazio probabilizzato, che prenda i valori $0, 1, 2, \dots, n$ con le probabilità

$$P\{X = k\} = \binom{n}{k} p^k (1 - p)^{n-k} \quad k = 0, 1, 2, \dots, n \quad (2.2)$$

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.

Le figure seguenti 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 densità sono tanto più asimmetriche quanto più p è vicino ai valori estremi 0 e 1.

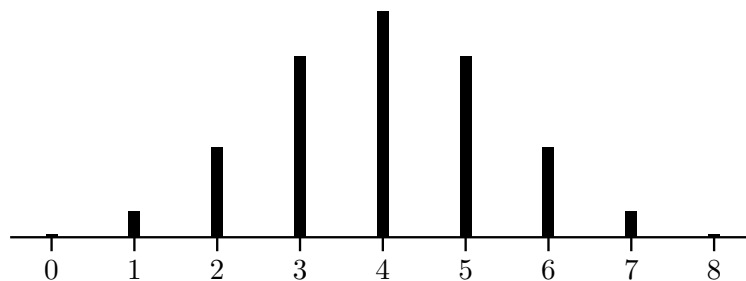


Figura 7.1 Andamento della densità $\mathcal{B}(8, 0.5)$. C'è una simmetria intorno al valore centrale $k = 4$.

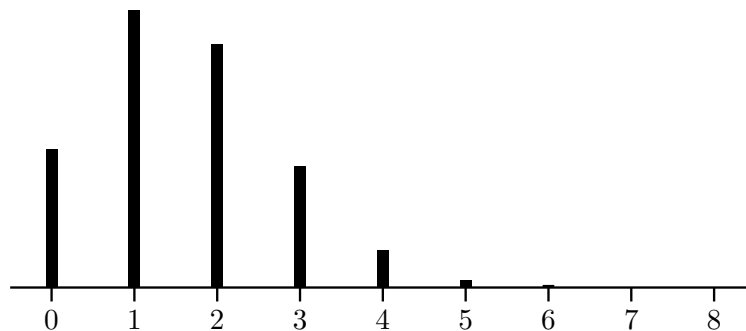


Figura 7.2 Andamento della densità $\mathcal{B}(8, 0.2)$: i valori 7 e 8 vengono assunti con probabilità prossima a 0.

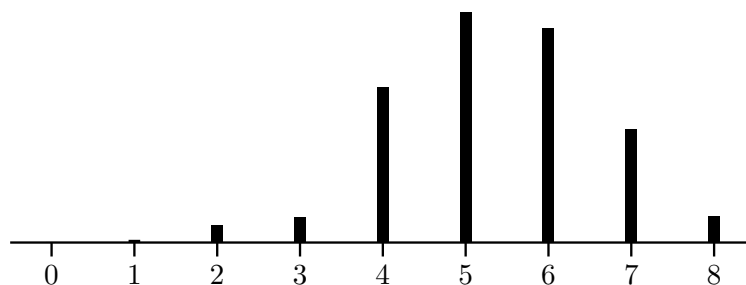


Figura 7.3 Andamento della densità $\mathcal{B}(8, 0.65)$. Si può vedere che con l'aumentare di p aumenta la probabilità di osservare valori grandi e diminuisce quella di osservare valori piccoli.

Consideriamo un esperimento aleatorio costituito da n prove ripetute e indipendenti (come ad esempio lanci successivi di una stessa moneta) in ciascuna delle quali sono possibili due risultati, che indicheremo convenzionalmente con 0 e con 1. 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 dato dalla (2.2). 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, ragioniamo nel modo seguente. Su uno spazio probabilizzato (Ω, \mathcal{A}, P) , siano X, Y due variabili aleatorie 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, si può dimostrare che è $\binom{n-1}{k} + \binom{n-1}{k-1} = \binom{n}{k}$ e dunque

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

cioè Z ha legge $\mathcal{B}(n, p)$. Ciò stabilito, 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 2.3 Un problema frequente, che si riconduce alle leggi binomiali, è il seguente: si considera 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, ...). Supponiamo che la percentuale di individui di tipo A all'interno della popolazione sia p . Da una siffatta popolazione, si scelgono n individui e si vede quanti di essi sono di tipo A . Poniamo $X_k = 1$ 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 opportuno 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 $1-p$ e dunque abbiano legge di Bernoulli $\mathcal{B}(1, p)$. 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.

Un'altra legge naturale in molte situazioni è la cosiddetta “legge di Poisson”. Si chiama la **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 gli 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 $\mathcal{P}(np)$, ovvero

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

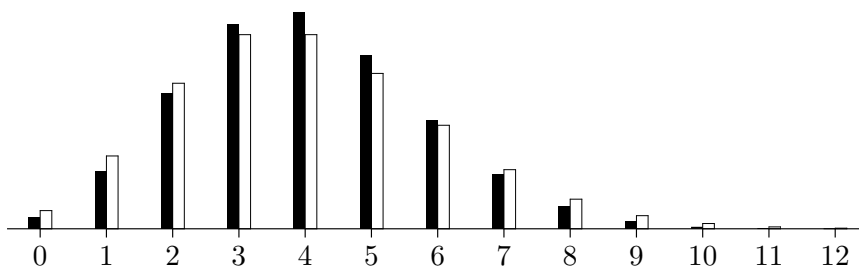


Figura 7.4 Confronto tra una legge binomiale $\mathcal{B}(20, 0.2)$ (corrispondente alle sbarre nere) ed una di Poisson di parametro $\lambda = 20 \cdot 0.2 = 4$ (corrispondente alle sbarre bianche).

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 $n \rightarrow \infty$,

$$\left(1 - \frac{\lambda}{n}\right)^{-k} \rightarrow 1,$$

$$\left(1 - \frac{\lambda}{n}\right)^n \rightarrow e^{-\lambda},$$

$$\frac{n(n-1)\cdots(n-k+1)}{n^k} \rightarrow 1.$$

(Ricordiamo che, per $n \rightarrow \infty$, il limite del rapporto di due polinomi aventi lo stesso grado è pari al rapporto dei coefficienti di grado massimo, qui entrambi eguali a 1.)

Esempio 2.4 Si vuole studiare la diffusione di un infestante nel terreno. A questo scopo, si suddivide un'area prescelta in n parcelle e quindi si passa a censire le piante di infestante presenti in ogni parcella. Per ogni indice i , indichiamo con X_i il numero di piante presenti nella parcella i -esima e vediamo cosa si possa dire della legge di X_i .

Se le piante sono N , in totale, e se supponiamo che ciascuna di esse "scelga" a caso e indipendentemente dalle altre piante una delle n parcelle, con legge uniforme, allora essa si troverà nella i -esima parcella con probabilità $1/n$. Se ora poniamo:

$$Y_k = \begin{cases} 1 & \text{se la } k\text{-esima pianta sceglie la } i\text{-esima parcella,} \\ 0 & \text{altrimenti,} \end{cases}$$

le variabili aleatorie Y_k prendono tutte il valore 1 con probabilità $1/n$ ed hanno dunque legge di Bernoulli di parametro $1/n$. Poiché il numero totale di piante nella i -esima parcella è rappresentato dalla variabile aleatoria $X_i = Y_1 + \dots + Y_N$, e le variabili aleatorie Y_k sono tra loro indipendenti, X_i ha legge binomiale $\mathcal{B}(N, 1/n)$. Nelle condizioni normali, per questo tipo di esperimenti, sia N che n sono abbastanza grandi; si può dunque dire che il numero di piante d'infestante in una singola parcella segue una legge approssimativamente di Poisson di parametro $\lambda = N/n$.

Naturalmente l'ipotesi che l'installazione di ogni singola pianta nelle parcelle dia luogo ad eventi indipendenti è, appunto, un'ipotesi: potrebbe succedere, infatti, che le piante tendano a concentrarsi oppure a mantenere una certa distanza tra l'una e l'altra. In questo caso l'ipotesi d'indipendenza cadrebbe e dunque il modello poissoniano appena sviluppato sarebbe inadeguato.

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.

2.3 La speranza di una variabile aleatoria discreta

Data, su uno spazio probabilizzato (Ω, \mathcal{A}, P) , una variabile aleatoria discreta, si dice che essa è *integrabile* se il numero

$$\sum_{x \in E} |x| P\{X = x\}$$

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

$$\mathbf{E}[X] = \sum_{x \in E} x P\{X = x\}. \quad (2.3)$$

La notazione $\mathbf{E}[X]$ (in cui la lettera \mathbf{E} ricorda le parole *espérance*, *Erwartungswert*, *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 . Nei casi in cui ciò potrebbe creare delle ambiguità, si ricorre alle notazioni $\mathbf{E}_P[X]$, $\mathbf{E}^P[X]$ o a notazioni consimili.

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. Essa è dunque la *media ponderata* dei valori x assunti: il generico valore x interviene nella media col peso $P\{X = x\}$.

Assegnato un evento A (elemento di \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

$$\mathbf{E}[c I_A] = cP(A),$$

per ogni evento A e ogni numero reale c . Evidentemente, poi, una 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, sono isonome, dunque hanno la stessa speranza.

