

Lezioni di matematica discreta e probabilità

Fabrizio Caselli

Contents

Introduzione al corso	5
Contatti e ricevimento	5
Esami	5
Struttura del corso.	5
Chapter 1. Cardinalità: insiemi finiti, numerabili e continui	7
Chapter 2. Le basi del calcolo combinatorio	9
1. Il prodotto cartesiano e le sequenze	9
2. L'insieme delle parti	10
3. Iniziamo a contare	11
4. Disposizioni e permutazioni	12
5. Combinazioni, sottoinsiemi di data cardinalità e anagrammi	13
Chapter 3. Combinatoria più avanzata	19
1. I numeri di Fibonacci	19
2. Il principio di inclusione-esclusione	20
3. Partizioni di un insieme	24
Chapter 4. Statistica descrittiva	29
1. Popolazione, campione e caratteri	29
2. Classi e istogrammi	30
3. Indici di posizione o centralità	31
4. Indici di dispersione	35
5. Correlazione	36
6. Altri indici di posizione	39
Chapter 5. Spazi di probabilità	43
1. Fenomeni deterministici e casuali: risultati ed eventi	43
2. Spazi di probabilità	44
3. La probabilità uniforme	45
4. La probabilità condizionale.	49
5. La formula delle probabilità totali e la formula di Bayes	50
6. Eventi indipendenti	53
Chapter 6. Variabili aleatorie discrete	55
1. Variabili aleatorie	55

2. Variabili discrete	56
3. La densità uniforme	58
4. La funzione caratteristica e la densità di Bernoulli	58
5. La densità binomiale	59
6. La densità ipergeometrica	60
7. La densità geometrica	61
8. La densità di Poisson	64
Chapter 7. Variabili discrete multidimensionali	67
1. Densità congiunte e marginali	67
2. Variabili aleatorie indipendenti	69
3. Trasformazioni di variabili aleatorie	70
Chapter 8. Il valore atteso e la varianza	75
1. Il valore atteso	75
2. Varianza	80
3. La disuguaglianza di Chebyshev e la legge dei grandi numeri	81
Chapter 9. Variabili aleatorie continue	85
1. Funzione di ripartizione e densità di variabili continue	85
2. La media e la varianza	86
3. Densità uniforme	87
4. Densità esponenziale	89
5. Trasformazioni di variabili	90
6. Densità normali	91
7. Il teorema del limite centrale	94
Chapter 10. Statistica inferenziale	99
1. Stime e intervalli di confidenza	99
2. Test statistici	101

Introduzione al corso

Queste note sono appunti del docente per le lezioni e sono rese disponibili agli studenti come guida per gli argomenti trattati. Non vanno pertanto considerate come dispense e non sostituiscono, ma integrano, i libri consigliati.

Contatti e ricevimento

Durante il corso mi trovate in ufficio (stanza 4022C) a Cesena il mercoledì dalle 15 alle 17. Potete venire liberamente senza preavviso, ma suggerisco sempre di mandare un messaggio tramite email o a lezione per vostra sicurezza. Al termine del corso potete essere ricevuti semplicemente bussando alla porta o prendendo un appuntamento via e-mail. (fabrizio.caselli@unibo.it) Su richiesta il ricevimento potrà svolgersi anche su Teams.

Esami

Solitamente l'esame è costituito da una prova scritta, dalla discussione della prova scritta e da un orale. La discussione della prova scritta si svolge nello stesso appello della prova scritta e al suo termine potrebbe essere data allo studente la possibilità di verbalizzare un voto che non sarà in ogni caso superiore a 24 senza necessità di sostenere la prova orale. In alternativa lo studente è tenuto a sostenere la prova orale nell'appello stesso o nell'appello successivo. Lo studente può anche decidere di non presentarsi per la discussione della prova scritta e di sostenere direttamente la prova orale nell'appello successivo.

Nella prova scritta viene chiesto di risolvere alcuni esercizi (tipicamente 3). Gli esercizi vengono valutati in trentesimi e si è ammessi all'orale se il risultato ottenuto è di almeno 15/30.

Nell'orale, di circa 10/15 minuti, verrà chiesto di descrivere uno o due aspetti teorici trattati durante il corso. Una volta effettuato l'orale, in caso di valutazione negativa o di ritiro, bisogna comunque rieffettuare lo scritto. Presentarsi allo scritto in un appello automaticamente annulla la prova scritta dell'appello precedente.

Struttura del corso.

