# Appunti dalle lezioni

per la parte di Statistica del corso di Matematica per BGA*

T. Ricciardi

versione 20 gennaio 2009

## Introduzione alla Statistica

La *Statistica* ha come principali obbiettivi la raccolta e la classificazione di dati, e la deduzione rigorosa di conclusioni a partire da questi. "Statistica" deriva da "stato": originariamente si trattava di dati di interesse per lo stato (ad es., numero di abitanti, nascite, decessi, ecc.).

Cominceremo con la *Statistica Descrittiva*, che si prefigge di illustrare e sintetizzare dati. Impareremo quindi a rappresentare dati mediante opportuni diagrammi (a punti, a barre, a torta,...) Impareremo poi ad estrarre da gruppi di dati alcune principali informazioni calcolando quantità quali media, moda, mediana, varianza, deviazione standard,...

Successivamente, studieremo qualche argomento di *Statistica Inferenziale*, che si occupa invece di trarre conclusioni dai dati.

Poiché il principale strumento matematico della Statistica Inferenziale è il *Calcolo delle Probabilità*, studieremo alcuni argomenti di base della relativa teoria.

Lo schema di questi appunti è pertanto il seguente:

**1.** Statistica descrittiva: organizzazione e sintesi di dati

**2.** Elementi di calcolo delle probabilità: concetto frequentistico di probabilità, calcolo combinatorio, eventi mutuamente esclusivi ed eventi indipendenti, probabilità condizionata, applicazioni alla genetica;

**3.** Statistica inferenziale: stima di parametri, test di ipotesi.

*Biologia Generale e Applicata. Queste note sono scaricabili dalla mia pagina web di didattica, raggiungibile dal sito www.docenti.unina.it.

# 1 Statistica descrittiva

## 1.1 Rappresentazione di dati

Cominciamo con un semplice esempio.

**Esempio 1.** *In una classe di 15 alunni i voti riportati ad un compito sono i seguenti: 6,5,5,8,6,6,4,8,7,6,7,4,9,7,8.*

**Definizione 1.** *La* frequenza *di un valore assunto da un insieme di dati è il numero di volte che questo valore compare nell'insieme di dati.*

Nell'esempio 1 qual è la frequenza del voto 7? (Risposta: 3).
Possiamo riportare una tabella delle frequenze. Si veda la Fig. 1.
Tracciamo un grafico a barre per le frequenze dell'esempio 1 (Fig. 2).

**Definizione 2.** *Supponiamo di avere $n$ dati e supponiamo che $f$ sia la frequenza di un certo valore assunto dai dati. Allora, si definisce* frequenza relativa *di tale dato il rapporto $f/n$.*

Per distinguerla dalla frequenza relativa, la frequenza $f$ è a volte detta anche *frequenza assoluta.*

**Esercizio 1.** *Riportare in una tabella le frequenze relative per l'esempio 1.*

Le frequenze relative sono particolarmente utili quando il numero di dati è grande. Osserviamo infatti che da $0 \leq f \leq n$ segue che $0 \leq f/n \leq 1$. Inoltre, la considerazione delle frequenze relative permette di confrontare campioni di taglie diverse.

I *diagrammi a torta* sono utili quando i dati sono di tipo categorico. Essi evidenziano le proporzioni fra le varie categorie.

**Esempio 2.** *Un'indagine sugli ultimi 200 studenti iscritti al primo anno di un certo corso di studi universitari rivela che 51 hanno frequentato il liceo classico, 93 il liceo scientifico, 10 l'istituto magistrale, mentre i rimanenti hanno frequentato altri istituti. Rappresentiamo questi dati mediante un diagramma a torta (Fig. 3).*

## 1.2 Le 3M della statistica descrittiva: media, mediana, moda

Anche se per motivi di praticità (ovvero per poter lavorare "con carta e matita") abbiamo considerato esempi con pochi dati, generalmente il numero di dati da analizzare è molto elevato, ed è pertanto importante individuare alcune grandezze che li sintetizzino. Le statistiche media, mediana e moda si usano per descrivere la "tendenza centrale" di un insieme di dati.

**Media**

**Definizione 3.** *Siano* $x_1, x_2, \ldots, x_n$ *n dati. La* media *di questi valori è data da:*

$$\overline{x} = \frac{1}{n}(x_1 + x_2 + \ldots + x_n) = \frac{1}{n}\sum_{i=1}^{n} x_i.$$

**Esempio 3.** *Calcoliamo la media dei voti riportati dagli alunni nell'esempio 1:*

$$\overline{x} = \frac{1}{15}(6+5+5+8+6+6+4+8+7+6+7+4+9+7+8) = \frac{96}{15} = 6,4.$$

È utile osservare che è più facile calcolare la media in termini di frequenze. Infatti, siano $v_1, v_2, \ldots, v_k$ i valori assunti dai dati. Sia $f_i$ la frequenza del valore $v_i$, $i = 1, \ldots, k$. Allora risulta

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{k} f_i v_i. \tag{1}$$

Verifichiamo la (1) nel caso dell'esempio 1. In questo caso, $n = 15$, $v_1 = 4$, $f_1 = 2$, $v_2 = 5$, $f_2 = 2$, $v_3 = 6$, $f_3 = 4$, $v_4 = 7$, $f_4 = 3$, $v_5 = 8$, $f_5 = 3$, $v_6 = 9$, $f_6 = 1$. Applicando la (1) troviamo che

$$\overline{x} = \frac{1}{15}(4\cdot2+5\cdot2+6\cdot4+7\cdot3+8\cdot3+9\cdot1) = \frac{1}{15}(8+10+24+21+24+9) = \frac{96}{15},$$

in accordo con quanto ottenuto precedentemente.

**Mediana**

**Definizione 4.** *Supponiamo di avere n dati numerici* $x_1, x_2, \ldots, x_n$*. Supponiamo che siano rappresentati in ordine crescente, ovvero che*

$$x_1 \leq x_2 \leq \ldots \leq x_n.$$

*Allora, la* mediana *è data da:*

$$\begin{cases} x_{\frac{n+1}{2}}, & \text{se } n \text{ è dispari} \\ \frac{1}{2}\left(x_{\frac{n}{2}} + x_{\frac{n}{2}+1}\right), & \text{se } n \text{ è pari.} \end{cases}$$

**Esempio 4.** *Calcoliamo la mediana per i dati dell'esempio 1. Riordiniamo i dati: 4,4,5,5,6,6,6,6,7,7,7,8,8,8,9. Poiché* $n = 15$ *è dispari, dalla Definizione 4 concludiamo che:*

$$mediana = x_{\frac{15+1}{2}} = x_8 = 6.$$

**Moda**

**Definizione 5.** *La* moda *di un insieme di dati, se esiste, è l'unico valore che ha frequenza massima. Se non vi è un solo valore con frequenza massima, ciasuno di essi è detto* valore modale.

Per esempio, nel caso degli alunni, esempio 1, abbiamo che la moda esiste ed è data dal valore 6.

**Esercizio 2.** *Supponiamo che le frequenze delle età dei membri di una orchestra sinfonica giovanile siano date dalla seguente tabella: 15,16,17,18,19,20; 2,5,11,9,14,13. Calcolare la media, la mediana e, se esiste, la moda di questi dati. (Risposta: media≈18,24; mediana=18,5; moda=19).*

## 1.3 Varianza e deviazione standard

Introduciamo adesso delle statistiche che misurano la "dispersione" dei dati intorno al valore medio.

**Definizione 6.** *Sia $x_1, x_2, \ldots, x_n$ un insieme di dati. Si dice* varianza *la quantità*

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2,$$

*dove $\overline{x}$ denota la media.*

Osserviamo che $s^2$ si puo' interpretare come "media" dei discostamenti al quadrato dalla media.

**Esempio 5.** *Calcolare la varianza dei due seguenti insiemi: $A:\ 3,4,6,7,10$; $B:\ -20,5,15,24$. (Risposte: A: media=6, varianza=7,5; B: media=6, varianza≈360,67)*

La seguente identità aiuta il calcolo della varianza:

**Proposizione 1.** *Sia dato un insieme di dati $x_1, x_2, \ldots, x_n$. Sia $\overline{x}$ la media. Allora, risulta:*

$$\sum_{i=1}^{n}(x_i - \overline{x})^2 = \sum_{i=1}^{n} x_i^2 - n\overline{x}^2. \tag{2}$$

*Dimostrazione.* Per definizione di $\overline{x}$, abbiamo che

$$\sum_{i=1}^{n} x_i = n\overline{x}.$$

Osserviamo inoltre che
$$\sum_{i=1}^{n} \overline{x} = \overbrace{\overline{x} + \overline{x} + \ldots + \overline{x}}^{n \text{ volte}} = n\overline{x}.$$
Pertanto, sviluppando il quadrato e tenendo conto delle identità appena dimostrate, abbiamo
$$\begin{aligned}\sum_{i=1}^{n}(x_i - \overline{x})^2 &= \sum_{i=1}^{n}(x_i^2 - 2x_i\overline{x} + \overline{x}^2) = \sum_{i=1}^{n} x_i^2 - 2\overline{x}\sum_{i=1}^{n} x_i + \sum_{i=1}^{n} \overline{x}^2 \\ &= \sum_{i=1}^{n} x_i^2 - 2n\overline{x}^2 + n\overline{x}^2 = \sum_{i=1}^{n} x_i^2 - n\overline{x}^2.\end{aligned}$$
□

Tenendo conto della Proposizione 1, concludiamo che la varianza di un insieme di dati si può equivalentemente calcolare usando la seguente formula:
$$s^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} x_i^2 - n\overline{x}^2\right). \tag{3}$$
**Esercizio 3.** *Calcolare la varianza per l'insieme A definito nell'esempio 5 usando la formula* (3).

# 2 Fondamenti di calcolo delle probabilità

Lanciando una moneta, uscirà testa o croce? Lanciando un dado, quale numero uscirà? Scegliendo a caso un individuo di una classe di ragazzi, che altezza avrà? È chiaro che domande di questo tipo non ammettono una risposta certa. È però altrettanto chiaro che lanciando più volte una moneta, uscirà testa un numero di volte non troppo diverso dal numero di volte che esce croce (se la moneta non è truccata).

Vogliamo in questo paragrafo dare un fondamento matematico alle situazioni appena descritte. Vedremo come definire una *probabilità* degli eventi "esce croce", "non esce 5".

Una probabilità è un "peso" compreso tra 0 e 1 che associamo ad un evento. Un evento impossibile (ad esempio "non esce né testa né croce") ha probabilità 0. L'evento certo (ad esempio, "esce o testa o croce") ha probabilità 1. In generale, un evento è tanto più probabile quanto più vicina a 1 è la sua probabilità.

Nelle applicazioni, il principale concetto di probabilità usato è quello frequentistico, dato dal rapporto di casi favorevoli su casi possibili, come descritto nel seguente paragrafo.

## 2.1 Concetto frequentistico di probabilità

Il seguente principio è alla base del concetto frequentistico di probabilità.