Sussiste, 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, 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 da

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

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

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

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

Su uno spazio probabilizzato (Ω, \mathcal{A}, P) , sia X una variabile aleatoria discreta, a valori nell'insieme finito E , e sia g una funzione di E in \mathbb{R} . Allora, affinché la variabile aleatoria $g(X)$ sia integrabile occorre e basta che il numero $\sum_{x \in E} |g(x)| P\{X = x\}$ sia finito, e se questa condizione è soddisfatta, si ha:

$$\mathbf{E}[g(X)] = \sum_{x \in E} g(x) P\{X = x\}.$$

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 "generali" della speranza, che vedremo nel prossimo paragrafo):

$$\mathbf{E}[X] = \sum_{n \geq 0} P\{X > n\}.$$

Infine, se X, Y sono due variabili aleatorie discrete integrabili e indipendenti, il loro prodotto è ancora una variabile aleatoria discreta integrabile; inoltre la sua speranza è data dalla relazione $\mathbf{E}[XY] = \mathbf{E}[X] \mathbf{E}[Y]$.

2.4 Il concetto generale di speranza, variabili aleatorie definite mediante densità

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 1.5). Sarà allora opportuno, almeno 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). A questo scopo, 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 definita nel paragrafo precedente.

Dato uno spazio probabilizzato (Ω, \mathcal{A}, P) , è possibile costruire, nella classe di tutte le variabili aleatorie *positive*, un'applicazione $X \mapsto \mathbf{E}[X]$ (detta **operatore di speranza**) che ad ogni variabile aleatoria reale positiva X associa un numero $\mathbf{E}[X]$ (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 $\mathbf{E}[X] = P(A)$.
2. Per ogni variabile aleatoria reale positiva X e per ogni numero reale c maggiore di zero, si ha $\mathbf{E}[cX] = c\mathbf{E}[X]$.
3. Se X, Y sono due variabili aleatorie reali positive, si ha $\mathbf{E}[X + Y] = \mathbf{E}[X] + \mathbf{E}[Y]$.
4. Se $X_1, X_2, \dots, X_n, \dots$ 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 $X_n(\omega) \uparrow X(\omega)$, allora si ha $\mathbf{E}[X_n] \uparrow \mathbf{E}[X]$.

Data ora una variabile aleatoria 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 $\mathbf{E}[X^+]$ e $\mathbf{E}[X^-]$, ed in tal caso si pone:

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

Il numero $\mathbf{E}[X]$ 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 precedente. 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 di \mathbb{R}_+ . Una siffatta variabile aleatoria si può scrivere nella forma

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

e di qui, utilizzando le proprietà appena elencate per la speranza, se ne deduce facilmente la formula (2.3).

Assegnata adesso, su uno spazio probabilizzato (Ω, \mathcal{A}, P) , una variabile aleatoria X , se esiste una funzione positiva f , tale che, per ogni intervallo I di \mathbb{R} , abbia senso l'integrale di f su I e risulti

$$P\{X \in I\} = \int_I f(x) dx, \tag{2.4}$$

si dice che X è una variabile aleatoria **dotata di densità**, e la funzione f si chiama una **densità** di X (secondo P). Dalla relazione (2.4) discende che, per una variabile aleatoria X , dotata di densità f , il calcolo delle probabilità di eventi del tipo $\{a \leq X \leq b\}$ si riconduce al calcolo di un integrale (Figura 9.1). In particolare, le regioni in cui f assume valori grandi sono le regioni nelle quali X prende valori con probabilità più elevata.

a b

Figura 9.1 L'area tratteggiata è pari alla probabilità che la variabile aleatoria X prenda valori nell'intervallo $I = [a, b]$.

Se X ammette f come densità, allora X è integrabile se e soltanto se l'integrale $\int_{\mathbb{R}} |x|f(x) dx$ è finito, e in tal caso si ha:

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

Più in generale: se X ammette f come densità, allora, per ogni variabile aleatoria della forma $g(X)$, con g funzione continua di \mathbb{R} in \mathbb{R} , affinché la variabile aleatoria $g(X)$ sia integrabile occorre e basta che l'integrale $\int_{\mathbb{R}} |g(x)|f(x) dx$ sia finito, e se questa condizione è soddisfatta, si ha:

$$\mathbf{E}[g(X)] = \int_{\mathbb{R}} g(x)f(x) dx. \quad (2.5)$$

Esempio 2.5 (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 $\mathbf{E}[X] = c^{-1} \int_a^b x dx = (a + b)/2$, cioè il punto medio dell'intervallo $[a, b]$.

2.5 Varianza e covarianza di una variabile aleatoria, la legge dei grandi numeri

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 a tale che la differenza $X - a$ sia centrata è evidentemente $a = \mathbf{E}[X]$; la differenza $X - \mathbf{E}[X]$ si chiama la variabile aleatoria centrata **associata** a X .

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

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

Intuitivamente, si può dire che la varianza di X è una misura della dispersione di X attorno al proprio valor medio a . Essa è nulla se e soltanto se la differenza $X - a$ 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 . Se X è una variabile aleatoria integrabile e dotata di densità, e se f è una densità per X , denotiamo con a la speranza di X . La formula (2.5) permette allora di scrivere:

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

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

$$\mathbf{Cov}(X, Y) = \mathbf{E}[(X - a)(Y - b)] = \mathbf{E}[XY] - ab.$$

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 $\mathbf{E}[XY] = \mathbf{E}[X]\mathbf{E}[Y]$, 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 2.3 *Su uno spazio probabilizzato, siano X, Y due variabili aleatorie integrabili e dotate di varianza finita. Si ha allora*

$$\mathbf{Var}[X + Y] = \mathbf{Var}[X] + \mathbf{Var}[Y] + 2\mathbf{Cov}(X, Y), \quad (2.6)$$

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 (2.6) da dimostrare 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:

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

Tanto basta per concludere. ■

Esempio 2.6 Su uno spazio probabilizzato, sia X una variabile aleatoria dotata di legge di Bernoulli di parametro p . Si ha allora, evidentemente, $\mathbf{E}[X^2] = \mathbf{E}[X] = p$ e quindi (ponendo $q = 1 - p$):

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

Esempio 2.7 Sia ora X una variabile aleatoria con legge binomiale di parametri 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 $\mathbf{E}[X] = np$ e $\mathbf{Var}[X] = npq$.

Esempio 2.8 Sia invece X una variabile aleatoria con legge di Poisson di parametro λ . Si ha allora

$$\mathbf{E}[X] = \sum_{k \geq 0} k e^{-\lambda} \frac{\lambda^k}{k!} = e^{-\lambda} \sum_{k \geq 1} \frac{\lambda^k}{(k-1)!}$$

e di qui, ponendo $n = k - 1$ nell'ultima somma, si trae

$$\mathbf{E}[X] = e^{-\lambda} \sum_{n \geq 0} \frac{\lambda^{n+1}}{n!} = \lambda e^{-\lambda} \sum_{n \geq 0} \frac{\lambda^n}{n!} = \lambda.$$

Con un calcolo simile si trova poi che è anche $\mathbf{Var}[X] = \lambda$. In altri termini: per una legge di Poisson il parametro reale λ coincide con la media e con la varianza.

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 enunciare questo importante teorema, occorre prima introdurre la nozione di "convergenza quasi certa" per una successione di variabili

aleatorie. A questo scopo, fissato uno spazio probabilizzato (Ω, \mathcal{A}, P) , sia $X_1, X_2, \dots, X_n, \dots$ una successione di variabili aleatorie e sia X un'altra variabile aleatoria. Diremo allora che la successione $X_1, X_2, \dots, X_n, \dots$ **converge quasi certamente** verso la variabile aleatoria X , e scriveremo $X_n \rightarrow X$, se esiste un evento quasi certo H (cioè un elemento H di \mathcal{A} con $P(H) = 1$) tale che, per ogni $\omega \in H$, sia $X_n(\omega) \rightarrow X(\omega)$.

Possiamo ora enunciare la legge dei grandi numeri:

Legge dei grandi numeri. Su uno spazio probabilizzato (Ω, \mathcal{A}, P) sia $X_1, X_2, \dots, X_n, \dots$ una successione di variabili aleatorie integrabili, indipendenti e dotate della medesima legge. Allora, indicata con a la comune speranza di queste variabili aleatorie, e posto $S_n = X_1 + \dots + X_n$, la successione (S_n/n) converge quasi certamente verso la costante a .

Nel caso particolare in cui la successione $X_1, X_2, \dots, X_n, \dots$ sia una successione di variabili aleatorie indipendenti e tutte dotate di legge di Bernoulli di parametro p , si ha $a = p$, 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 tende alla probabilità di successo in una singola prova.