Il corso è suddiviso in quattro parti

- (1) Calcolo Combinatorio, 18 ore;
- (2) Statistica descrittiva, 7 ore;
- (3) Probabilità, 30 ore;
- (4) Statistica inferenziale, 5 ore.

Facciamo una breve descrizione di questi quattro settori in cui è suddiviso il corso.

- Il calcolo combinatorio ha il compito di *contare* quanti elementi contiene un insieme dato

- La statistica descrittiva ha il compito di *presentare* i dati raccolti in una certa indagine in modo sintetico, comunicativo e rappresentativo.
- La probabilità studia le possibilità che possa realizzarsi un certo evento casuale (o aleatorio), partendo da un modello che si dà per buono. E' una scienza esatta.
- La statistica inferenziale ha il compito di dedurre caratteristiche di una certa popolazione partendo dalla conoscenza di un certo campione scelto a caso. Siccome il risultato dipende dal campione siamo in presenza di un evento casuale e avremo bisogno della probabilità per studiare la statistica inferenziale.

Vediamo un esempio concreto di applicazione nel caso del lancio di una moneta.

- Un compito del calcolo combinatorio è quello di contare quante sono le sequenze possibili di testa e croce di lunghezza 100 in modo che ci siano complessivamente 55 testa e 45 croce.
- Un compito della statistica descrittiva può essere quello di comunicare, magari attraverso un grafico, che in 100 lanci di una moneta abbiamo ottenuto 55 volte testa e 45 volte croce.
- Un compito della probabilità può essere quello di prevedere, assumendo che la moneta non sia truccata, quanto è probabile che in 100 lanci di una moneta si ottengano 55 teste;
- Un compito della statistica inferenziale può essere quello di stimare la probabilità che una moneta sia truccata, sapendo che su 100 lanci effettuati si sono verificate 55 teste e 45 croci.

In altre parole, la probabilità cerca di prevedere quello che può accadere in un esperimento e con quale “probabilità” *prima* che l'esperimento venga effettuato; la statistica descrittiva e inferenziale hanno il compito di trarre delle conclusioni *dopo* l'effettuazione dell'esperimento stesso. Il calcolo combinatorio lo utilizzeremo soprattutto come strumento per calcolare determinate probabilità.

CHAPTER 1

Cardinalità: insiemi finiti, numerabili e continui

Daremo per buoni i concetti di insieme, sottoinsieme, complementare, funzione, dominio, codominio, funzioni iniettive, suriettive e biiettive. Dare per buono vuol dire che bisogna conoscere questi concetti, non che non è necessario conoscerli. In particolare utilizzeremo le parentesi graffe per descrivere gli elementi di un insieme. Ad esempio $A = \{1, 2, 3\}$ vuol dire che A è l'insieme costituito dagli elementi 1,2,3 e che l'ordine di questi non conta, cioè avremmo potuto anche scrivere $A = \{2, 3, 1\}$.

Come già accennato in precedenza, il calcolo combinatorio si occupa principalmente di “contare” quanti elementi contiene un insieme finito.

DEFINIZIONE. Un insieme A si dice finito se esiste $n \in \mathbb{N}$ tale che A contiene esattamente n elementi distinti o, equivalentemente, se esiste una funzione biunivoca $f : \{1, 2, \dots, n\} \rightarrow A$. In questo caso diciamo che la sua cardinalità è n e scriviamo $|A| = n$.

Osserviamo che anche l'insieme vuoto è finito ed è l'unico insieme di cardinalità 0. Diciamo che un insieme è infinito se non è finito. Esempi di insiemi infiniti che conosciamo già bene sono $\mathbb{N}, \mathbb{Z}, \mathbb{Q}, \mathbb{R}, \mathbb{R}^2$.

DEFINIZIONE. Diciamo che due insiemi (infiniti) hanno la stessa cardinalità se esiste una corrispondenza biunivoca tra di loro.

Diamo ora alcuni cenni alla teoria della cardinalità di insiemi infiniti, soprattutto per rendersi conto che ci sono insiemi infiniti che hanno la stessa cardinalità per essendo uno contenuto nell'altro e che effettivamente esistono insiemi infiniti aventi cardinalità distinte.

DEFINIZIONE. Un insieme A si dice numerabile se ha la stessa cardinalità di \mathbb{N} . Talvolta per dire che A è numerabile si scrive $|A| = \aleph_0$ (leggi alef zero, ma non lo faremo mai).