**Principio 1** (Frequenza empirica)**.** *Se abbiamo $n$ risultati possibili ed egualmente probabili, se $s$ di questi risultati sono successi e $i = n - s$ sono insuccessi, allora la probabilità di successo è data da:*

$$p = \frac{s}{n}$$

*e la probabilità di insuccesso è data da*

$$q = \frac{i}{n} = \frac{n-s}{n} = 1 - \frac{s}{n} = 1 - p.$$

Questo principio riflette l'osservazione empirica che dati $n$ risultati equiprobabili di un esperimento, in un numero elevato di prove ripetute il rapporto tra il numero di volte che si osserva un particolare risultato ed il numero totale di prove si avvicina a $1/n$.

**Esercizio 4** (Lancio di moneta)**.** *Si lancia una moneta (non truccata). Qual è la probabilità che esca testa?*

*Soluzione.* Abbiamo due risultati possibili, testa e croce. Testa è un successo, croce è un insuccesso. Dunque, la probabilità che esca testa è $1/2$ e la probabilità che esca croce è $1/2$. □

**Esercizio 5.** *Si estrae una carta da un mazzo regolare. Calcolare le probabilità dei seguenti eventi:*

*(a) La carta è rossa.*

*(b) La carta è di picche.*

*(c) La carta è un re.*

*(d) la carta* non *è l'asso di cuori.*

[Risposte: (a) 1/2; (b) 1/4; (c) 1/13; (d) 51/52]

**Esercizio 6.** *Si estrae una pallina da un sacco contenente 3 palline bianche, 4 palline rosse, 5 palline nere. Calcolare le probabilità dei seguenti eventi:*

*(a) La pallina è bianca;*

*(b) La pallina è bianca o rossa;*

*(c) La pallina non è rossa.*

[Risposte: (a) 1/4; (b) 7/12; (c) 2/3]

**Esercizio 7.** *Si lanciano due dadi. Calcolare le probabilità dei seguenti eventi:*

*(a) Il totale è 7;*

*(b) Il totale è 8;*

*(c) Il totale è maggiore di 10;*

*(d) Esce lo stesso numero su entrambi i dadi.*

*Soluzione.* I possibili risultati dell'esperimento sono le coppie di numeri interi compresi tra 1 e 6:

$$\{(h, k) : h, k \in \{1, 2, 3, 4, 5, 6\}\}.$$

Abbiamo quindi 36 risultati possibili.
(a). I successi sono dati da: (1,6), (2,5), (3,4), (4,3), (5,2), (6,1)}. Quindi $p = 6/36 = 1/6$.
(b) I successi sono dati da: (2,6), (3,5), (4,4), (5,3), (6,2). Quindi $p = 5/36$.
(c) I successi sono: (5,6), (6,6), (6,5). Quindi $p = 3/36 = 1/12$.
(d) I successi sono: (1,1), (2,2), (3,3), (4,4), (5,5), (6,6). Quindi $p = 1/6$. □

## 2.2 Un po' di calcolo combinatorio

Per situazioni appena più complesse di quelle fin qui considerate abbiamo bisogno di un po' di "calcolo combinatorio", che ci aiuta a "contare" i casi favorevoli. In particolare, introduciamo i concetti di *permutazione*, *combinazione*, *fattoriale* e *coefficiente binomiale*. Alla base delle considerazioni contenute in questo paragrafo è il seguente

**Principio 2** (Enumerazione)**.** *Supponiamo che un primo evento si può verificare in $n_1$ modi diversi; dopo di questo un secondo evento si può verificare in $n_2$ modi diversi; dopo di questo un terzo evento si può verificare in $n_3$ modi diversi; . . . Allora questi eventi si possono verificare nell'ordine suddetto in $n_1 \cdot n_2 \cdot n_3 \cdot \ldots$ modi diversi.*

## Permutazioni

**Definizione 7.** *Una* permutazione *di un insieme di* $n$ *oggetti è un elenco ordinato degli elementi dell'insieme.*

**Esempio** *abc*, *bac*, *cba* sono permutazioni dell'insieme $\{a, b, c\}$.

Vogliamo contare le possibili permutazioni di un insieme di $n$ oggetti (distinti). Il seguente principio è fondamentale e sarà usato più volte.

**Proposizione 2.** *Il numero di permutazioni di un insieme di* $n$ *oggetti è dato da*

$$1 \cdot 2 \cdot 3 \cdot \ldots \cdot (n-1) \cdot n,$$

*ovvero dal prodotto degli interi compresi tra 1 ed* $n$.

*Dimostrazione.* Possiamo scegliere il primo elemento dell'elenco in $n$ modi, il secondo in $n-1$ modi, il terzo in $n-3$ modi, e così via. Per il Principio 2 il numero di modi in cui possiamo elencare $n$ oggetti distinti è $n!$. □

**Notazione.** Si denota

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

ed il simbolo $n!$ si legge "$n$ fattoriale". Per convenzione, si definisce $0! = 1$.

**Esercizio 8.** *Calcolare: 3!; 4!; 5!*

[Risposte: 6; 24; 120].

**Esercizio 9.** *Calcolare: 7!/6!; (r+1)!/(r-1)! dove* $r \in N, r \geq 1$.

[Risposte: 7; r(r+1)].

## Combinazioni

**Definizione 8.** *Sia* $r \leq n$. *Una* combinazione *di* $n$ *elementi presi* $r$ *alla volta è un sottoinsieme di* $r$ *elementi di un insieme contenente* $n$ *elementi.*

Ad esempio, le combinazioni delle 4 lettere $a, b, c, d$ prese 3 alla volta sono: *abc*, *abd*, *abd*, *bcd*.

**Proposizione 3.** *Il numero di combinazioni di* $n$ *elementi presi* $r$ *alla volta è:*

$$\frac{n!}{r!(n-r)!}.$$

*Dimostrazione.* Calcoliamo prima il numero di modi di elencare $r$ elementi scelti tra $n$ in un fissato ordine. Il primo elemento può essere scelto in $n$ modi. Il secondo in $n-1$ modi. Il terzo in $n-2$ modi. E così via, fino ad arrivare all'$r$-mo elemento che può essere scelto in $n-r+1$ modi. Per il Principio 2, il numero di modi di elencare $r$ elementi scelti tra $n$ in un fissato ordine è dato da

$$n(n-1)(n-2)\dots(n-r+1) = \frac{n!}{(n-r)!}.$$

Poichè una combinazione di $n$ oggetti presi $r$ alla volta *non* dipende dal particolare ordine degli $r$ oggetti scelti, per ottenere il numero di combinazioni cercato dobbiamo dividere questo numero per il numero di permutazioni di $r$ oggetti, cioè per $r!$. In conclusione, il numero di combinazioni di $n$ oggetti presi $r$ alla volta è dato da

$$\frac{n!}{r!(n-r)!}.$$

□

**Notazione.** Si denota

$$\binom{n}{r} = \frac{n!}{r!(n-r)!}$$

Il simbolo $\binom{n}{r}$ è detto *coefficiente binomiale*, e si legge "$n$ su $r$".

Si verifica facilmente che

$$\binom{n}{r} = \binom{n}{n-r}.$$

**Esercizio 10.** *Ad un bambino si dà il permesso di scegliere 3 giocattoli da uno scaffale che ne contiene 12. In quanti modi può farlo?*

[Risposta: In $\binom{12}{3} = 220$ modi].

**Esercizio 11.** *Una signora offre una cena a 6 ospiti. In quanti modi può sceglierli tra 10 amici?*

[Risposta: In $\binom{10}{6} = 210$ modi].

I coefficienti binomiali sono utili anche per sviluppare l'espressione $(a+b)^n$. Vale infatti:

**Teorema 1** (Teorema binomiale)**.** *Risulta*

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}$$

*per ogni* $a, b \in \mathbb{R}$ *e per ogni* $n \in \mathbb{N}$.

*Dimostrazione.* Scriviamo:

$$(a+b)^n = (a+b)(a+b)\dots(a+b).$$

Otteniamo il termine $a^k b^{n-k}$ quando da ciascuna parentesi scegliamo $k$ volte il termine $a$ e $n-k$ volte il termine $b$. Questa scelta si può fare in $\binom{n}{k}$ modi diversi. $\square$

**Esercizio 12.** *Verificare il Teorema 1 per* $n = 1, 2, 3$.

**Esercizio 13.** *Un sacchetto contiene 2 palline bianche e 3 palline nere. Si estrae una pallina 5 volte, rimpiazzandola ogni volta. Calcolare le probabilità dei seguenti eventi:*

*(a) Le prime 4 sono bianche e l'ultima è nera.*

*(b) Esce una pallina bianca esattamente 4 volte.*

*(c) Esce una pallina bianca almeno una volta.*

*Soluzione.* (a). Contiamo prima in quante volte possiamo estrarre palline. Alla prima estrazione abbiamo 5 possibilità. Alla seconda 5, alla terza ancora 5, ecc. Per il Principio 2, abbiamo quindi $5^5 = 3125$ possibilità. Contiamo adesso in quanti modi possiamo ottenere bianco le prime 4 volte e nero l'ultima. Alla prima estrazione abbiamo 2 possibilità. Alla seconda, terza, quarta, abbiamo sempre 2 possibilità. Alla quinta estrazione abbiamo 3 possibilità. Per il Principio 2, abbiamo quindi $2^4 \cdot 3 = 48$ possibilità. In conclusione, la probabilità cercata è $48/3125$.
(b). Analogamente a quanto visto in (a), la probabilità di estrarre quattro palline bianche in quattro particolari estrazioni, e una pallina nera all'estrazione rimanente è $48/3125$. Dobbiamo moltiplicare questo numero per il numero di modi in cui possiamo scegliere quattro estrazioni tra cinque, cioè per $\binom{5}{4} = \binom{5}{1} = 5$. La probabilità cercata è $5 \cdot 48/3125 = 48/625$.
(c). L'evento "esce bianco almeno una volta" è complementare all'evento "esce 5 volte nero". Analogamente a quanto fatto al punto (a), abbiamo che la probabilità di estrarre 5 volte nero è $3^5/5^5 = 243/3125$. Poichè la somma delle probabilità di due eventi complementari è 1, concludiamo che la probabilità di estrarre bianco almeno una volta è $1 - 243/3125 = 2882/3125$. $\square$

**Osservazione** Nell'esercizio 13-(a) abbiamo calcolato la probabilità come rapporto dei $2^4 \cdot 3 = 48$ successi e dei $5^5 = 3125$ casi possibili. Osserviamo che

$$\frac{2^4 \cdot 3}{5^5} = \left(\frac{2}{5}\right)^4 \cdot \left(\frac{3}{5}\right).$$

I punti (a)–(b) dell'esercizio 13 sono quindi casi particolari della seguente proprietà generale.