Esempio 2.9 (Il metodo Montecarlo) Sia f una funzione reale limitata, definita sull'intervallo $[0, 1]$ e sia X_1, \dots, X_n, \dots una successione di variabili aleatorie indipendenti, tutte dotate della stessa legge uniforme su $[0, 1]$. Allora, la successione $f(X_1), \dots, f(X_n), \dots$ è ancora formata da variabili aleatorie indipendenti, tutte di speranza eguale a $\mathbf{E}[f(X_1)]$. Per la legge dei grandi numeri, allora,

$$\frac{f(X_1) + \dots + f(X_n)}{n} \tag{2.7}$$

converge quasi certamente verso il numero

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

Questa osservazione suggerisce un metodo 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 (2.7). 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

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.

2.2 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 in modo tale 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.

2.3 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 in modo tale 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.

2.4 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.

2.5 Un individuo lancia due dadi non truccati. Determinare la legge della variabile aleatoria che rappresenta la somma dei numeri usciti sulla facce dei dadi.

2.6 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?

2.7 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 e calcolarne la speranza. Calcolare poi la probabilità che sia uscita una pallina rossa, sapendo che è uscita una sola testa.

2.8 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 $Y = 2X - 1$.

2.9 Su uno spazio probabilizzato (Ω, \mathcal{A}, P) , sia X una variabile aleatoria discreta, a valori in $E = \{0, 1, 2, 3\}$, dotata della densità discreta seguente:

x	0	1	2	3
$f(x)$	1/6	1/3	1/6	1/3

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

2.10 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: $X + Y$, $X - 2Y$, $|X - Y|$.

2.11 Su uno spazio probabilizzato (Ω, \mathcal{A}, P) sia X una variabile aleatoria uniformemente ripartita su $[-1, 1]$. Trovare la densità della variabile aleatoria $Y = X^2$.

2.12 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 .

2.13 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.

2.14 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.

3. Il teorema limite centrale

3.1 Le leggi normali, del chi-quadro e di Student

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:

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

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 m e varianza σ^2 . La sua densità è data dalla funzione

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

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 che sia dotata di legge normale, si dice anche una variabile aleatoria **gaussiana**.

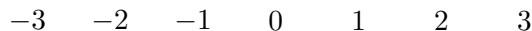


Figura 11.1 Il grafico della legge $\mathcal{N}(0, 1)$.

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

Teorema 3.1 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 3.2 Siano X, Y due variabili aleatorie indipendenti, la prima delle quali 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 $Z = X + Y$ ha legge $\mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$.

$$\sigma^2 = 0.5$$

$$\sigma^2 = 2$$
$$\mu$$

Figura 11.2 Confronto tra le densità normali per diversi valori di σ^2 .

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 3.1) e non $\mathcal{N}(0, 2)$ (come dovrebbe essere se valesse il Teorema 3.2 senza l'ipotesi d'indipendenza).

L'importanza delle leggi normali in Statistica è dovuta ad un risultato del Calcolo delle probabilità (il cosiddetto *teorema limite centrale*) 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 misurazione si può supporre che il risultato dell’operazione sia eguale al “vero” valore da misurare, più un termine casuale (l’errore) che è dovuto alla risultante di molti effetti che perturbano gli strumenti di misura e le operazioni di lettura, ciascuno dei quali dà un piccolo contributo all’errore finale. In assenza di un errore sistematico, si può pensare dunque che il risultato della misurazione sia una variabile aleatoria della forma $\mu + X$, dove μ è il vero valore da misurare mentre X è il termine di errore che dunque è una variabile aleatoria normale di media 0 (poiché supponiamo che non vi sia un errore sistematico). Dunque, è naturale rappresentare il risultato della misurazione come una variabile aleatoria Y dotata di legge normale $\mathcal{N}(\mu, \sigma^2)$. I parametri μ e σ^2 sono da stimare (sarà questo il 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 normale) viene spesso utilizzato per giustificare a priori il fatto che le osservazioni di una variabile aleatorie seguano una legge normale.

Un’altra legge importante per la Statistica è la cosiddetta “legge del chi-quadro”. Se Z_1, \dots, Z_n sono variabili aleatorie 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 l’uso che faremo delle leggi del chi-quadro non sarà necessario conoscere l’espressione esplicita della densità. Sarà comodo, però, conoscerne la speranza. Poiché, per ciascun indice j , risulta $\mathbf{E}[Z_j^2] = \mathbf{Var}[Z_j] = 1$, si ricava subito

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

In altri termini, per una variabile aleatoria con legge $\chi^2(n)$ la speranza coincide con il numero di gradi di libertà. La figura seguente descrive l’andamento di una densità $\chi^2(n)$.

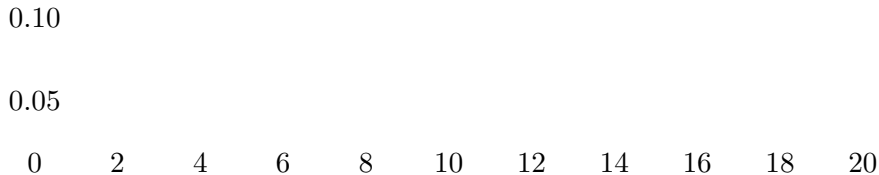


Figura 11.1 Andamento di una densità $\chi^2(7)$. Questo è l’andamento tipico delle densità $\chi^2(n)$ per $n \geq 2$. Il massimo si trova sempre un po’ prima della media, che è eguale al numero di gradi di libertà (7 in questo caso).

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** 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 $t(n)$. Sia per le leggi di Student che per le leggi del chi-quadro sarà invece importante saper calcolare le *funzioni di ripartizione*, che introdurremo nel prossimo paragrafo.

3.2 Funzioni di ripartizione

In molte occasioni, sarà importante calcolare delle probabilità del tipo $P\{X \leq t\}$, dove X è una variabile aleatoria reale. In particolare, questo problema si presenterà per variabili aleatorie con leggi $\mathcal{N}(0, 1)$, chi-quadro e di Student.

A questo scopo, introduciamo un'utile definizione. Sia X una variabile aleatoria definita su uno spazio probabilizzato (Ω, \mathcal{A}, P) . Si chiama la **funzione di ripartizione** di X , la funzione F , di \mathbb{R} in $[0, 1]$, così definita:

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

Si riconosce immediatamente che si tratta di una funzione crescente che, in generale, non è invertibile. Per ogni numero reale α , con $0 < \alpha < 1$, chiameremo **quantile di ordine α** relativo alla variabile aleatoria X il più piccolo numero reale x tale che risulti

$$P\{X \leq x\} = \alpha$$

e lo denoteremo con il simbolo x_α . Denoteremo poi con Φ la funzione di ripartizione di una variabile aleatoria di legge normale ridotta, e con ϕ_α il relativo quantile di ordine α . Tradizionalmente, i quantili della legge normale ridotta, così come quelli delle leggi del chi-quadro e di Student che vedremo tra breve, si ottengono da apposite tavole numeriche.

Uno sguardo alle tavole (poste in fondo alle dispense) mostra che la probabilità che una variabile aleatoria gaussiana, dotata di legge normale ridotta, prenda valori al di fuori dell'intervallo $[-3, 3]$ è molto bassa. Osservando che la funzione $x \mapsto e^{-x^2/2}$ è pari, si vede facilmente che risulta $P\{X \leq -t\} = P\{X \geq t\}$. Da questo fatto segue

$$P\{X \leq -\phi_\alpha\} = P\{X \geq \phi_\alpha\} = 1 - P\{X \leq \phi_\alpha\} = 1 - \alpha,$$

da cui segue la relazione

$$P\{|X| \geq \phi_{1-\alpha/2}\} = \alpha \tag{3.1}$$

della quale ci serviremo spesso nel seguito. Per provare la (3.1), basta osservare che, perché sia $|X| \geq \phi_{1-\alpha/2}$ dev'essere $X \geq \phi_{1-\alpha/2}$ oppure $X \leq -\phi_{1-\alpha/2}$ e dunque

$$P\{|X| \geq \phi_{1-\alpha/2}\} = P\{X \geq \phi_{1-\alpha/2}\} + P\{X \leq -\phi_{1-\alpha/2}\} = \frac{\alpha}{2} + \frac{\alpha}{2} = \alpha.$$

Con i simboli $t_\alpha(n)$ e $\chi_\alpha^2(n)$ si indicano i quantili di ordine α delle variabili aleatorie di legge rispettivamente $t(n)$ 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. Poiché anche la densità della legge $t(n)$ è simmetrica, per essa continuano a valere le due relazioni trovate per i quantili della legge normale:

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

Questo discorso non si può applicare invece per la legge $\chi^2(n)$ perché la sua densità non è simmetrica.

3.3 Il teorema limite centrale

Abbiamo già accennato nel paragrafo 10, sia pure senza enunciarlo, al teorema limite centrale e ad alcune sue conseguenze. Andiamo ora ad enunciarlo precisamente, ed a dedurne un'altra conseguenza che ci permetterà di introdurre alcuni utili risultati di approssimazione.