ESERCIZIO 1.1. Mostrare che ogni sottoinsieme infinito di \mathbb{N} è numerabile.

Per stabilire se un insieme A è numerabile dobbiamo riuscire ad “elencare” in un elenco infinito gli elementi di A dicendo chi è il primo, chi il secondo e così via. Questo è esattamente quello che produce una corrispondenza biunivoca di un insieme con \mathbb{N} .

ESEMPIO 1.2. L'insieme \mathbb{Z} dei numeri interi è numerabile. Infatti possiamo scrivere

$$\mathbb{Z} = \{0, +1, -1, +2, -2, \dots\}.$$

L'insieme $\mathbb{Q}_{>0}$ costituito dai numeri razionali positivi è anche numerabile, infatti,

$$\mathbb{Q}_{>0} = \{0, 1, \frac{1}{2}, 2, \frac{1}{3}, \frac{2}{3}, \frac{3}{2}, 3, \frac{1}{4}, \frac{3}{4}, \frac{4}{3}, 4, \dots\}.$$

Gli elementi di \mathbb{Q} li abbiamo elencati in questo modo: abbiamo prima scritto tutti i numeri razionali che si possono scrivere utilizzando 1 (cioè solo 1), poi abbiamo scritto in ordine crescente tutti i numeri razionali che si possono ottenere utilizzando anche il numero 2 (cioè $1/2$ e 2), poi quelli che utilizzano anche il 3 cioè $\frac{1}{3}, \frac{2}{3}, \frac{3}{2}, 3$ e così via. È chiaro che in questo modo possiamo elencare tutti i numeri razionali positivi. Chiaramente questo non è l'unico modo per farlo e anzi ce ne sono infiniti.

ESERCIZIO 1.3. Mostrache che anche \mathbb{Q} è numerabile.

DEFINIZIONE. Un insieme finito o numerabile si dice *discreto*.

In questo corso, nella parte di probabilità, quando parliamo di insieme numerabile, si potrà sempre pensare all'insieme \mathbb{Z} dei numeri interi.

Potremmo essere tentati di pensare a questo punto che tutti gli insiemi infiniti siano numerabili. Ciò non è vero come mostra il prossimo risultato.

PROPOSIZIONE 1.4. *L'insieme B di tutte le sequenze binarie infinite non è numerabile.*

DIMOSTRAZIONE. Supponiamo per assurdo che sia possibile elencare tutte le sequenze binarie e quindi scrivere

$$B = \{\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3, \dots\},$$

dove quindi le \mathbf{b}_i sono *tutte* le sequenze binarie.

Costruiamo a questo punto una nuova sequenza binaria \mathbf{b} nel seguente modo: la prima coordinata di \mathbf{b} è diversa dalla prima coordinata di \mathbf{b}_1 : ad esempio se \mathbf{b}_1 comincia con 0 allora \mathbf{b} comincia con 1 e viceversa.

La seconda coordinata di \mathbf{b} è diversa dalla seconda coordinata di \mathbf{b}_2 .

La terza coordinata di \mathbf{b} è diversa dalla terza coordinata di \mathbf{b}_3 e così via. Si ha che la sequenza \mathbf{b} così costruita è diversa dalla sequenza \mathbf{b}_i per ogni i avendone la coordinata di posto i diversa. Abbiamo quindi una contraddizione perché avremmo costruito una sequenza binaria infinita che non sta in $\{\mathbf{b}_1, \mathbf{b}_2, \dots\}$. □

Con un po' di attenzione e sfruttando la scrittura binaria di un numero reale si può dimostrare, ma non lo faremo nei particolari, che \mathbb{R} è in corrispondenza biunivoca con l'insieme delle sequenze binarie infinite e che quindi *non* è numerabile e diremo che \mathbb{R} e ogni insieme in corrispondenza biunivoca con esso ha la cardinalità del continuo. In questo caso si può scrivere $|\mathbb{R}| = \aleph_1$.

ESERCIZIO 1.5. Ogni intervallo non banale di \mathbb{R} ha anche la cardinalità del continuo.

Non ci addentriamo ulteriormente nella teoria della cardinalità di insiemi infiniti, teoria che richiederebbe un corso intero per essere ben approfondita.

ESERCIZIO 1.6 (difficile, per appassionati). Dimostrare che \mathbb{R}^2 ha la stessa cardinalità di \mathbb{R} .