**Proposizione 4.** *Sia $p$ la probabilità di successo di un esperimento. Ripetiamo l'esperimento $n$ volte.*

*(i) La probabilità di ottenere $k$ successi e $n-k$ insuccessi in un determinato ordine è $p^k(1-p)^{n-k}$.*

*(ii) La probabilità di ottenere $k$ successi e $n-k$ insuccessi in un ordine qualunque è $\binom{n}{k}p^k(1-p)^{n-k}$.*

## 2.3 Eventi mutuamente esclusivi ed eventi indipendenti

Nella pratica ci si trova spesso a considerare eventi costituiti da "combinazioni" di eventi più semplici. Le combinazioni più semplici di eventi si ottengono mediante le operazioni logiche di "oppure" ed "e". In questi casi, due regole fondamentali del calcolo delle probabilità affermano che le probabilità dei corrispondenti eventi composti sono dati dalla somma e dal prodotto delle probabilità dei singoli eventi, rispettivamente. Precisiamo qui di seguito questi concetti.

**Definizione 9.** *Due eventi sono detti* mutuamente esclusivi *se il verificarsi dell'uno impedisce il verificarsi dell'altro.*

Ad esempio, la comparsa di un jack e di una regina in una estrazione da un mazzo regolare di carte sono eventi mutuamente esclusivi. Al contrario, la comparsa di un jack e di una carta di picche non sono eventi mutuamente esclusivi (perché se si estrae il jack di picche si verificano entrambi).

**Regola 1** (della "somma" o dell'"oppure")**.** *La probabilità che si verifichi uno tra due eventi mutuamente esclusivi è la somma delle probabilità dei singoli eventi.*

Verifichiamo la Regola 1 nel semplice caso di lancio di un dado. Gli eventi $A =$"esce 1" e $B =$"esce 2" sono mutuamente esclusivi. Osserviamo che, usando il ragionamento frequentistico del Principio 1, otteniamo che $P(A) = 1/6$, $P(B) = 1/6$ e $P($"esce 1 oppure 2"$) = 1/3 = P(A) + P(B)$, in accordo con quanto affermato nella Regola 1.

**Definizione 10.** *Due eventi sono detti* indipendenti *se il verificarsi dell'uno non influenza il verificarsi dell'altro.*

Ad esempio, in un lancio di due dadi, il risultato del primo dado non condiziona il risultato del secondo. Invece, se estraiamo due carte da un mazzo, senza rimettere nel mazzo la prima carta estratta, la probabilità che la seconda carta sia rossa dipende dall'essere rossa o meno la prima carta.

**Regola 2** (del "prodotto" o dell'"e"). *La probabilità che si verifichino due eventi indipendenti in successione o contemporaneamente è data dal prodotto delle probabilità dei singoli eventi.*

Verifichiamo la Regola 2 nel caso del lancio di due dadi. Sia $A =$"esce 1 al primo lancio" e $B =$"esce 1 al secondo lancio". Usando il ragionamento frequentistico concludiamo facilmente che $P(A) = 1/6$, $P(B) = 1/6$ e $P($"esce 1 al primo lancio e 1 al secondo lancio"$) = 1/36 = P(A) \cdot P(B)$, in accordo con quanto affermato nella Regola 2.

**Notazione.** Siano $A$ e $B$ eventi. Per ragioni che saranno chiarite nel paragrafo 2.7, denotiamo l'evento "si verifica $A$ oppure $B$" con $A \cup B$; denotiamo l'evento "si verificano $A$ e $B$" con $A \cap B$.

Mediante tale notazione, la Regola 1 afferma che se $A$ e $B$ sono eventi mutuamente esclusivi, allora $P(A \cup B) = P(A) + P(B)$; la Regola 2 afferma che se $A$ e $B$ sono eventi indipendenti, allora $P(A \cap B) = P(A) \cdot P(B)$.

**Esercizio** La fibrosi cistica (CF) è la malattia genetica recessiva più frequente tra gli Europei. Circa una persona su 2500 è portatrice di almeno una copia del gene CF, che si presenta con uguale frequenza negli uomini e nelle donne. Supponiamo che due genitori siano entrambi portatori sani del gene CF. Allora, per le leggi di Mendel, la probabilità di avere un figlio normale è 1/4; la probabilità di avere un figlio portatore sano è 1/2; la probabilità di avere un figlio affetto da CF è 1/4. Calcolare le seguenti probabilità di avere:

(a) due figli (di sesso qualunque) non portatori del gene CF;

(b) un figlio maschio portatore sano;

(c) due figlie femmine, una portatrice sana ed una affetta da CF;

(d) due figlie femmine affette da CF.

**Esercizio** Per studiare una popolazione di tritoni *(Triturus cristatus)*, 150 tritoni vengono marcati e rimessi in un lago. Una settimana dopo vengono ripescati 351 tritoni, e si rileva che 54 sono marcati. Stimare il numero totale di tritoni presenti nel lago. [R: 975]

**Esercizio** Un'indagine rileva che 19 su 60 uomini sono fumatori e che 12 su 40 donne sono fumatrici. Calcolare le seguenti probabilità:

(a) un individuo scelto a caso sia un uomo fumatore;

(b) un individuo scelto a caso sia un fumatore;

(c) un uomo scelto a caso sia fumatore;

(d) un fumatore scelto a caso sia uomo.

## 2.4 Probabilità condizionata

Ci occupiamo in questo paragrafo di calcolare la probabilità che si verifichi un evento $A$, sapendo che si è verificato un evento $B$. Tale probabilità è detta probabilità di $A$ condizionata dall'evento $B$ e si denota con $P(A|B)$.

Ad esempio, supponiamo che un gruppo d persone consista di 5 donne e di 5 uomini, e che sono fumatori 2 uomini e 1 donna. Ragionando frequentisticamente, calcoliamo facilmente le probabilità dei seguenti eventi:

(a) un individuo scelto a caso sia una donna fumatrice. $P(\text{donna fumatrice}) = 1/10$

(b) una donna scelta a caso sia fumatrice. $P(\text{fumatrice}|\text{donna}) = 1/5$

(c) un individuo scelto a caso sia donna. $P(\text{donna}) = 1/2$.

Osserviamo adesso che possiamo scrivere:

$$P(\text{fumatrice}|\text{donna}) = \frac{P(\text{donna fumatrice})}{P(\text{donna})}$$

In generale, diamo la seguente definizione.

**Definizione 11.** *Siano $A$ e $B$ due eventi, con $P(B) \neq 0$. La* probabilità di $A$ condizionata da $B$ *è per definizione:*

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

**Esercizio** Si lancia un dado. Esce un numero pari. Calcolare la probabilità che tale numero sia $\geq 4$, sia ragionando frequentisticamente, sia usando la definizione di probabilità condizionata.

*Soluzione.* I numeri pari di un dado sono 2, 4, 6. I numeri pari di un dado che sono anche $\geq 4$ sono 4 e 6. Dunque, la probabilità richiesta è 2/3.

Alternativamente:

$$P(\geq 4|\text{pari}) = \frac{P(\geq 4 \cap \text{pari})}{P(\text{pari})} = \frac{2/6}{3/6} = \frac{2}{3}.$$

Il seguente risultato sarà utilizzato nel Paragrafo 2.5, per dimostrare il Principio di Hardy-Weinberg in genetica.

**Teorema 2** (delle alternative)**.** *Siano $A_i$, $i = 1, 2, \dots, k$ eventi mutuamente esclusivi ed esaustivi, cioè tali che $A_i$ e $A_j$ sono a due a due mutuamente esclusivi per ogni $i, j = 1, 2, \dots, k$ e tali che $A_1 \cup A_2 \cup \dots \cup A_k = \Omega$, dove $\Omega$ è l'insieme di tutti gli esiti possibili. Allora, per ogni evento $A$ risulta:*

$$P(A) = \sum_{i=1}^{k} P(A|A_i)P(A_i).$$

Gli eventi $A_i$, $i = 1, 2, \dots, k$ nel Teorema 2 costituiscono un "sistema completo di alternative", da cui il nome del teorema.

Il seguente risultato, detto Legge di Bayes, sarà utilizzato nel Paragrafo 2.6 per verificare l'attendibilità di un test diagnostico. Tale risultato consente di calcolare $P(B|A)$ a partire da $P(A|B)$, $P(A)$ e $P(B)$. Interpretando $A$ come causa e $B$ come effetto, esso fornisce la"probabilità di una causa".

**Teorema 3** (Legge di Bayes)**.** *Risulta:*

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

*Dimostrazione.* Per definizione, abbiamo

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

Poiché $P(B \cap A) = P(A \cap B)$, possiamo scrivere:

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

□

## 2.5 Applicazione del teorema delle alternative: Il principio di Hardy-Weinberg

È noto che alcune caratteristiche di specie viventi sono determinate dai geni presenti sui cromosomi. Negli organismi "diploidi", tra cui gli esseri umani, i cromosomi si presentano in coppie. Un gene su un cromosoma ha quindi un suo gene corrispondente sul cromosoma "omologo". In generale un gene si presenta in più forme, dette "alleli". Supponiamo che un gene abbia due alleli $A$ e $a$. Per ogni coppia di cromosomi omologhi abbiamo quindi 3 combinazioni di alleli: $AA$, $Aa$ e $aa$ (genotipi). Un figlio riceve un gene dal padre e un gene dalla madre. La probabilità di ricevere dal padre il primo gene è uguale alla probabilità di ricevere in alternativa il secondo. Lo stesso vale per il gene ricevuto dalla madre. Supponiamo che le probabilità che un genitore appartenga ai genotipi $AA$, $Aa$ e $aa$ siano $u > 0$, $2v > 0$ e $w > 0$, rispettivamente. Poichè gli eventi "essere $AA$", "essere $Aa$" e "essere $aa$" sono mutuamente esclusivi ed esauriscono tutte le possibilità, la somma delle loro singole probabilità deve essere uguale alla probabilità dell'evento certo. Quindi deve essere $u + 2v + w = 1$. Supponiamo adesso che padre e madre vengano scelti casualmente. Per la Regola 2, risulta che:

la probabilità che padre sia $AA$ e madre sia $AA$ è $u^2$;
la probabilità che padre sia $AA$ e madre sia $Aa$ è $2uv$;
la probabilità che padre sia $Aa$ e madre sia $AA$ è $2uv$;
la probabilità che padre sia $Aa$ e madre sia $Aa$ è $4v^2$.

Risulta inoltre che:

la probabilità che da padre $AA$ e madre $AA$ nasca un figlio $AA$ è 1;
la probabilità che da padre $AA$ e madre $Aa$ nasca un figlio $AA$ è 1/2;
la probabilità che da padre $Aa$ e madre $AA$ nasca un figlio $AA$ è 1/2;
la probabilità che da padre $Aa$ e madre $Aa$ nasca un figlio $AA$ è 1/4.