Teorema limite centrale. Consideriamo una successione $X_1, X_2, \dots, X_n, \dots$ di variabili aleatorie indipendenti, definite su un opportuno spazio probabilizzato, tutte dotate della stessa legge di speranza a e varianza finita σ^2 . Poniamo

$$S_n = X_1 + \dots + X_n, \quad T_n = \frac{S_n - \mathbf{E}[S_n]}{\sqrt{\mathbf{Var}[S_n]}} = \frac{S_n - na}{\sigma\sqrt{n}}.$$

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

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

L'importanza e l'utilità di questo risultato risiedono nel fatto che non si fa nessuna ipotesi sulla legge delle variabili aleatorie X_1, \dots, X_n, \dots in questione, purché esse abbiano speranza e varianza entrambe finite.

Un'applicazione tipica di questo risultato è la seguente: supponiamo di voler calcolare la probabilità $P\{S_n \leq t\}$, dove X_1, \dots, X_n sono variabili aleatorie indipendenti e tutte dotate della stessa legge di speranza a e varianza σ^2 (entrambe finite), e dove si sia posto $S_n = X_1 + \dots + X_n$. Allora, per n grande, basta approssimare il numero

$$P\{S_n \leq t\} = P\left\{\frac{S_n - na}{\sigma\sqrt{n}} \leq \frac{t - na}{\sigma\sqrt{n}}\right\}$$

con il numero

$$\Phi\left(\frac{t - na}{\sigma\sqrt{n}}\right)$$

(calcolabile numericamente mediante le tavole della legge normale). Questa approssimazione è comunemente chiamata **approssimazione normale**. In particolare, se X_1, \dots, X_n sono variabili aleatorie indipendenti e tutte dotate della stessa legge di Bernoulli $\mathcal{B}(1, p)$, la variabile aleatoria S_n ha legge binomiale $\mathcal{B}(n, p)$. D'altra parte, il calcolo delle probabilità della forma $P\{S_n \leq t\}$ è abbastanza complicato: è preferibile dunque utilizzare l'approssimazione normale. Poiché risulta $\mathbf{E}[S_n] = np$ e $\mathbf{Var}[S_n] = npq$ (dove si sia posto $q = 1 - p$), l'approssimazione normale, in questo caso, diviene:

$$P\{S_n \leq t\} \approx \Phi\left(\frac{t - np}{\sqrt{npq}}\right). \quad (3.2)$$

Esempio 3.1 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:

$$P\{S_{100} \geq 65\} = \sum_{k=65}^{100} \binom{100}{k} \frac{1}{2^{100}}. \quad (3.3)$$

Questo calcolo è evidentemente molto laborioso e difficilmente realizzabile senza l'uso di un calcolatore. Invece, utilizzando l'approssimazione normale (3.2), si ottiene:

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

Il calcolo esatto della (3.3) avrebbe dato come risultato $P\{S_{100} \geq 65\} = 0.00176$. Osserviamo però che, poiché S_n assume soltanto valori interi, si ha $\{S_{100} \geq 65\} = P\{S_{100} \geq 64.5\}$ e dunque, ripetendo i calcoli, si ottiene

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

che è una migliore approssimazione del valore vero 0.00176. In generale, per delle variabili aleatorie a valori interi si ottiene una migliore approssimazione prendendo $P\{S_n \leq t + 1/2\}$ piuttosto che $P\{S_n \leq t\}$, quando t sia un numero intero.

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 sia lontana dalla legge $\mathcal{N}(0, 1)$. Per esempio, se le variabili aleatorie X_1, \dots, X_n hanno legge di Bernoulli $\mathcal{B}(1, p)$ 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. 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.

Figura 13.1 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.

Figura 13.2 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.

Figura 13.3 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

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

3.2 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) $P\{X \leq 9.3\}$, (b) $P\{X \geq 10\}$ (c) $P\{6.5 \leq X \leq 7.5\}$.

3.3 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 .

3.4 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 .

3.5 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 $P\{|X - 2| \leq 2\}$.

3.6 Su uno spazio probabilizzato (Ω, \mathcal{A}, P) , sia X una variabile aleatoria gaussiana con media μ , con $\mu > 0$, e varianza $\sigma^2 = f(\mu)$. Determinare la funzione f in maniera tale che la probabilità $P\{X \leq 0\}$ non dipenda da μ .

3.7 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.

3.8 Un distributore di caffè fornisce è tarato in maniera tale la fornire 25 cc di caffè con una varianza di 4 cc². Determinare la probabilità che il distributore fornisca una tazza con più di 29 cc di caffè.

3.9 Lo scorso anno, nel corso di Istituzioni di matematica, 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?

3.10 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?

3.11 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.

3.12 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 0.2 mg. Considerato che una scatola contiene 40 pasticche, perché una scatola venga commercializzata, essa deve contenere non meno di 30 mg di principio attivo, determinare la percentuale di confezioni commercializzabili.

3.13 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.

3.14 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.

3.15 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?

3.16 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 numero 7 si presenti almeno 15 volte.

3.17 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?

3.18 Tra i novecento 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?

3.19 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.

3.20 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.

4. La statistica inferenziale

4.1 Introduzione

Nelle scienze sperimentali (chimiche, fisiche, biologiche, ...) il ricercatore si trova spesso in presenza di dati che deve elaborare per poterli interpretare: questo è il problema della statistica. Tradizionalmente essa si divide in due parti: la statistica *descrittiva* e quella *inferenziale*. La prima si occupa di organizzare e riassumere in modo significativo i dati raccolti, e qui termina il suo compito. La seconda, invece, utilizzando i metodi e le nozioni del calcolo delle probabilità, cerca di fare delle previsioni, o di ottenere dei risultati estendibili all'intera popolazione, pur partendo solo da un piccolo campione effettivamente osservato.

Esempio 4.1 Una popolazione è composta da due tipi d'individui: quelli di tipo A e quelli di tipo B (Esempio 2.3). Supponiamo però di non conoscere il rapporto effettivo tra il numero d'individui di tipo A 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 delle quantità casuali x_1, \dots, x_n che possono assumere soltanto i valori 0 oppure 1. Poiché queste osservazioni sono casuali, si può pensare che i dati raccolti x_1, \dots, x_n siano i valori assunti da n variabili aleatorie X_1, \dots, X_n indipendenti e definite su un opportuno spazio probabilizzato. Di queste, però, abbiamo soltanto l'informazione che esse hanno legge di Bernoulli, della quale non conosciamo il parametro p .

Esempio 4.2 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, questi dati raccolti si potranno pensare come i valori assunti da n variabili aleatorie X_1, \dots, X_n indipendenti e definite su un opportuno spazio probabilizzato. Di esse sappiamo che si tratta di variabili aleatorie gaussiane, ma non sono note la speranza e la varianza. Lo scopo stesso della misurazione è quello di stimare quale sia il valore della comune speranza delle variabili aleatorie.

Dunque, in un problema di statistica, siamo in presenza di un esperimento aleatorio che produce un certo numero n di quantità osservate x_1, \dots, x_n che si modellizzano come i valori assunti da n variabili aleatorie X_1, \dots, X_n definite su un opportuno spazio probabilizzato, la cui legge dipende da un parametro sconosciuto θ del quale sappiamo soltanto che appartiene ad un certo insieme Θ .

Saremo spesso condotti a considerare quantità come “la probabilità che una funzione delle osservazioni prenda valori nell'intervallo I ”, oppure “la speranza di una funzione delle osservazioni”. Queste quantità dipendono dalla legge delle osservazioni, e dunque dal parametro sconosciuto θ . Per dare un significato rigoroso a queste quantità supporremo che le variabili aleatorie X_1, \dots, X_n che modellizzano le osservazioni siano delle variabili aleatorie definite su uno spazio del tipo

$$(\Omega, \mathcal{A}, (P_\theta)_{\theta \in \Theta}) \quad (4.1)$$

dove P_θ è una probabilità sullo spazio probabilizzabile (Ω, \mathcal{A}) , che descrive l'insieme di tutti i possibili risultati dell'esperimento aleatorio e tutti gli eventi osservabili. Se θ è un elemento di Θ , se Y è una variabile aleatoria integrabile, definita sullo spazio probabilizzato $(\Omega, \mathcal{A}, P_\theta)$, la sua speranza si denota con $\mathbf{E}_\theta[Y]$ (da leggersi: “la speranza di Y per il valore θ del parametro”).

Ogni terna del tipo (4.1) si chiama un *modello statistico*. La prima operazione da affrontare in un problema di tipo statistico consiste dunque nel costruire un modello statistico adatto a

descrivere ragionevolmente il problema. In pratica si tratterà di stabilire quali siano i possibili valori del parametro sconosciuto θ e quali siano le leggi delle variabili aleatorie X_1, \dots, X_n secondo ciascuna delle probabilità P_θ . Una situazione molto frequente è quella (che compare negli esempi fatti) in cui le osservazioni X_1, \dots, X_n siano costituite da una sequenza di n variabili aleatorie indipendenti e tutte dotate della medesima legge μ_θ , con $\theta \in \Theta$. Chiameremo questo modello statistico un ***campione*** di taglia n e di legge μ_θ .