Sia $P_1(AA)$ la probabilità che un figlio sia $AA$. Osserviamo che, in una scelta casuale di due genitori, gli eventi "padre $AA$ e madre $AA$", "padre $AA$ e madre $Aa$", "padre $Aa$ e madre $AA$", "padre $Aa$ e madre $Aa$" e "padre aa" e "madre aa" sono mutuamente esclusivi ed esaustivi. Indicato con $P_1(AA)$ la probabilità che un figlio sia $AA$, per il Teorema 2 concludiamo che:

$$\begin{aligned} P_1(AA) =& P_1(AA|\mathrm{p}\,AA \cap \mathrm{m}\,AA)P(\mathrm{p}\,AA \cap \mathrm{m}\,AA) \\ &+ P_1(AA|\mathrm{p}\,AA \cap \mathrm{m}\,Aa)P(\mathrm{p}\,AA \cap \mathrm{m}\,Aa) \\ &+ P_1(AA|\mathrm{p}\,Aa \cap \mathrm{m}\,AA)P(\mathrm{p}\,Aa \cap \mathrm{m}\,AA) \\ &+ P_1(AA|\mathrm{p}\,Aa \cap \mathrm{m}\,Aa)P(\mathrm{p}\,Aa \cap \mathrm{m}\,Aa) \\ &+ P_1(AA|\mathrm{p}\,aa)P(\mathrm{p}\,aa) + P_1(AA|\mathrm{m}\,aa)P(\mathrm{m}\,aa), \end{aligned}$$

dove con le notazioni p $AA$, m $AA$,. . . denotiamo gli eventi "padre $aa$", "madre aa". . . Poiché $P_1(AA|\mathrm{p}\,aa) = 0$ e $P_1(AA|\mathrm{m}\,aa) = 0$, concludiamo che

$$P_1(AA) =1 \cdot u^2 + \frac{1}{2} \cdot 2uv + \frac{1}{2} \cdot 2uv + \frac{1}{4} \cdot 4v^2 = u^2 + 2uv + v^2 = (u + v)^2.$$

Per simmetria (o ragionando analogamente), indicato con $P_1(aa)$ la probabilità che un figlio sia $aa$, otteniamo:

$$P_1(aa) = (w + v)^2.$$

Indichiamo con $P_1(Aa)$ la probabilità che un figlio sia $Aa$. Poichè deve essere $P_1(AA) + P_1(Aa) + P_1(aa) = 1$, concludiamo che:

$$P_1(Aa) = 1 - P_1(AA) - P_1(aa) = 1 - (u + v)^2 - (w + v)^2.$$

Ricordando che $u + 2v + w = 1$ possiamo semplificare la formula precedente:

$$P_1(Aa) = [(u + v) + (w + v)]^2 - (u + v)^2 - (w + v)^2 = 2(u + v)(w + v).$$

A questo punto poniamo $P_1(AA) = u_1$, $P_1(Aa) = 2v_1$ e $P_1(aa) = w_1$ e ripetiamo il ragionamento per la seconda generazione. Indicati con $u_2, 2v_2$ e $w_2$ le probabilità che una persona della seconda generazione sia $AA$, $Aa$ e $aa$ rispettivamente, troviamo che:

$$u_2 = (u_1 + v_1)^2, \qquad v_2 = (u_1 + v_1)(w_1 + v_1), \qquad w_2 = (v_1 + w_1)^2.$$

Sostituendo in questa formula $u_1 = (u + v)^2$, $v_1 = (u + v)(w + v)$ e $w_1 = (w + v)^2$, troviamo che

$$u_2 = (u_1+v_1)^2 = [(u+v)^2+(u+v)(w+v)]^2 = (u+v)^2(u+2v+w) = (u+v)^2 = u_1.$$

Analogamente, troviamo $w_2 = w_1$ e quindi $v_2 = v_1$. Cioè, otteniamo il seguente notevole risultato, dovuto al matematico G.H. Hardy ed al biologo W. Weinberg nel 1908: *la distribuzione dei genotipi non cambia nel corso delle successive generazioni.* Osserviamo che questa schematizzazione assume scelte casuali di padri e madri e non tiene conto di mutazioni, immigrazioni, ecc.

**Osservazione** Alla base del principio appena descritto vi è la seguente semplice proprietà algebrica: se $u, v, w \geq 0$ sono tali che $u+2v+w = 1$, posto $u_1 = (u + v)^2$, $v_1 = (u + v)(w + v)$, $w_1 = (w + v)^2$, risulta $u_1 + 2v_2 + w_1 = 1$, $(u_1+v_1)^2 = (u+v)^2$, $(u_1+v_1)(w_1+v_1) = (u+v)(w+v)$, $(w_1+v_1)^2 = (w+v)^2$.

## 2.6 Applicazione della Legge di Bayes: test diagnostici

Un test diagnostico consiste generalmente nella verifica di opportune reazioni biochimiche su un campione sanguigno del paziente, allo scopo di stabilire rapidamente se vi è una significativa probabilità che il paziente sia affetto da una certa malattia. Poiché nessun test è perfetto, vi saranno sempre delle persone malate sulle quali il test dà esito negativo, e delle persone sane sulle quali il test dà esito positivo. È pertanto di fondamentale importanza stimare il *valore predittivo del test*, ovvero la probabilità che una persona che risulta positiva al test sia effettivamente malata. Dal punto di vista sperimentale, risulta però più immediato sottoporre al test un grande numero di persone che si sa per altre via essere malate, e verificare quale percentuale di esse risulta positiva al test; e analogamente, sottoporre al test un grande numero di persone che si sa per altre via essere sane, e verificare quale percentuale di esse risulta negativa al test. Ovvero, indicando con $M, S$ gli eventi "malato" e "sano", rispettivamente, e con $T^+, T^-$ gli eventi "positivo al test" e "negativo al test", sperimentalmente si stimano le seguenti probabilità:

| | |
|---|---|
| $P(T^+\|M)$ | detta *sensibilità del test* |
| $P(T^-\|S)$ | detta *specificità del test.* |

Invece, il *valore predittivo del test* è dato da $P(M|T^+)$. Vediamo come la Legge di Bayes, Teorema 3 risulta utile per calcolarlo (con l'ausilio di almeno un altro dato). A tale scopo osserviamo che la popolazione totale risulta suddivisa in quattro categorie: $M\cap T^+$="veri positivi", ovvero malati positivi al test; $M\cap T^-$ ="falsi negativi", ovvero malati negativi al test; $S\cap T^+$ ="falsi positivi", ovvero individui sani positivi al test; $S\cap T^-$ ="veri negativi", ovvero individui sani negativi al test. La conoscenza delle quattro probabilità $P(M\cap T^+)$, $P(M\cap T^-)$, $P(S\cap T^+)$, $P(S\cap T^-)$ permetterebbe di calcolare subito $P(M|T^+)$. Infatti:

$$P(M|T^+) = \frac{P(M\cap T^+)}{P(T^+)} = \frac{P(M\cap T^+)}{P(S\cap T^+) + P(M\cap T^+)}.$$

La sola conoscenza della sensibilità e della specificità, invece, corrispondono a due equazioni lineari nelle quattro probabilità suddette. Infatti, supponiamo di sapere che la sensibilità è $\alpha$ e la specificità è $\beta$. Da

$$P(T^+|M) = \frac{P(T^+\cap M)}{P(M)} = \frac{P(T^+\cap M)}{P(M\cap T^+) + P(M\cap T^-)}$$

otteniamo

$$P(T^+\cap M) = \alpha[P(M\cap T^+) + P(M\cap T^-)]. \tag{4}$$

Similmente, otteniamo

$$P(T-+\cap S) = \beta[P(S\cap T^-) + P(S\cap T^+)]. \tag{5}$$

La normalizzazione implica inoltre che

$$P(M\cap T^+) + P(M\cap T^-) + P(S\cap T^+) + P(S\cap T^-) + = 1. \tag{6}$$

Le equazioni (4)–(5)–(6) costituiscono un sistema di tre equazioni lineari nelle quattro incognite $P(M\cap T^+)$, $P(M\cap T^-)$, $P(S\cap T^+)$, $P(S\cap T^-)$. Abbiamo bisogno quindi di un'ulteriore informazione per calcolare $P(M|T^+)$. Potrebbe essere nota, ad esempio, l'incidenza della malattia $P(M) = P(M \cap T^+) + P(M \cap T^-)$. Oppure, mediante uno screening di massa, potremmo avere una stima di $P(T^+) = P(T^+ \cap M) + P(T^+ \cap S)$. Vediamo degli esempi.

**Esempio.** Per un test diagnostico è noto che la sensibilità è del 95% e che la specificità è del 90%. Da uno screening di massa si rileva inoltre che lo 0.2 della popolazione risulta positiva al test. Determinare il valore predittivo del test.

*Soluzione.* Abbiamo le quattro equazioni:

$$\begin{cases} P(T^+ \cap M) = 0.95[P(T^+ \cap M) + P(T^- \cap M)] \\ P(T^- \cap S) = 0.9[P(T^+ \cap S) + P(T^- \cap S)] \\ P(T^+ \cap S) + P(T^+ \cap M) = 0.2 \\ P(M \cap T^+) + P(M \cap T^-) + P(S \cap T^+) + P(S \cap T^-) + = 1. \end{cases}$$

Dalle prime due equazioni ricaviamo che $P(T^+ \cap M) = 19P(T^- \cap M)$ e $P(T^- \cap S) = 9P(T^+ \cap S)$. Abbiamo inoltre che

$$P(T^- \cap S) = P(T^-) - P(T^- \cap M) = 1 - P(T^+) - P(T^- \cap M) = 0.8 - P(T^- \cap M)$$

e

$$P(T^+ \cap S) = P(T^+) - P(T^+ \cap M) = 0.2 - P(T^+ \cap M).$$

Otteniamo quindi che

$$0.8 - P(T^- \cap M) = 9[0.2 - P(T^+ \cap M)] = 1.8 - 9P(T^+ \cap M).$$

Siamo quindi ridotti al sistema di due equazioni:

$$\begin{cases} P(T^+ \cap M) = 19P(T^- \cap M) \\ 0.8 - P(T^- \cap M) = 1.8 - 9P(T^+ \cap M \end{cases}$$

da cui ricaviamo che $P(T^+ \cap M) = 0.11$ e $P(T^- \cap M) = 0.006$. In particolare,

$$P(M) = P(T^+ \cap M) + P(T^- \cap M) = 0.116$$

ed infine otteniamo il valore predittivo del test:

$$P(M|T^+) = \frac{P(T^+ \cap M)}{P(T^+)} = \frac{0.11}{0.2} = 0.55.$$

□

## 2.7 Spazi di probabilità

Motivati dagli esempi precedenti, basati sui conteggi di casi favorevoli e non, in quanto segue illustriamo brevemente la cosiddetta *teoria assiomatica della probabilità*. Tale teoria si basa su una definizione di probabilità astratta, che prescinde da concetti empirici, combinatorici, frequentistici, ecc.

Consideriamo un esperimento casuale, ossia un esperimento i cui possibili risultati sono noti, ma non sono prevedibili con certezza (lancio di moneta, estrazione di una carta da un mazzo, ecc.). Supponiamo che i possibili risultati del nostro esperimento casuale siano in numero *finito*.

Definiamo *spazio campione* l'insieme dei possibili risultati dell'esperimento. Denotiamo lo spazio campione con $\Omega$. I sottoinsiemi di $\Omega$ sono gli *eventi*. Più precisamente, identifichiamo col sottoinsieme $A \subset \Omega$ l'evento "il risultato $\omega$ dell'esperimento appartiene ad $A$". Denotiamo con $\mathcal{P}(\Omega)$ l'insieme degli eventi. Una funzione $P : \mathcal{P}(\Omega) \to [0, 1]$ è detta *distribuzione di probabilità* se verifica i seguenti assiomi:

$$P(\Omega) = 1 \tag{7}$$

$$P(A \cup B) = P(A) + P(B) \quad \text{per ogni } A, B \subset \Omega \text{ tali che } A \cap B = \emptyset. \tag{8}$$

La coppia $(\Omega, P)$ è detta *spazio di probabilità* (finito). L'assioma 7 è detto "normalizzazione"; l'assioma 8 è detto "additività".

**Esempio 1.** Riprendiamo l'esperimento del lancio di moneta dell'Esercizio 4. Abbiamo $\Omega = \{T, C\}$. Verifichiamo che i possibili eventi corrispondono a sottoinsiemi di $\Omega$. I possibili eventi sono dati da: "esce testa"$=\{T\}$, "esce croce"$=\{C\}$, "esce testa o croce"$=\{T, C\}$. Tra i sottoinsiemi di $\Omega$ risulta anche l'evento impossibile "non esce né testa né croce"$=\{\emptyset\}$. Inoltre, $P(\{T\}) = 1/2$, $P(\{C\}) = 1/2$, $P(\{T, C\}) = 1$, $P(\emptyset) = 0$.

Osserviamo esplicitamente che la Regola 2 corrisponde all'assioma 8. Inoltre, tenendo presente la Regola 2 risulta naturale la seguente definizione.

**Definizione 12.** *Due eventi $A$ e $B$ sono detti* indipendenti *se $P(A \cap B) = P(A) \cdot P(B)$.*

La Definizione 11 di probabilità condizionata vale anche in uno spazio di probabilità astratto $(\Omega, P)$. In particolare, se $A, B \subset \Omega$ sono eventi e se $P(B) \neq 0$, la probabilità di $A$ condizionata da $B$ è per definizione:

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

**Esercizio** Dimostrare che le seguenti affermazioni sono equivalenti:

(i) $A$ e $B$ sono eventi indipendenti;

(ii) $P(A|B) = P(A)$;

(ii) $P(B|A) = P(B)$.

Per semplificare la notazione, denotiamo d'ora in avanti

$$P(\{\omega\}) = P(\omega) \qquad \text{per ogni } \omega \in \Omega.$$

Con questa notazione, è chiaro che possiamo scrivere

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

La funzione $\omega \to P(\omega)$ è detta *distribuzione di probabilità*.

**Distribuzione binomiale di parametri** $(n, p)$ La distribuzione binomiale è la distribuzione di probabilità definita sull'insieme $\Omega = \{0, 1, 2, \ldots, n\}$ da

$$P(k) = \binom{n}{k} p^k (1-p)^{n-k}, \qquad k = 0, 1, 2, \ldots$$

Per quanto affermato nella Proposizione 4, la distribuzione binomiale di parametri $(n, p)$ fornisce la probabilità di ottenere esattamente $k$ successi in un esperimento di $n$ prove ripetute, dove la probabilità di successo in ogni singola prova è $p$.

Osserviamo che per il Teorema 1 risulta $P(\{0, 1, 2, \ldots, n\}) = 1$, così che l'assioma di normalizzazione (7) è verificato.

### Cenni agli spazi di probabilità numerabili

In alcuni esperimenti casuali è naturale riferirsi a spazi campione del tipo $\Omega = \mathbb{N}$, dove $\mathbb{N}$ denota l'insieme di tutti i numeri naturali. Generalizzando la definizione di spazio di probabilità finito, diciamo che $(\Omega, P)$ è uno *spazio di probabilità numerabile* se $\Omega = \mathbb{N}$ e $P$ è una funzione che ad ogni sottoinsieme $A \subset \Omega$ associa un valore $P(A) \in [0, 1]$ con le seguenti proprietà:

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

Per ogni successione di sottoinsiemi $A_i$ a due a due disgiunti,

$$P(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i).$$

(Si veda l'Appendice per la definizione di questa somma di infiniti numeri). Uno spazio di probabilità si dice *discreto* se è finito o numerabile.

**Distribuzione di Poisson** Una importante distribuzione di probabilità numerabile è la distribuzione di Poisson. Sia $\lambda > 0$. La distribuzione di Poisson di parametro $\lambda$ è definita su $\Omega = \mathbb{N} \cup \{0\}$ ponendo:

$$P(k) = e^{-\lambda} \frac{\lambda^k}{k!} \qquad \text{per } k = 0, 1, 2, 3, \ldots$$

La distribuzione di Poisson descrive "eventi rari".

## 2.8 Variabili casuali

Generalmente negli esperimenti casuali concreti, siamo interessati non tanto nel risultato dell'esperimento in sé, quanto in delle quantità numeriche che dipendono dal risultato. Per esempio, in uno studio statistico di una popolazione interessano non tanto le persone scelte, ma quantità quali età, statura, voto riportato ad un esame, ecc.

**Definizione 13.** *Sia $(\Omega, P)$ uno spazio di probabilità finito. Una funzione $X : \Omega \to \mathbb{R}$ è detta* variabile casuale.

Sia $\Omega = (\omega_1, \omega_2, \ldots, \omega_n)$ lo spazio campione e sia $X : \Omega \to \mathbb{R}$. Sia

$$X(\Omega) = \{x_1, x_2, \ldots, x_h\} = \{X(\omega_j),\ j = 1, 2, \ldots, n\}$$

l'immagine di $\Omega$ mediante $X$. Per $j = 1, 2, \ldots, h$ indichiamo con $\{X = x_j\}$ il sottoinsieme di $\Omega$ definito da $\{\omega \in \Omega : X(\omega) = x_j\}$. Una utile rappresentazione di $X$ si ottiene tracciando dei segmenti verticali di lunghezza $P(\{X = x_j\})$ a partire dai punti $x_j$, $j = 1, 2, \ldots, h$.

**Definizione 14.** *La funzione $x_j \to P(\{X = x_j\})$ definita su $X(\Omega)$ è detta* distribuzione di probabilità *di $X$.*

**Esempio** Riprendiamo l'esempio del lancio di moneta equa dell'Esercizio 4 e dell'Esempio 1. Consideriamo la variabile casuale definita da $X(T) = 1$, $X(C) = 0$. Risulta $X(\Omega) = \{0, 1\}$, $\{X = 1\} = \{T\}$, $\{X = 0\} = \{C\}$. Quindi, la distribuzione di probabilità di $X$ è data da $P(\{X = 1\}) = P(T) = 1/2$ e $P(\{X = 0\}) = P(C) = 1/2$.
**Esempio** Consideriamo adesso una moneta truccata. Abbiamo quindi ancora $\Omega = \{T, C\}$ e supponiamo che, per esempio, $P(T) = 1/3$ e $P(C) = 2/3$. Consideriamo ancora la variabile casuale definita da $X(T) = 1$, $X(C) = 0$. Risulta $X(\Omega) = \{0, 1\}$, $\{X = 1\} = \{T\}$, $\{X = 0\} = \{C\}$. Concludiamo che la distribuzione di probabilità di $X$ è data da $P(\{X = 1\}) = P(T) = 1/3$ e $P(\{X = 0\}) = P(C) = 2/3$.

**Esempio** Consideriamo il lancio di un dado. In questo caso $\Omega = \{1, 2, 3, 4, 5, 6\}$. Consideriamo la variabile casuale $X : \Omega \to \mathbb{R}$ definita da $X(j) = j$, per ogni $j \in \Omega$. Allora $\{X = j\} = \{j\}$ e $P(\{X = j\}) = P(j) = 1/6$ per ogni $j = 1, 2, 3, 4, 5, 6$.
**Esempio 2.** Riprendiamo l'esempio del lancio di due dadi dell'Esercizio 7. In questo caso $\Omega = \{(a, b) : \ a, b, \in \{1, 2, 3, 4, 5, 6\}\}$ e $P((a, b)) = 1/36$ per ogni $(a, b) \in \Omega$. La funzione $X : \Omega \to \mathbb{R}$ definita da: $X((a, b)) = a + b$ è una variabile casuale. Per rappresentare $X$ calcoliamo $X(a, b)$ per ogni $(a, b) \in \Omega$.

$$P(\{X = 2\}) = \frac{1}{36};\ P(\{X = 3\}) = \frac{2}{36} = \frac{1}{18};\ P(\{X = 4\}) = \frac{3}{36} = \frac{1}{12};$$
$$P(\{X = 5\}) = \frac{4}{36} = \frac{1}{9}; P(\{X = 6\}) = \frac{5}{36};\ P(\{X = 7\}) = \frac{6}{36} = \frac{1}{6};$$
$$P(\{X = 8\}) = \frac{5}{36};\ P(\{X = 9\}) = \frac{4}{36} = \frac{1}{9};\ P(\{X = 10\}) = \frac{3}{36} = \frac{1}{12};$$
$$P(\{X = 11\}) = \frac{2}{36} = \frac{1}{18};\ P(\{X = 12\}) = \frac{1}{36}.$$

I seguenti due concetti sono fondamentali.

**Definizione 15** (Valore medio)**.** *Sia $\Omega = \{\omega_1, \omega_2, \dots, \omega_n\}$ uno spazio campione finito. Sia $X : \Omega \to \mathbb{R}$ una variabile casuale. Il* valore medio *di $X$, denotato $E(X)$ è definito da:*

$$\begin{aligned} E(X) =& X(\omega_1)P(\omega_1) + X(\omega_2)P(\omega_2) + \cdots + X(\omega_n)P(\omega_n) \\ =& \sum_{j=1}^{n} X(\omega_j)P(\omega_j). \end{aligned}$$

Il valore medio ha il significato di media dei valori assunti da $X$, pesati secondo la probabilità $P$.

**Definizione 16** (Varianza)**.** *Sia $\Omega = \{\omega_1, \omega_2, \dots, \omega_n\}$ uno spazio campione finito. Sia $X : \Omega \to \mathbb{R}$ una variabile casuale. La* varianza *di $X$ è definita da:*

$$\begin{aligned} D^2(X) =& (X(\omega_1) - E(X))^2 P(\omega_1) + (X(\omega_2) - E(X))^2 P(\omega_2) + \cdots \\ & + (X(\omega_n) - E(X))^2 P(\omega_n) \\ =& \sum_{j=1}^{n} (X(\omega_j) - E(X))^2 P(\omega_j). \end{aligned}$$

La varianza ha il significato di media dei quadrati dei discostamenti dei valori assunti da $X$ dal valore medio di $X$, pesati secondo la probabilità $P$.