È del tutto naturale porsi delle domande sul parametro sconosciuto θ e considerare il fenomeno aleatorio che ne dipende (nel caso dell'Esempio 4.1, l'estrazione degli n individui dalla popolazione) semplicemente come un esperimento che si compie al solo scopo di trarne qualche indicazione sul parametro θ .

Più precisamente, quando si sia in presenza di una situazione quale quella sopra descritta, uno dei problemi che spesso si possono affrontare è il *problema della stima*. Esso consiste nella scelta di uno *stimatore*, cioè di un'applicazione T di Ω in Θ . Questa applicazione rappresenta la strategia seguente: ci s'impegna, qualunque sarà la realizzazione ω dell'esperimento, ad attribuire convenzionalmente al parametro sconosciuto θ il "valore stimato" $T(\omega)$. Il problema consiste nello scegliere lo stimatore T in modo da minimizzare certe quantità, di natura probabilistica, ad esso legate (ed espresse mediante le probabilità P_θ).

Questo è, per essere più precisi, il problema della *stima puntuale*. Un genere di stima leggermente differente consiste nello scegliere un'applicazione S di Ω in $\mathcal{P}(\Theta)$; questa applicazione rappresenta 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 $S(\omega)$ (*intervallo di fiducia*).

Un terzo tipo di problema è quello dei *test d'ipotesi*. Sia data un'*ipotesi* concernente il vero 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* per questa ipotesi, cioè una partizione (D, D^c) di Ω ; questa partizione rappresenta la strategia seguente: ci s'impegna, qualunque sia la realizzazione ω dell'esperimento, a rifiutare l'ipotesi se ω appartiene a D , ad accettarla se ω appartiene a D^c .

La differenza fondamentale tra il probabilista puro e lo statistico è che quest'ultimo 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'impagna *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 .

4.2 Problemi di stima

Consideriamo un campione X_1, \dots, X_n di taglia n . Come abbiamo detto, il primo dei problemi dello statistico è quello di ricavare, dalle osservazioni x_1, \dots, x_n , alcune informazioni sul parametro θ o, più in generale, su una funzione $\psi(\theta)$ del parametro. A questo scopo, si chiama ***statistica*** ogni variabile aleatoria della forma

$$T = f(X_1, \dots, X_n).$$

Data una funzione ψ di Θ in \mathbb{R} , chiameremo **stimatore** di $\psi(\theta)$, ogni statistica T . Intuitivamente, dare uno stimatore T significa fissare la regola che, se i dati raccolti dalle osservazioni sono x_1, \dots, x_n , si stima la quantità sconosciuta $\psi(\theta)$ con il numero $f(x_1, \dots, x_n)$. Notiamo, comunque, che il valore assunto dallo stimatore è un'approssimazione del parametro $\psi(\theta)$. In effetti, uno stimatore, come ogni statistica, è una variabile aleatoria (è una funzione delle osservazioni) e dunque non assumerà quasi mai il valore $\psi(\theta)$ da stimare, anche se, naturalmente, si spera che esso prenda valori da esso non lontani.

Per la definizione che abbiamo dato, *qualunque* funzione delle osservazioni è uno stimatore. Occorre quindi disporre di qualche criterio con cui stabilire quali funzioni delle osservazioni sono “buoni” stimatori. Noi non entreremo nel dettaglio di questo argomento, perché ci limiteremo qui a considerare soltanto gli stimatori “naturali” dei semplici problemi che affronteremo (soprattutto la media e la varianza empirica). Per dare un cenno, comunque, introduciamo la seguente definizione.

Diremo che una statistica T è uno stimatore **corretto** (o **non distorto**) del parametro $\psi(\theta)$ se risulta

$$\mathbf{E}_\theta [T] = \psi(\theta) \quad \text{per ogni } \theta \in \Theta.$$

In altri termini: T può prendere valori diversi da $\psi(\theta)$, ma, se è corretto, la “media” dei suoi valori dev'essere $\psi(\theta)$, qualunque sia θ .

Esempio 4.3 Consideriamo un campione X_1, \dots, X_n di taglia n , e supponiamo che, per ogni θ , una (e quindi ciascuna) di queste variabili aleatorie sia integrabile. Uno stimatore corretto per $\mathbf{E}_\theta [X_1]$ è fornito dalla **media empirica** definita da

$$\bar{X} = \frac{X_1 + \dots + X_n}{n}.$$

Per riconoscerlo, basta osservare che, dalle proprietà della speranza, si ha:

$$\mathbf{E}_\theta [\bar{X}] = \mathbf{E}_\theta \left[\frac{X_1 + \dots + X_n}{n} \right] = \frac{1}{n} \left\{ \mathbf{E}_\theta [X_1] + \dots + \mathbf{E}_\theta [X_n] \right\} = \mathbf{E}_\theta [X_1].$$

Esempio 4.4 Consideriamo ora un campione X_1, \dots, X_n di taglia n , e supponiamo che, per ogni θ , una (e quindi ciascuna) di queste variabili aleatorie abbia varianza finita. Cerchiamo uno stimatore per $\mathbf{Var}_\theta [X_1]$. Iniziamo col supporre che la speranza di X_1 sia nota (cioè indipendente da θ) e denotiamola con m . Si ha allora:

$$\mathbf{E}_\theta \left[\frac{1}{n} \sum_{i=1}^n (X_i - m)^2 \right] = \frac{1}{n} \left\{ \mathbf{E}_\theta [(X_1 - m)^2] + \dots + \mathbf{E}_\theta [(X_n - m)^2] \right\} = \mathbf{Var}_\theta [X_1]$$

e dunque la variabile aleatoria

$$\frac{1}{n} \sum_{i=1}^n (X_i - m)^2 \tag{4.2}$$

è uno stimatore corretto di $\mathbf{Var}_\theta [X_1]$. D'altra parte, nei casi concreti, di solito, la speranza m non è nota, per cui non è possibile calcolare la (4.2). Per trovare uno stimatore che funzioni anche in questo caso, allora, proviamo a sostituire m con la media empirica e vediamo quali sono le proprietà della variabile aleatoria

$$\hat{S}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$$

come stimatore di $\mathbf{Var}_\theta[X_1]$. Ricordato l'eguaglianza $\sum_{i=1}^n X_i = n\bar{X}$, si ha:

$$\sum_{i=1}^n (X_i - \bar{X})^2 = \sum_{i=1}^n X_i^2 - 2\bar{X} \sum_{i=1}^n X_i + \sum_{i=1}^n X_i^2 = \sum_{i=1}^n X_i^2 - n\bar{X}^2.$$

Inoltre, per definizione di varianza, si ha:

$$\mathbf{E}_\theta[X_1^2] = \mathbf{Var}_\theta[X_1] + \mathbf{E}_\theta[X_1]^2, \quad \mathbf{E}_\theta[\bar{X}^2] = \frac{1}{n} \mathbf{Var}_\theta[X_1] + \mathbf{E}_\theta[X_1]^2.$$

Se ne deduce

$$\mathbf{E}_\theta[n\widehat{S}^2] = n\mathbf{Var}_\theta[X_1] + n\mathbf{E}_\theta[X_1]^2 - \mathbf{Var}_\theta[X_1] - n\mathbf{E}_\theta[X_1]^2 = (n-1)\mathbf{Var}_\theta[X_1]$$

e dunque $\widehat{S}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$ non è uno stimatore corretto. Se, però, si modifica la costante di normalizzazione, allora la variabile aleatoria

$$S^2 = \frac{n}{n-1} \widehat{S}^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2,$$

detta la **varianza empirica**, è uno stimatore corretto della varianza.

Sia T uno stimatore per il parametro $\psi(\theta)$ e sia α un numero reale, con $0 < \alpha < 1$. Si chiama **intervallo di fiducia di livello** $1 - \alpha$ un intervallo della forma $[T - \delta, T + \delta]$, se, per ogni θ , risulta

$$P_\theta\{|T - \psi(\theta)| < \delta\} \geq 1 - \alpha.$$

In altri termini, un intervallo di fiducia di livello α per $\psi(\theta)$ è un intervallo della retta reale per il quale si possa affermare che, con probabilità $1 - \alpha$, il vero valore $\psi(\theta)$ cada in questo intervallo. Evidentemente, si tratta di un "intervallo aleatorio" nel senso che dipende dalle osservazioni e dunque, in ultima analisi, dalla particolare realizzazione dell'esperimento.

In pratica, il valore di α sarà scelto abbastanza piccolo (i valori tipici sono 0.05 oppure 0.01); un intervallo di fiducia sarà dunque tale che il parametro da stimare si trovi al suo interno con un'altissima probabilità.