**Esempio** Consideriamo il lancio di moneta. Al solito, abbiamo $\Omega = \{T, C\}$, $P(T) = p$, $P(C) = 1 - p$ e consideriamo la variabile casuale$X : \Omega \to \mathbb{R}$ definita da $X(T) = 1$, $X(C) = 0$. Calcoliamo la media di $X$:

$$E(X) = X(T)P(T) + X(C)P(C) = 1 \cdot p + 0 \cdot (1 - p) = p.$$

Osserviamo che, nel caso particolare $p = 1/2$ (moneta equa) otteniamo $E(X) = 1/2$.

Calcoliamo la varianza di $X$.

$$\begin{aligned} D^2(X) =& [X(T) - E(X)]^2 P(T) + [X(C) - E(X)]^2 P(C) \\ =& (1-p)^2 p + (0-p)^2(1-p) = p(1-p)(1-p+p) = p(1-p). \end{aligned}$$

Nel caso di moneta equa otteniamo $D^2(X) = 1/4$.
**Osservazione** Sia $(\Omega, P)$ uno spazio di probabilità finito e sia $X : \Omega \to \mathbb{R}$. una variabile casuale. Osserviamo che $X$ determina un nuovo spazio campione $X(\Omega)$ su cui è definita la probabilità:

$$P(x) = P(X = x) = \sum_{X(\omega)=x} P(\omega)$$

per ogni $x \in X(\Omega)$. Generalmente ha più interesse questa probabilità determinata da $X$ su $X(\Omega)$ che lo spazio originario $(\Omega, P)$. Pertanto, spesso si specifica solo la probabilità $P(X = x)$, ovvero, la distribuzione di probabilità di $X$. È facile verificare che la media e la varianza di $X$ si possono definire in termini di $P(X = x)$, senza fare riferimento a $(\Omega, P)$ come segue:

$$E(X) = \sum_{x \in X(\Omega)} x P(X = x), \qquad D^2(X) = \sum_{x \in X(\Omega)} [x - E(X)]^2 P(X = x).$$

**Esempio** Scriviamo esplicitamente la probabilità $P(X = x)$ nel caso di lancio di due dadi. Abbiamo: $\Omega = \{(h, k) : h, k \in \{1, 2, 3, 4, 5, 6\}\}$ e $P((h, k)) = 1/36$ per ogni $(h, k) \in \Omega$. La variabile casuale $X$ è data da: $X(h, k) = h + k$. Allora, $X(\Omega) = \{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$ e $P(X = 2) = 1/36$, $P(X = 3) = 2/36$, come già fatto nell'Esempio 2.

Per calcolare esplicitamente la varianza, risulta molto utile la seguente formula.

**Proposizione 5.** *Sia $X$ una variabile casuale e sia $E(X)$ la sua media. Sia inoltre $E(X^2)$ la media della variabile casuale $X^2$. Risulta:*

$$D^2(X) = E(X^2) - [E(X)]^2. \tag{9}$$

*Dimostrazione.* Per definizione di $D^2(X)$, abbiamo:

$$\begin{aligned} D^2(X) &= \sum_{x\in X(\Omega)} (x - E(X))^2 \, P(X = x) \\ &= \sum_{x\in X(\Omega)} \left\{x^2 - 2xE(X) + [E(X)]^2\right\} P(X = x) \\ &= \sum_{x\in X(\Omega)} x^2 P(X = x) - 2E(X) \sum_{x\in X(\Omega)} xP(X = x) \\ &\qquad\qquad + [E(X)]^2 \sum_{x\in X(\Omega)} P(X = x). \end{aligned}$$

Per definizione di media, abbiamo $\sum_{x\in X(\Omega)} xP(X = x) = E(X)$. Per l'assioma di normalizzazione (7) abbiamo $\sum_{x\in X(\Omega)} P(X = x) = P(\Omega) = 1$. Infine, calcoliamo:

$$\sum_{x\in X(\Omega)} x^2 P(X = x) = \sum_{x^2\in X^2(\Omega)} x^2 P(X^2 = x^2) = E(X^2).$$

Concludiamo quindi che

$$D^2(X) = E(X^2) - 2[E(X)]^2 + [E(X)]^2 = E(X^2) - [E(X)]^2,$$

come volevasi dimostrare. □

Ad esempio, ricalcoliamo la varianza nel caso precedente di lancio di moneta. La media della variabile casuale $X^2$ è data da

$$E(X^2) = X^2(T)P(T) + X^2(C)P(C) = p.$$

Per la formula (9), otteniamo:

$$D^2(X) = p - p^2 = p(1 - p),$$

in accordo con quanto ottenuto precedentemente.

**Media e varianza di variabile casuale binomiale.** Calcoliamo la media e la varianza di una variabile casuale binomiale $X$ di parametri $(n, p)$. $X$ rappresenta quindi il numero di successi in $n$ prove ripetute, essendo $p$ la probabilità del singolo successo. I valori che assume $X$ sono quindi $\{0, 1, 2, \ldots, n\}$ e la distribuzione di $X$ è data da:

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

Mostriamo che

$$E(X) = np \qquad D^2(X) = np(1-p). \tag{10}$$

Calcoliamo la media di $X$.

$$\begin{aligned} E(X) =& \sum_{k=0}^{n} k \binom{n}{k} p^k (1-p)^{n-k} = \sum_{k=1}^{n} k \binom{n}{k} p^k (1-p)^{n-k} \\ =& \sum_{k=1}^{n} k \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} = \sum_{k=1}^{n} \frac{n!}{(k-1)!(n-k)!} p^k (1-p)^{n-k}. \end{aligned}$$

Ponendo $h = k - 1$, possiamo scrivere

$$\begin{aligned} E(X) =& \sum_{h=0}^{n-1} \frac{n!}{h!(n-h-1)!} p^{h+1} (1-p)^{n-h-1} \\ =& np \sum_{h=0}^{n-1} \frac{(n-1)!}{h!(n-1-h)!} p^h (1-p)^{n-1-h} = np \sum_{h=0}^{n-1} \binom{n-1}{h} p^h (1-p)^{n-1-h} \\ =& np(p+1-p)^{n-1} = np, \end{aligned}$$

dove la penultima uguaglianza segue dal Teorema 1 con $a = p$, $b = 1 - p$. Vogliamo adesso calcolare la varianza di $X$ usando la formula (9). A tal fine, calcoliamo $E(X^2)$. Per definizione,

$$E(X^2) = \sum_{k=0}^{n} k^2 \binom{n}{k} p^k (1-p)^{n-k} = \sum_{k=1}^{n} k^2 \binom{n}{k} p^k (1-p)^{n-k}.$$

Scrivendo $k^2 = k(k-1) + k$, otteniamo che:

$$\begin{aligned} E(X^2) =& \sum_{k=1}^{n} k(k-1) \binom{n}{k} p^k (1-p)^{n-k} + \sum_{k=0}^{n} k \binom{n}{k} p^k (1-p)^{n-k} \\ =& \sum_{k=1}^{n} k(k-1) \binom{n}{k} p^k (1-p)^{n-k} + E(X). \end{aligned}$$

Sviluppiamo il primo addendo nell'ultimo rigo:

$$\begin{aligned} \sum_{k=1}^{n} k(k-1) \binom{n}{k} p^k (1-p)^{n-k} =& \sum_{k=2}^{n} k(k-1) \binom{n}{k} p^k (1-p)^{n-k} \\ =& \sum_{k=2}^{n} \frac{n!}{(k-2)!(n-k)!} p^k (1-p)^{n-k}. \end{aligned}$$

Ponendo $h = k - 2$, possiamo scrivere:

$$\begin{aligned}
\sum_{k=1}^{n} k(k-1)\binom{n}{k} p^k(1-p)^{n-k} &= \sum_{h=0}^{n-2} \frac{n!}{h!(n-h-2)!} p^{h+2}(1-p)^{n-h-2} \\
&= n(n-1)p^2 \sum_{h=0}^{n-2} \frac{(n-2)!}{h!(n-2-h)!} p^2(1-p)^{n-2-h} \\
&= n(n-1)p^2 \sum_{h=0}^{n-2} \binom{n-2}{h} p^2(1-p)^{n-2-h} \\
&= n(n-1)p^2(p+1-p)^{n-2} = n(n-1)p^2.
\end{aligned}$$

In conclusione, abbiamo che

$$E(X^2) = n(n-1)p^2 + E(X) = n(n-1)p^2 + np.$$

Sostituendo nella formula (9),

$$D^2(X) = n(n-1)p^2 + np - (np)^2 = -np^2 + np = np(1-p).$$

**Variabili casuali indipendenti** Per definizione, si dice che due variabili casuali sono *indipendenti* se:

(11) $$P(X = a, Y = b) = P(X = a)P(Y = b).$$

Si confronti questa definizione con quella di eventi indipendenti. Il tipico esempio di variabili casuali indipendenti è legato agli esperimenti di prove ripetute.

**Esercizio** Supponiamo di lanciare una moneta 2 volte. Sia $X_1$ la variabile casuale che vale 1 se al primo lancio esce testa e 0 se al primo lancio esce croce. Sia $X_2$ la variabile casuale che vale 1 se al secondo lancio esce testa e 0 se al secondo lancio esce croce. Dimostrare che $X_1$ e $X_2$ sono indipendenti.

*Soluzione.* Abbiamo:

$$P(X_1 = 1) = P(X_1 = 0) = \frac{1}{2}, \qquad P(X_2 = 1) = P(X_2 = 0) = \frac{1}{2}.$$

Inoltre, contando casi favorevoli e non, otteniamo che

$$\begin{aligned}
P((X_1, X_2) = (T, T)) &= P((X_1, X_2) = (T, C)) = P((X_1, X_2) = (C, T)) \\
&= P((X_1, X_2) = (C, C)) = \frac{1}{4}.
\end{aligned}$$

Abbiamo quindi, per esempio,

$$P((X_1, X_2) = (T, T)) = \frac{1}{4} = \frac{1}{2} \cdot \frac{1}{2} = P(X_1 = T)P(X_2 = T).$$

I rimanenti casi sono analoghi. □

## 2.9 Cenni alle variabili casuali continue

In molti casi i risultati possibili degli esperimenti sono numeri appartenenti ad un *intervallo*. Pensiamo, ad esempio, alla durata media del funzionamento di una macchina. Siamo quindi naturalmente condotti a considerare variabili casuali che hanno per codominio un intervallo. Tali variabili casuali sono dette *continue*.

Rispetto al caso di variabili discrete, il caso di variabili casuali continue presenta notevoli complicazioni dal punto di vista matematico. Pertanto, ci limitiamo ad illustrare le idee principali in maniera intuitiva.