Esempio 4.5 Sia X_1, \dots, X_n un campione di n leggi di Bernoulli di parametro sconosciuto θ . Cerchiamo un intervallo di fiducia di livello $1 - \alpha$. Con le notazioni usuali, si ha, usando l'approssimazione normale, si riconosce immediatamente che, siccome la variabile aleatoria

$$Y = \frac{\bar{X} - \theta}{\sqrt{\theta(1-\theta)}} \sqrt{n}$$

ha approssimativamente legge normale ridotta, osservato che il termine sconosciuto $\theta(1-\theta)$ è inferiore a $1/4$, si ha:

$$P_\theta\{|\bar{X} - \theta| \geq \delta\} = P_\theta\left\{|Y| \geq \frac{\delta\sqrt{n}}{\sqrt{\theta(1-\theta)}}\right\} \leq P_\theta\{|Y| \geq 2\delta\sqrt{n}\} = \alpha.$$

Di qui, ricordata la (3.1), si trae:

$$2\delta\sqrt{n} = \phi_{1-\alpha/2}.$$

E dunque $[\bar{X} - \frac{\phi_{1-\alpha/2}}{2\sqrt{n}}, \bar{X} + \frac{\phi_{1-\alpha/2}}{2\sqrt{n}}]$ è (approssimativamente) un intervallo di fiducia di livello $1 - \alpha$ per il parametro sconosciuto θ .

4.3 Stima della media e della varianza per campioni gaussiani

Consideriamo un campione di taglia n di variabili aleatorie gaussiane X_1, \dots, X_n . L'obiettivo di questo paragrafo è quello di costruire un intervallo di fiducia per la loro speranza.

Un primo semplicissimo caso (ma raramente utile nella pratica) è quello in cui si suppone che le osservazioni X_1, \dots, X_n abbiano tutte legge $\mathcal{N}(\mu, \sigma^2)$, dove σ è un numero fissato e conosciuto. In questo caso, per i teoremi 3.1 e 3.2, la variabile aleatoria

$$\frac{\bar{X} - \mu}{\sigma} \sqrt{n}$$

ha legge normale ridotta $\mathcal{N}(0, 1)$. Si ha allora, ricordando la (3.1):

$$1 - \alpha = P_\mu \left\{ \left| \frac{\bar{X} - \mu}{\sigma} \sqrt{n} \right| \leq \phi_{1-\alpha/2} \right\} = P_\mu \left\{ |\bar{X} - \mu| \leq \frac{\sigma}{\sqrt{n}} \phi_{1-\alpha/2} \right\}$$

e dunque

$$\left[\bar{X} - \frac{\sigma}{\sqrt{n}} \phi_{1-\alpha/2}, \bar{X} + \frac{\sigma}{\sqrt{n}} \phi_{1-\alpha/2} \right]$$

è un intervallo di fiducia di livello $1 - \alpha$ per μ . In pratica, però, la varianza σ^2 è raramente nota, e quindi non è possibile calcolare esplicitamente l'intervallo. 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 di σ^2 . Allo scopo di rispondere a questa domanda, introduciamo un importante teorema del calcolo delle probabilità. Precisamente:

Teorema di Cochran. *Siano X_1, \dots, X_n variabili aleatorie indipendenti e tutte dotate della medesima legge normale $\mathcal{N}(\mu, \sigma^2)$ e poniamo come al solito:*

$$\bar{X} = \frac{X_1 + \dots + X_n}{n}, \quad S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2.$$

Poniamo anche:

$$W = \frac{(n-1)S^2}{\sigma^2}, \quad T = \frac{\bar{X} - \mu}{\sigma} \sqrt{n}.$$

Allora la variabile aleatoria W ha legge $\chi^2(n-1)$ e T ha legge $t(n-1)$. Inoltre, queste due variabili aleatorie sono tra loro indipendenti.

Con l'ausilio del teorema di Cochran possiamo velocemente trovare un intervallo di fiducia per μ . Basta per questo osservare che si ha:

$$1 - \alpha = P_\theta \left\{ |T| \leq t_{1-\alpha/2}(n-1) \right\} = P_\theta \left\{ \left| \frac{\bar{X} - \mu}{S} \sqrt{n} \right| \leq t_{1-\alpha/2}(n-1) \right\}$$

che vuole dire che

$$\left[\bar{X} - \frac{S}{\sqrt{n}} t_{1-\alpha/2}(n-1), \bar{X} + \frac{S}{\sqrt{n}} t_{1-\alpha/2}(n-1) \right]$$

è un intervallo di fiducia per μ di livello $1 - \alpha$. Confrontando questo intervallo con quello dato dalla (3.1) (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 di sostituire ai quantili della legge normale quelli della legge di Student $t(n-1)$ (che sono un po' più grandi).

Esempio 4.6 La tabella seguente riporta cento misurazioni della velocità della luce nell'aria, effettuate dal grande fisico sperimentale Michelson 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 l'intervallo di fiducia per la velocità della luce nell'aria al livello 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:

$$\bar{X} = 852.40, \quad S^2 = 6242.67$$

e quindi l'intervallo di fiducia, al nostro abituale livello 0.95, è

$$[852.4 - 15.67, 852.4 + 15.67] = [836.73, 868.07]$$

che si può anche esprimere dicendo che la velocità della luce nell'aria è 299852.4 ± 15.67 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.

Concludiamo questo paragrafo calcolando un intervallo di fiducia per la varianza di un campione gaussiano. A questo scopo, iniziamo con l'osservare che la varianza σ^2 è un numero positivo, così come il suo stimatore S^2 . Inoltre, sappiamo dal teorema di Cochran che la variabile aleatoria $(n-1)S^2/\sigma^2$ ha legge $\chi^2(n-1)$. Si ha dunque:

$$1 - \alpha = P_{\theta} \left\{ \frac{n-1}{\sigma^2} S^2 \geq \chi_{\alpha}^2(n-1) \right\} = P_{\theta} \left\{ \sigma^2 \leq \frac{(n-1)S^2}{\chi_{\alpha}^2(n-1)} \right\},$$

che significa che un intervallo di fiducia di livello $1 - \alpha$ per la varianza è:

$$\left[0, \frac{(n-1)S^2}{\chi_{\alpha}^2(n-1)} \right].$$

Per calcolare un intervallo più preciso, ripetiamo il ragionamento appena fatto:

$$\begin{aligned} 1 - \alpha &= P_{\theta} \left\{ \chi_{\alpha/2}^2(n-1) \leq (n-1)S^2/\sigma^2 \leq \chi_{1-\alpha/2}^2(n-1) \right\} \\ &= P_{\theta} \left\{ \frac{(n-1)S^2}{\chi_{1-\alpha/2}^2(n-1)} \leq \sigma^2 \leq \frac{(n-1)S^2}{\chi_{\alpha/2}^2(n-1)} \right\} \end{aligned}$$

da cui si ottiene che

$$\left[\frac{(n-1)S^2}{\chi_{1-\alpha/2}^2(n-1)}, \frac{(n-1)S^2}{\chi_{\alpha/2}^2(n-1)} \right] \tag{4.3}$$

è un intervallo di fiducia di livello $1 - \alpha$ per σ^2 .

4.4 Test d'ipotesi statistiche

Un tipico problema di statistica consiste nello stabilire se il parametro incognito θ sia di un certo tipo oppure no. È molto comune in numerosi campi di applicazione (medicina, scienza, tecnologia, industria, ...) di dover trovare una risposta del tipo “sì” o “no” ad una domanda. Il nuovo farmaco è realmente efficace? Il partito A ha effettivamente più elettori del partito B ? La nuova macchina è davvero più efficiente del vecchio modello? Per dare risposta a queste domande, in genere, si effettua un esperimento aleatorio, o una misurazione, in modo da procurarsi delle osservazioni (di solito un campione) per mezzo delle quali si cerca di dare risposta alla domanda posta.

Tutte queste problematiche si possono esprimere, dopo aver costruito un opportuno modello statistico del fenomeno, chiedendosi se il parametro θ si trovi in un certo sottoinsieme Θ_0 di Θ oppure no. Questa è appunto la situazione generale: in un problema di test, ci si trova in presenza di una partizione $\{\Theta_0, \Theta_1\}$ di Θ e si vuole stabilire se il vero valore di θ appartiene a Θ_0 oppure a Θ_1 . Tradizionalmente, gli insiemi Θ_0 e Θ_1 non hanno un carattere simmetrico: il primo si chiama l'*ipotesi* mentre il secondo l'*alternativa*.

Lo scopo di un test è quello di scegliere tra due possibilità: se respingere oppure no l'ipotesi. Questa decisione equivale a stabilire qual è l'insieme delle realizzazioni che conducono al rigetto dell'ipotesi. Chiameremo quest'insieme la *regione critica* del test. In generale, qualunque sia la scelta della regione critica, 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, come abbiamo detto, 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 è peggio che non mettere in commercio un farmaco funzionante).

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_1 , 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 4.7 La settimana successiva al suicidio di un famoso personaggio televisivo, in una 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 il fatto che una persona che 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 suicidi) 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 possa modellizzare con una variabile aleatoria X avente legge di Poisson di parametro 8.

Dire 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 diverso da 8. Usiamo dunque come modello statistico un campione (di taglia 1) di legge di Poisson di parametro θ , con $\theta \in \Theta = \mathbb{R}_+$.

In questo caso, l'ipotesi è $\Theta_0 =]0, 8]$ 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 $\sup_{0 < \theta \leq 8} P_\theta\{X \geq k\} = P_8\{X \geq k\} \leq 0.05$. Calcolando numericamente la funzione di ripartizione della legge di Poisson di parametro 8, si trae

$$\begin{aligned} P_8\{X \geq 12\} &= 0.112 \\ P_8\{X \geq 13\} &= 0.064 \\ P_8\{X \geq 14\} &= 0.034 \end{aligned}$$

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 (al meno al livello 0.05).

4.5 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 di μ , e cerchiamo di determinare un numero δ maggiore di zero in modo tale che, se l'ipotesi è vera, allora si abbia

$$P_\theta\left\{|\bar{X} - \mu_0| > \delta\right\} = \alpha.$$

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:

$$P_{\theta}\{|\bar{X} - \mu| > \delta\} = P_{\theta}\{|T| \geq \sqrt{n}\delta/S\}$$

e, come abbiamo visto nel paragrafo 3.2, questa quantità vale α se risulta $\sqrt{n}\delta/S = t_{1-\alpha/2}(n-1)$ ovvero se

$$\delta = \frac{S}{\sqrt{n}}t_{1-\alpha/2}(n-1). \quad (4.4)$$

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 che la media empirica \bar{X} differisca da μ_0 per una quantità maggiore di δ , dove δ è dato appunto dalla (4.4). Oppure, in maniera più semplice, basterà calcolare la statistica T e rigettare l'ipotesi se $|T|$ risulta più grande di $t_{1-\alpha/2}(n-1)$.

Esempio 4.8 L'altezza media degli uomini di un paese era di 170 cm nel 1957. Su $n = 100$ 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

$$|T| = \left| \frac{\bar{X} - \mu_0}{S} \right| \sqrt{n}$$

e di confrontarla con il numero $t_{0.975}(99)$. Sostituendo i valori, si ha:

$$\left| \frac{\bar{X} - \mu_0}{S} \sqrt{n} \right| = \left| \frac{171 - 170}{4} \cdot 10 \right| = 2.5$$

mentre il quantile è $t_{0.975}(99) = 1.98$. Quindi l'ipotesi è respinta e si può 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 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

$$P_{\theta}\{T_0 \geq t_{1-\alpha}(n-1)\} \leq P_{\theta}\{T \geq t_{1-\alpha}(n-1)\} = \alpha$$

ovvero $\{T_0 \geq t_{1-\alpha}(n-1)\}$ 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”.

4.6 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 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 di σ^2 , e cerchiamo di determinare un numero δ maggiore di zero in maniera tale che, se l'ipotesi è vera, allora si abbia

$$P_\theta\{S^2 > \delta\} = \alpha.$$

Per un tale valore di δ , l'evento $\{S^2 > \delta\}$ 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 che (per il teorema di Cochran), 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, si ha $\sigma^2 \leq \sigma_0^2$, si trae:

$$P_\theta\{S^2 > \delta\} = P_\theta\left\{W > \frac{\delta(n-1)}{\sigma^2}\right\} \leq P_\theta\left\{W > \frac{\delta(n-1)}{\sigma_0^2}\right\}.$$

Ora, quest'ultima quantità vale α se risulta $\delta(n-1)/\sigma_0^2 = \chi_{1-\alpha}^2(n-1)$ ovvero se

$$\delta = \frac{(n-1)\chi_{1-\alpha}^2\sigma_0^2}{n-1}. \quad (4.5)$$

Per questo valore di δ , dunque, l'evento $\{S^2 > \delta\}$ 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 (4.5). Oppure, in maniera più semplice, basterà calcolare la statistica $W = \frac{(n-1)S^2}{\sigma_0^2}$ e rigettare l'ipotesi se il valore trovato risulta maggiore di $\chi_{1-\alpha}^2(n-1)$.

Esempio 4.9 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 $\chi_{0.99}(24)$. 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

$$W = \frac{(n-1)S^2}{\sigma_0^2}$$

ha legge $\chi^2(n-1)$ e dunque, ripetendo i passaggi fatti alla fine del paragrafo 16 per costruire l'intervallo di fiducia per la varianza, si ottiene che

$$\left\{ S < \frac{\chi_{\alpha/2}^2(n-1)\sigma_0^2}{n-1} \right\} \cup \left\{ S > \frac{\chi_{1-\alpha/2}^2(n-1)\sigma_0^2}{n-1} \right\}$$

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”.

4.7 Il test del chi–quadrato

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 $\theta = (\theta_1, \dots, \theta_m)$, con $\theta_1 + \dots + \theta_m = 1$ e poniamo, per ciascun indice j compreso tra 1 e m ,

$$P_\theta\{X_1 = x_j\} = \theta_j.$$

Così, legge del campione è determinata non appena si conosca il vettore θ , che rappresenta proprio la “densità discreta” della suddetta legge secondo P_θ . Noi vogliamo stabilire se il campione segue la legge corrispondente ad un certo parametro $\theta_0 = (p_1, \dots, p_m)$ 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 θ_0 ” contro l'alternativa “il campione segue una legge differente da quella determinata da θ_0 ”. A questo scopo, per ciascun indice j compreso tra 1 e m , poniamo

$$O_j(\omega) = \text{Card}\{i \in \{1, \dots, n\} : X_i(\omega) = x_j\}.$$

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 θ_0 . Poniamo infine

$$T = \sum_{j=1}^m \frac{(O_j - E_j)^2}{E_j}. \quad (4.6)$$

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 perché sia $E_j = np_j \geq 5$ per ciascun indice j .

Esempio 4.10 Un dado viene lanciato 2400 volte con i seguenti risultati:

1	2	3	4	5	6
450	421	395	358	387	389

Si può affermare, ad un livello $\alpha = 0.05$, che il dado è equilibrato?

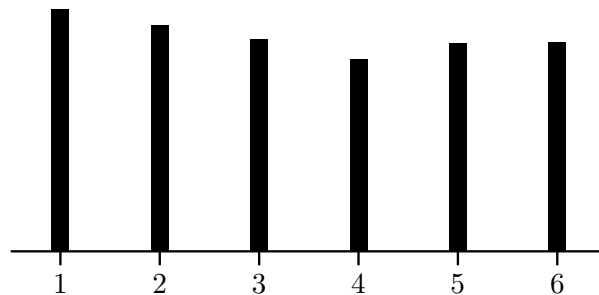


Figura 20.1 Andamento degli effettivi empirici per i 2400 lanci di dado dell'Esempio 4.10.

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 possa assumere 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, $P_\theta\{Y_1 = k\} = P_\theta\{X_1 \in I_k\} = p_k$. 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 $\theta_0 = (p_1, \dots, p_m)$.

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 a dei compromessi, da valutare caso per caso.

Esempio 4.11 Nella tabella sottostante sono riportati 63 numeri. Si può affermare, ad un livello $\alpha = 0.05$, che si tratta di un campione estratto da una legge gaussiana $\mathcal{N}(0, 1)$?

-0.69	1.52	-0.42	-0.39	-0.82	0.36	-0.45	1.25	-2.05	-0.97	-0.15
1.41	0.27	-1.12	0.42	-0.98	0.99	-0.59	1.56	0.70	-0.72	-0.09
3.11	-0.43	-0.73	0.51	-0.79	0.72	-1.25	2.01	-0.95	0.52	-1.02
2.07	0.74	0.40	3.05	0.75	0.03	0.63	0.29	0.99	0.05	0.58
0.32	0.12	0.88	-1.28	0.19	0.73	-0.91	0.79	-1.13	-0.63	-0.83
-0.17	0.87	-0.02	3.14	-0.21	0.59	0.92	-0.81			

Si tratta 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, si avrà:

$$p_1 = \Phi(x_1), \quad p_m = 1 - \Phi(x_{m-1}),$$

$$p_k = \Phi(x_k) - \Phi(x_{k-1}) \quad (k = 2, 3, \dots, m-1).$$

Per semplificare il conto (anche se non sarebbe necessario), supponiamo che sia $p_k = 1/m$ per ciascun indice k . In questo modo, come 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, $63/m \geq 5$, o ciò ch'è lo stesso, $m \leq 63/5 = 12.6$. 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 $k = 1, 2, \dots, 9$.

$\phi_{0.1}$	$\phi_{0.2}$	$\phi_{0.3}$	$\phi_{0.4}$	$\phi_{0.5}$	$\phi_{0.6}$	$\phi_{0.7}$	$\phi_{0.8}$	$\phi_{0.9}$
-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 2); 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	2	6.3	-4.3	2.93
2	8	6.3	1.7	0.46
3	9	6.3	2.7	1.16
4	4	6.3	-2.3	0.84
5	5	6.3	-1.3	0.27
6	4	6.3	-2.3	0.84
7	8	6.3	1.7	0.46
8	9	6.3	2.7	1.16
9	6	6.3	-0.3	0.01
10	8	6.3	1.7	0.46

Si tratta adesso di calcolare la statistica T e di confrontarla con il numero $\chi_{0.95}^2(9)$. Sostituendo i valori, si ha $T = 8.59$ 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.