Sia $(\Omega, P)$ uno spazio di probabilità finito, sia $X : \Omega \to \mathbb{R}$ una variabile casuale e sia $X(\Omega) = \{x_1, x_2, \ldots, x_h\}$. Abbiamo visto che la distribuzione di probabilità di $X$ è una funzione definita sull'insieme finito $X(\Omega)$ a valori in $[0, 1]$. Posto inoltre

$$p_j = P(\{X = x_j\}) \qquad j = 1, 2, \ldots, h,$$

per la proprietà di normalizzazione (7) abbiamo che:

$$p_1 + p_2 + \ldots + p_h = \sum_{j=1}^{h} p_j = P(\Omega) = 1.$$

Inoltre, indicato con $\{a \leq X \leq b\} = \{\omega \in \Omega : a \leq X(\omega) \leq b)\}$, abbiamo che per ogni $a, b \in \mathbb{R}$ con $a \leq b$ risulta:

$$P(\{a \leq X \leq b\}) = \sum \{P(\omega_j) : \; a \leq X(\omega_j) \leq b\}.$$

A questo punto non è difficile immaginare la situazione in cui la distribuzione di probabilità di una variabile casuale $X$ sia definita su un *intervallo* da una *densità di probabilità*. Più precisamente, sia $f : \mathbb{R} \to [0, +\infty)$ continua e tale che

$$\int_{-\infty}^{+\infty} f(t) \; \mathrm{d}t = 1$$

(si veda l'Appendice per la definizione di integrali su intervalli infiniti). Una variabile casuale $X$ tale che per ogni $a, b \in \mathbb{R}$ con $a \leq b$

$$P(a \leq X \leq b) = \int_a^b f(t) \; \mathrm{d}t$$

è detta *variabile casuale continua*. A partire dalla densità di probabilità definiamo il valore medio e la varianza di una variabile casuale continua.

**Definizione 17.** *Sia $X$ una variabile casuale continua avente densità di probabilità $f$. Il valore medio di $X$, se esiste, è definito da:*

$$E(X) = \int_{-\infty}^{+\infty} x f(x)\,\mathrm{d}x.$$

*La varianza di $X$, se esiste, è definita da:*

$$D^2(X) = \int_{-\infty}^{+\infty} (x - E(X))^2 f(x)\,\mathrm{d}x.$$

Osserviamo esplicitamente che la funzione $F_X : \mathbb{R} \to [0, 1]$ definita da:

$$F_X(x) = \int_{-\infty}^{x} f(t)\ \mathrm{d}t$$

è detta funzione di distribuzione di $X$ e viene usata frequentemente al posto della densità $f(x)$.

Le seguenti densità sono di fondamentale importanza.

### Densità normale o di Gauss

Una variabile casuale continua è detta normale o di Gauss se la sua densità di probabilità è del tipo:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\{-\frac{(x-m)^2}{2\sigma^2}\},$$

dove $\sigma > 0$ e $m \in \mathbb{R}$. In questo caso risulta:

$$E(X) = m; \qquad\qquad D^2(X) = \sigma^2.$$

Questa densità è fondamentale in quanto descrive gli errori di misurazione. La dimostrazione rigorosa di tale proprietà è stata data da Kolmogoroff (Teorema Centrale del Limite).

### Densità chi-quadrato

Sia $\nu \in \mathbb{N}$. Una variabile casuale continua è detta chi-quadrato con $\nu$ gradi di libertà se la sua densità di probabilità è del tipo:

$$f(x) = \begin{cases} \frac{x^{\nu/2-1} e^{-x/2}}{2^{\nu/2}\Gamma(\nu/2)} & \text{se } x > 0 \\ 0 & \text{altrimenti,} \end{cases}$$

dove $\Gamma$ denota la funzione Gamma di Eulero (si veda l'Appendice). In questo caso risulta:

$$E(X) = \nu; \qquad\qquad D^2(X) = 2\nu.$$

Questa densità si usa nel test di ipostesi "chi-quadrato".

# 3 Statistica inferenziale

Una delle principali finalità della statistica è quella di descrivere un insieme grande attraverso lo studio di una sua piccola parte, detta *campione*. Per fare ciò in maniera rigorosa, si postula che la proprietà che si vuole studiare sia regolata da una distribuzione di probabilità. Ovvero, si suppone che scegliendo a caso degli elementi dall'insieme grande e valutando con un criterio numerico su di essi la proprietà oggetto di studio, si ottengono delle variabili casuali indipendenti, aventi tutte una stessa distribuzione di probabilità.

Tale distribuzione, la cui esistenza è postulata, è generalmente l'incognita del problema.

In tale contesto, la statistica affronta tre principali problemi:

**(i)** Determinare la "forma generale" o modello della distribuzione di probabilità incognita. (Ad esempio, stabilire che è gaussiana, binomiale, ecc.) Questo tipo di problema è detto *modellizzazione*.

**(ii)** Essendo noto il modello, determinarne uno o più parametri. Ad esempio, sapendo che la distribuzione è gaussiana, determinarne la media $m$ e/o la varianza $\sigma$. Questo tipo di problema è detto *stima di parametri*.

**(iii)** Fornire criteri per verificare l'attendibilità del modello usato, ovvero per stabilire se i dati osservati sono compatibili con esso, con un assegnato livello di significatività. Questo tipo di problema è detto *verifica di ipotesi*.

Vedremo in quanto segue degli esempi di problemi di tipo (ii) e di tipo (iii).

## 3.1 Esempio di stima di parametri: il principio di massima verosimiglianza

Vogliamo determinare se una moneta è truccata o meno. A tal fine, lanciamola un elevato numero di volte (ad es. 100 000). Il nostro insieme grande è quindi l'insieme di 100 000 risultati di lancio. Non potendo (o non volendo) considerare ad uno ad uno i risultati, scegliamone a caso 100. Supponiamo che nei 100 lanci selezionati a caso sia uscito 46 volte testa e 54 volte croce.

Questi valori ci inducono a ritenere che, se anche la moneta fosse truccata, non dovrebbe esserlo eccessivamente.

Per avere un'idea di ragionamento statistico, supponiamo di sapere inoltre, per altra via, che la probabilità che esca testa o è $0,5$, o è $0,4$. La probabilità di osservare 46 volte testa e 54 croce nell'ordine particolare in cui sono usciti è data da:

$$\mathcal{L}(0,5) = (0,5)^{46}(0,5)^{54} = (0,5)^{100} = 7,88 \cdot 10^{-31},$$

se $P(\text{testa}) = 0,5$. Invece, se $P(\text{testa}) = 0,4$, tale probabilità è data da:

$$\mathcal{L}(0,4) = (0,4)^{46}(0,6)^{54} = 5,18 \cdot 10^{-31}.$$

Abbiamo quindi che $\mathcal{L}(0,5) > \mathcal{L}(0,4)$. A questo punto è naturale ritenere più verosimile $P(\text{testa}) = 0,5$, invece che $P(\text{testa}) = 0,4$, visto che tale valore rende più alta la probabilità dei dati ottenuti.
L'esempio appena descritto è un caso particolare del fondamentale "principio di massima verosimiglianza" ("maximum likelihood principle", in inglese, da cui il simbolo $\mathcal{L}$).

## 3.2 Esempio di test di ipotesi: il test chi-quadrato

Diamo prima un'idea di questo metodo con un esempio. Vogliamo studiare la percentuale di stranieri in una grande città. Supponiamo che da altre informazioni siamo portati a congetturare che il 30% degli abitanti siano stranieri. Scegliamo a caso 100 abitanti. Supponiamo di aver constatato che 20 di questi abitanti sono stranieri, ed i rimanenti 80 sono italiani. Ci chiediamo: l'ipotesi che il 30% degli abitanti è straniero è compatibile con la nostra osservazione?

In base ai valori riscontrati, siamo portati a pensare che la stima del 30% non sia attendibile. Vediamo adesso con un test statistico ("test chi-quadrato") che, in base ai nostri dati, con probabilità di errore dello 0,5%, dobbiamo rifiutare l'ipotesi che il 30% sia la percentuale corretta. A tale scopo, definiamo l'ipotesi nulla

$$H_0 : \text{gli stranieri costituiscono il 30\% degli abitanti}$$

e l'ipotesi alternativa:

$$H_1 : \text{gli stranieri } \textit{non} \text{ costituiscono il 30\% degli abitanti.}$$

Vogliamo stabilire se $H_0$ va rifiutata. Decidiamo di accontentarci di un "livello di significatività" del 5%. Cioè, accettiamo di rifiutare $H_0$, con una probabilità di errore del 5% che $H_0$ sia invece vera.

Sotto l'ipotesi $H_0$, la probabilità che un abitante a caso sia italiano è $p = 7/10$ e la probabilità che sia straniero è $1 - p = 3/10$. Poichè abbiamo scelto 100 abitanti come campione, denotiamo $n = 100$. Poichè abbiamo osservato 80 italiani e 20 stranieri, poniamo $n_1 = 80$ e $n_2 = 20$. Per il test chi-quadrato abbiamo bisogno delle seguenti quantità, di cui daremo in seguito una giustificazione intuitiva:

$$np = 100 \cdot \frac{7}{10} = 70 \qquad\qquad n(1-p) = 100 \cdot \frac{3}{10} = 30.$$

Consideriamo la seguente quantità:

$$\chi^2 = \frac{(n_1 - np)^2}{np} + \frac{(n_2 - n(1-p))^2}{n(1-p)} = \frac{(80-70)^2}{70} + \frac{(20-30)^2}{30} = 4,762.$$

Questa può interpretarsi come una misura del discostamento globale tra il risultato delle osservazioni e quanto previsto dall'ipotesi nulla. Poichè abbiamo due classi di individui (italiani e non), confrontiamo questo valore di $\chi^2$ con la densità chi-quadrato con un grado di libertà (il grado di libertà è dato dal numero di classi in considerazione, diminuito di 1). Poichè abbiamo scelto un livello di significatività del 5%, prendiamo dalla tavola dei "quantili" chi-quadrato $\chi^2_{1;0,05}$. Troviamo (v. Tabella) che $\chi^2_{1;0,05} = 3,841$. Abbiamo quindi che:

$$\chi^2 = 4,762 > 3,841 = \chi^2_{1;0,05}.$$

Il test ci dice che $H_0$ *va rifiutata*. Osserviamo che se avessimo scelto un livello di significatività del 2,5%, avremmo invece trovato:

$$\chi^2 = 4,762 < 5.024 = \chi^2_{1;0,025}$$

ed il test *non* avrebbe portato al rifiuto di $H_0$. Cioè, se scegliamo il livello di significatività 2,5%, l'aver constatato la presenza di 80 italiani e 20 stranieri non consente di rifiutare l'ipotesi che la popolazione è costituita dal 30% di stranieri.

## Test chi-quadrato: regola generale

In generale, supponiamo che la proprietà della popolazione che studiamo è descritta dalla variabile casuale incognita $X$. Siano $a_1, a_2, \ldots, a_k$ i valori che assume $X$. Vogliamo confrontare l'ipotesi nulla

$$H_0 : P(X = a_j) = p_j \qquad \text{per ogni } j = 1, 2, \ldots, k$$

con l'ipotesi alternativa:

$$H_0 : P(X = x_j) \neq p_j \qquad \text{per qualche } j = 1, 2, \ldots, k$$

Supponiamo di esaminare un campione di $n$ elementi della popolazione scelti a caso.