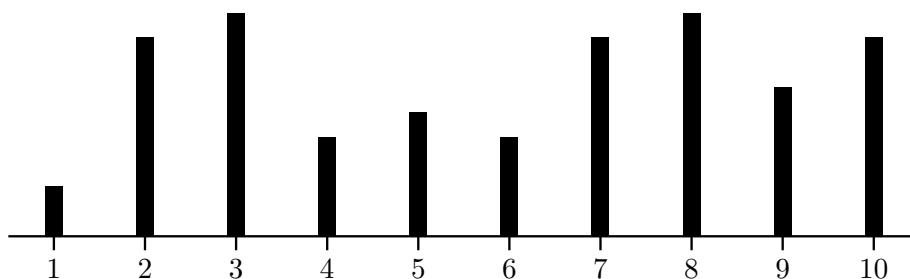


Figura 20.2 Andamento degli effettivi empirici per i numeri dell'Esempio 4.11.

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 dicostamento dalla legge teorica è notevole, oppure se la taglia del campione è grande.

Sarebbe molto utile se si potesse 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 $Q(\theta_1, \dots, \theta_r)$ dipendente dagli r parametri $\theta_1, \dots, \theta_r$.

Per prima cosa, occorre stimare i parametri $\theta_1, \dots, \theta_r$ con degli stimatori corretti, 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 $Q(\hat{\theta}_1, \dots, \hat{\theta}_r)$. Allora, se l'ipotesi è vera, cioè se il campione segue una legge appartenente alla famiglia $Q(\theta_1, \dots, \theta_r)$, 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 4.12 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:

num. di tane = x_j	0	1	2	3	4
num. di ettari = O_j	19	30	20	14	12

Se si suppone che le volpi scelgano il luogo dove fare 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 “ X ha legge $\mathcal{P}(\lambda)$ ”. Poiché il parametro λ non è noto, esso dev'essere stimato mediante un “opportuno” stimatore corretto. Poiché il parametro λ coincide con la speranza $\mathbf{E}[X]$, si riconosce subito che un suo stimatore corretto è dato dalla media empirica \bar{X} . Nel nostro caso, dunque, si ha:

$$\bar{X} = \frac{0 \cdot 19 + 1 \cdot 30 + 2 \cdot 20 + 3 \cdot 14 + 4 \cdot 12}{95} = 1.68.$$

Poniamo dunque $\lambda = 1.68$ e calcoliamo gli effettivi teorici.

$$\begin{aligned} P\{X = 0\} &= e^{-\lambda} = 0.19, & E_0 &= 95 \cdot 0.19 = 17.63 \\ P\{X = 1\} &= \lambda e^{-\lambda} = 0.31, & E_1 &= 95 \cdot 0.31 = 29.69 \\ P\{X = 2\} &= \frac{\lambda^2}{2} e^{-\lambda} = 0.26, & E_2 &= 95 \cdot 0.26 = 25.01 \\ P\{X = 3\} &= \frac{\lambda^3}{3!} e^{-\lambda} = 0.15, & E_3 &= 95 \cdot 0.15 = 14.4 \\ P\{X \geq 4\} &= 1 - P\{X < 4\} = 0.09, & E_4 &= 95 \cdot 0.09 = 8.63 \end{aligned}$$

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 avrà approssimativamente legge $\chi^2(3)$. Essa dovrà dunque 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.

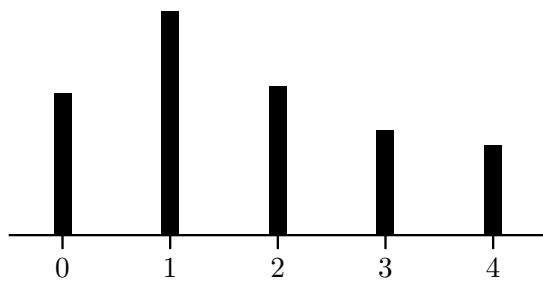


Figura 20.2 Andamento degli effettivi empirici il numero di tane per ettaro dell'Esempio 4.12.

Esercizi

4.1 Si effettua per 100 volte la misura della lunghezza di un tavolo, ottenendo così una media $\bar{X} = 112$ cm e una varianza $S^2 = 40$ cm². Determinare un intervallo di fiducia per la media di livello 0.95.

4.2 Trovare un intervallo di fiducia di livello $\alpha = 0.1$ per la media del seguente campione di misure ripetute con varianza assegnata $\sigma^2 = 16$:

4.31	1.30	1.62	2.27	3.27	4.23	0.21	1.98	3.31	2.97
------	------	------	------	------	------	------	------	------	------

Calcolare poi lo stesso intervallo di fiducia se si suppone che la varianza σ^2 sia sconosciuta.

4.3 I dati riportati nella tabella seguente riportano i valori della pressione arteriosa di un paziente (esprese in mm Hg)

126.2	106.3	113.8	120.5	128.5	123.2	127.8	106.4	108.7	109.3
123.2	111.5	126.1	119.9	121.8	124.1	107.4	124.5	119.7	118.6

Calcolare un intervallo di fiducia di livello 0.99 per la media.

4.4 La tabella seguente riporta le misurazioni (esprese in km/h) della velocità di una vettura

207	183	188	200	205	202	205	186	181	189	210	191
206	199	208	210	209	193	181	201	207	199	183	178
215	203	209	194	212	200	223	206	197			

Calcolare un intervallo di fiducia di livello 0.90 per la media.

4.5 Il sindaco di una città posta vicino ad una grande fabbrica vuole conoscere la media giornaliera della quantità di emissione di ossido di zolfo emessi dalla fabbrica; un campione casuale di 16 giorni ha dato una media di 530 kg di emissioni, con uno scarto quadratico medio di circa 300 kg al giorno; al livello 0.90, trovare un intervallo di fiducia per la media.

4.6 Vengono compiute 22 misurazioni del numero di ottani della benzina prodotta da varie compagnie, rilevando così uno scarto quadratico medio di 0.8. Qual è un intervallo di fiducia di livello $\alpha = 0.025$ per la varianza?

4.7 Per misurare l'indice di rifrazione di un vetro, si ripetono 5 misure, supponendo che il risultato di ciascuna misura abbia legge normale, e sia nota la varianza. Si ottiene così, ad un livello $\alpha = 0.1$, l'intervallo di fiducia [1.52, 1.54].

- Individuare il valor medio ottenuto dal campione, e la varianza utilizzata nel calcolo eseguito.
- Se, a parità di livello, si vuole ottenere una precisione doppia, quante misure si debbono ancora fare?
- Se, invece, si vuole la stessa precisione ad un livello $\alpha = 0.05$, quante misure si debbono ancora fare?

4.8 Una fabbrica produce chiodi di metallo di peso medio 25 g. 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?

4.9 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 $\alpha = 0.01$?

4.10 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?

4.11 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 e considera un campione casuale di 36 giorni e osserva che in media sono stati 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?

4.12 Il responsabile di una compagnia di trasporti ritiene che il carico medio consegnato sia 450 tonnellate. Il responsabile di magazzino contesta l'affermazione e pertanto registra un campione casuale di 25 trasporti e trova che il carico medio corrispondente a 446 tonnellate, con uno scarto quadratico medio di 0.25 tonnellate. Ad un livello $\alpha = 0.05$ l'affermazione del responsabile può essere rigettata?

4.13 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. Ad un livello $\alpha = 0.01$ si può dire che i due parametri sono cambiati?

4.14 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 ancora correttamente oppure no?

4.15 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?

4.16 In cento pagine dattiloscritte da una segretaria, sono stati contrassegnati i seguenti numeri di errori per pagina:

num. di errori:	0	1	2	3	4	5	6
num. 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?

4.17 Nella tabella sottostante, sono riportati i valori della velocità del vento al suolo (espressa in nodi) registrati a La Spezia 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.

4.18 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:

num. di pezzi difettosi:	0	1	2	3	4	5	6
num. 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.

4.19 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?

4.20 In 100 periodi di tempo di un minuto sono stati conteggiati i seguenti raggi cosmici:

conteggio:	0	1	2	3	4	5	6	7	8 o più
frequenze:	7	17	29	20	16	8	1	2	0

Verificare se questa distribuzione empirica segue una legge di Poisson ad un livello $\alpha = 0.01$.

Tavole numeriche

La funzione di ripartizione della legge $\mathcal{N}(0, 1)$

$$\phi_\alpha$$

x	0	1	2	3	4	5	6	7	8	9
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

I quantili di uso più frequente:

$$\phi_{0.95} = 1.644854 \quad \phi_{0.975} = 1.959964$$

I quantili delle leggi $t(n)$ di Student

$$t_\alpha(n)$$

	0.95	0.975	0.99	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 $\chi^2(n)$

$$\chi_\alpha(n)$$

	0.01	0.025	0.05	0.95	0.975	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:

$$\chi_\alpha^2(n) \approx \frac{1}{2} \left(\phi_\alpha + \sqrt{2n-1} \right)^2$$