L'idea del test è: se $P(X = a_j) = p_j$, allora il numero di volte che osservo $a_j$ deve riflettere in qualche modo questo valore $p_j$. Introduciamo quindi delle nuove variabili casuali definite come segue. $N_1$ è il numero di osservazioni di $a_1$; $N_2$ è il numero di osservazioni di $a_2$; in generale $N_j$ è il numero di osservazioni di $a_j$, $j = 1, 2, \ldots, k$. L'osservazione del campione produce la $k$-pla $(n_1, n_2, \ldots, n_k)$. Le variabili casuali $N_j$, $j - 1, 2, \ldots, k$, hanno *distribuzione binomiale*. Infatti, la probabilità $P(N_j = n_j)$ è la probabilità che esca $n_j$ volte $a_j$ in $n$ scelte a caso (prove ripetute). Quindi,

$$P(N_j = n_j) = \binom{n}{n_j} P(X = a_j)^{n_j} (1 - P(X = a_j))^{n - n_j},$$

$j - 1, 2, \ldots, k$. Sotto l'ipotesi nulla $H_0$, abbiamo quindi

$$P(N_j = n_j) = \binom{n}{n_j} p_j^{n_j} (1 - p_j)^{n - n_j}.$$

Ancora, sotto l'ipotesi nulla $H_0$, la media di $N_j$ è data da (ricordare la formula (10)) $E(N_j) = np_j$. Non sorprende quindi che il test prenda come valutazione dell'errore la quantità:

$$\chi^2 = \frac{(n_1 - np_1)^2}{np_1} + \frac{(n_2 - np_2)^2}{np_2} + \ldots + \frac{(n_k - np_k)^2}{np_k} = \sum_{j=1}^{k} \frac{(n_j - np_j)^2}{np_j}.$$

Questa quantità va confrontata con il corrispondente quantile $\chi^2_{k-1;\alpha}$ della densità chi-quadrato con $k - 1$ gradi di libertà, determinato dal livello di significatività $\alpha$ scelto. In generale,il test individua la seguente regione di rifiuto:

$$\mathcal{C} = \{(n_1, \ldots, n_k) : \ \chi^2 \geq \chi^2_{k-1;\alpha}\}.$$

In altre parole, se l'osservazione $(n_1, n_2, \ldots, n_k)$ porta ad un valore $\chi^2$ tale che $\chi^2 \geq \chi^2_{k-1;\alpha}$, si conclude che (con il livello di significatività scelto) l'ipotesi $H_0$ è incompatibile con le ossevazioni, e quindi va rifiutata.

Questo test è dovuto a E.S. Pearson.

**Distribuzione chi-quadrato: valori di $\chi^2_{\alpha;\nu}$**

| $\nu \setminus \alpha$ | .995 | .99 | .975 | .95 | .05 | .025 | .01 | .005 |
|---|---|---|---|---|---|---|---|---|
| 1 | 0.0000393 | 0.000157 | 0.000982 | 0.00393 | 3.841 | 5.024 | 6.635 | 7.879 |
| 2 | 0.0100 | 0.0201 | 0.0506 | 0.103 | 5.991 | 7.378 | 9.210 | 10.597 |
| 3 | 0.0717 | 0.115 | 0.216 | 0.352 | 7.815 | 9.348 | 11.345 | 12.838 |
| 4 | 0.207 | 0.297 | 0.484 | 0.711 | 9.488 | 11.143 | 13.277 | 14.860 |
| 5 | 0.412 | 0.554 | 0.831 | 1.145 | 11.070 | 12.832 | 15.086 | 16.750 |
| 6 | 0.676 | 0.872 | 1.237 | 1.635 | 12.592 | 14.449 | 16.812 | 18.548 |
| 7 | 0.989 | 1.239 | 1.690 | 2.167 | 14.067 | 16.013 | 18.475 | 20.278 |
| 8 | 1.344 | 1.646 | 2.180 | 2.733 | 15.507 | 17.535 | 20.090 | 21.955 |
| 9 | 1.735 | 2.088 | 2.700 | 3.325 | 16.919 | 19.023 | 21.666 | 23.589 |
| 10 | 2.156 | 2.558 | 3.247 | 3.940 | 18.307 | 20.483 | 23.209 | 25.188 |
| 11 | 2.603 | 3.053 | 3.816 | 4.575 | 19.675 | 21.920 | 24.725 | 26.757 |
| 12 | 3.074 | 3.571 | 4.404 | 5.226 | 21.026 | 23.337 | 26.217 | 28.300 |
| 13 | 3.565 | 4.107 | 5.009 | 5.892 | 22.362 | 24.736 | 27.688 | 29.819 |
| 14 | 4.075 | 4.660 | 5.629 | 6.571 | 23.685 | 26.119 | 29.141 | 31.319 |
| 15 | 4.601 | 5.229 | 6.262 | 7.261 | 24.996 | 27.488 | 30.578 | 32.801 |
| 16 | 5.142 | 5.812 | 6.908 | 7.962 | 26.296 | 28.845 | 32.000 | 34.267 |
| 17 | 5.697 | 6.408 | 7.564 | 8.672 | 27.587 | 30.191 | 33.409 | 35.718 |
| 18 | 6.265 | 7.015 | 8.231 | 9.390 | 28.869 | 31.526 | 34.805 | 37.156 |
| 19 | 6.844 | 7.633 | 8.907 | 10.117 | 30.144 | 32.852 | 36.191 | 38.582 |
| 20 | 7.434 | 8.260 | 9.591 | 10.851 | 31.410 | 34.170 | 37.566 | 39.997 |
| 21 | 8.034 | 8.897 | 10.283 | 11.591 | 32.671 | 35.479 | 38.932 | 41.401 |
| 22 | 8.643 | 9.542 | 10.982 | 12.338 | 33.924 | 36.781 | 40.289 | 42.796 |
| 23 | 9.260 | 10.196 | 11.689 | 13.091 | 35.172 | 38.076 | 41.638 | 44.181 |
| 24 | 9.886 | 10.856 | 12.401 | 13.848 | 36.415 | 39.364 | 42.980 | 45.558 |
| 25 | 10.520 | 11.524 | 13.120 | 14.611 | 37.652 | 40.646 | 44.314 | 46.928 |
| 26 | 11.160 | 12.198 | 13.844 | 15.379 | 38.885 | 41.923 | 45.642 | 48.290 |
| 27 | 11.808 | 12.879 | 14.573 | 16.151 | 40.113 | 43.194 | 46.963 | 49.645 |
| 28 | 12.461 | 13.565 | 15.308 | 16.928 | 41.337 | 44.461 | 48.278 | 50.993 |
| 29 | 13.121 | 14.256 | 16.047 | 17.708 | 42.557 | 45.722 | 49.588 | 52.336 |
| 30 | 13.787 | 14.953 | 16.791 | 18.493 | 43.773 | 46.979 | 50.892 | 53.672 |

# Appendice

## Somma di serie

Sia $a_n$, $n \in \mathbb{N}$, una successione. La somma degli elementi $a_n$ della successione, se esiste, è definita dal seguente limite:

$$\sum_{n=1}^{\infty} a_n = \lim_{N\to\infty} \sum_{n=1}^{N} a_n. \tag{12}$$

Questa "somma di infiniti numeri" è detta *serie*. Quando il limite esiste ed è finito, si dice che la serie *converge*.

## Integrali su intervalli infiniti

Sia $f : \mathbb{R} \to \mathbb{R}$ una funzione continua. Per ogni $T > 0$ possiamo considerare l'integrale definito:

$$\int_0^T f(x)\ \mathrm{d}x.$$

L'integrale di $f$ sull'intervallo $[0, +\infty)$, se esiste, è definito da:

$$\int_0^{+\infty} f(x)\ \mathrm{d}x = \lim_{T\to+\infty} \int_0^T f(x)\ \mathrm{d}x.$$

Analogamente, l'integrale di $f$ sull'intervallo $(-\infty, 0]$, se esiste, è definito da:

$$\int_{-\infty}^{0} f(x)\ \mathrm{d}x = \lim_{T\to+\infty} \int_{-T}^{0} f(x)\ \mathrm{d}x.$$

A questo punto, se esistono entrambi gli integrali, possiamo definire:

$$\int_{-\infty}^{\infty} f(x)\ \mathrm{d}x = \int_{-\infty}^{0} f(x)\ \mathrm{d}x + \int_0^{+\infty} f(x)\ \mathrm{d}x.$$

## La funzione Gamma di Eulero

Per ogni $t > 0$ la funzione Gamma è definita da:

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

**Proposizione 6.** *La funzione Gamma soddisfa le seguenti proprietà:*

*(i)* $\Gamma(1) = 1$*;*

*(ii)* $\Gamma(t+1) = t\Gamma(t)$ *per ogni* $t > 0$*;*

*(iii)* $\Gamma(n+1) = n!$ *per ogni* $n \in \mathbb{N}$*;*

*(iv)* $\Gamma(1/2) = \sqrt{\pi}$*.*

*Dimostrazione.* (i). Risulta:

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

(ii). Integrando per parti,

$$\begin{aligned}\Gamma(t+1) &= \int_0^{+\infty} x^t e^{-x}\, \mathrm{d}x = -\int_0^{+\infty} x^t (e^{-x})'\, \mathrm{d}x \\ &= -x^t e^{-x}\Big|_0^{\infty} + \int_0^{+\infty} t x^{t-1} e^{-x}\, \mathrm{d}x = t\int_0^{+\infty} x^{t-1} e^{-x}\, \mathrm{d}x \\ &= t\Gamma(t).\end{aligned}$$

(iii). Usando la formula (ii), calcoliamo i valori di $\Gamma(n+1)$ per i primi 4 numeri naturali:

$$\begin{aligned}\Gamma(2) &= \Gamma(1+1) = 1 \cdot \Gamma(1) = 1 = 1! \\ \Gamma(3) &= \Gamma(2+1) = 2 \cdot \Gamma(2) = 2 = 2! \\ \Gamma(4) &= \Gamma(3+1) = 3 \cdot \Gamma(3) = 3 \cdot 2 = 3! \\ \Gamma(5) &= \Gamma(4+1) = 4 \cdot \Gamma(4) = 4 \cdot 3 \cdot 2 = 4!\end{aligned}$$

La formula generale si può ottenere per induzione su $n$.

Omettiamo la dimostrazione di (iv), che richiede l'integrazione in due variabili. □

# Bibliografia

F. Ayres Jr., *First Year College Mathematics*, Schaum, New York, 1958.
D. Benedetto, M. Degli Esposti e C. Maffei, *Matematica per le scienze della vita*, Casa Editrice Ambrosiana, 2008.
A.J. Cann, *Maths from Scratch for Biologists*, Wiley, Chichester, 2003.
U. Krengel, *Einführung in die Wahrscheinlichkeitstheorie und Statistik*, Vieweg, Braunschweig/Wiesbaden 1991.
S. Ross, *Probabilità e statistica per l'ingegneria e le scienze*, Apogeo, Milano, 2008.