CHAPTER 2

Le basi del calcolo combinatorio

Per imparare a contare quanti elementi ha un insieme è importante saper costruire insieme complicati partendo da insiemi elementari e viceversa, saper ridurre un insieme complicato ad insiemi più elementari. Lo scopo di questa lezione è proprio imparare questi tipi di operazioni tra insiemi.

1. Il prodotto cartesiano e le sequenze

Un modo semplice di costruire un insieme complicato partendo da due insiemi è il seguente.

DEFINIZIONE. Il *prodotto cartesiano* $A \times B$ tra due insiemi A e B è l'insieme delle coppie ordinate (cioè in cui l'ordine conta!) (a, b) dove $a \in A$ e $b \in B$.

Dire "l'ordine conta" significa che la coppia (a, b) è diversa dalla coppia (b, a) . Come abbiamo fatto già nel primo anno utilizzeremo le parentesi tonde per indicare che l'ordine è importante.

Nel caso in cui $A = B$ scriveremo semplicemente A^2 al posto di $A \times A$. Un esempio di prodotto cartesiano è \mathbb{R}^2 ampiamente studiato nel corso di algebra e geometria. Se consideriamo $A = \{1, 2, 3\}$ e $B = \{3, 4\}$ abbiamo

$$A \times B = \{(1, 3), (1, 4), (2, 3), (2, 4), (3, 3), (3, 4)\}.$$

(Notare qui l'uso delle parentesi graffe fuori e delle tonde dentro). Il prodotto cartesiano tra due insiemi si può generalizzare al caso di n insiemi. Siano A_1, \dots, A_n n insiemi. Allora il loro prodotto cartesiano è dato da

$$A_1 \times \dots \times A_n := \{(a_1, \dots, a_n) : a_i \in A_i, \text{ per ogni } i = 1, \dots, n\}.$$

Riprendendo gli insiemi $A = \{1, 2, 3\}$ e $B = \{3, 4\}$ dell'esempio precedente abbiamo ad esempio

$$B \times A \times B = \{(3, 1, 3), (3, 1, 4), (3, 2, 3), (3, 2, 4), (3, 3, 3), (3, 3, 4), (4, 1, 3), (4, 1, 4), (4, 2, 3), (4, 2, 4), (4, 3, 3), (4, 3, 4)\}$$

Anche in questo caso scriveremo A^n anzichè $A \times \dots \times A$ (n volte).

DEFINIZIONE. Una *sequenza (o lista) finita* di lunghezza n di elementi di A è un elemento (a_1, a_2, \dots, a_n) del prodotto cartesiano A^n .

Una sequenza infinita $\mathbf{a} = (a_1, a_2, \dots)$ (dove i puntini di sospensione indicano infiniti elementi) viene anche detta *successione* di elementi di A . Una successione di elementi di A può essere definita più formalmente come una funzione $\mathbf{a} : \mathbb{N} \rightarrow A$ ponendo $a_1 = \mathbf{a}(1)$, $a_2 = \mathbf{a}(2)$ e così via.

Vediamo alcuni esempi di sequenze:

- una sequenza di 5 quadratini colorati con rosso, blu e nero rappresenta una sequenza di lunghezza 5 nell'insieme $A = \{\text{rosso}, \text{blu}, \text{nero}\}$;

- le parole di 5 lettere possono essere interpretate come sequenze di lunghezza 5 nell'insieme di tutte le lettere.
- le sequenze binarie di una data lunghezza sono sequenze in $\{0, 1\}$. In particolare i byte sono sequenze di lunghezza 8 in $\{0, 1\}$.

2. L'insieme delle parti

Facciamo ora un ulteriore livello di astrazione ed iniziamo a pensare agli insiemi come ad elementi di un altro insieme.

DEFINIZIONE. Se A è un insieme, l'*insieme delle parti* $\mathcal{P}(A)$ è l'insieme i cui elementi sono tutti i sottoinsiemi di A , inclusi \emptyset e A stesso.

Ad esempio, se $A = \{1, a, \pi\}$ abbiamo

$$\mathcal{P}(A) = \{\emptyset, \{1\}, \{a\}, \{\pi\}, \{1, a\}, \{1, \pi\}, \{a, \pi\}, A\} :$$

in questo caso l'insieme $\mathcal{P}(A)$ ha quindi cardinalità 8. Notare l'uso delle parentesi graffe sia all'interno che all'esterno. Dimostriamo subito una proposizione di fondamentale importanza.

PROPOSIZIONE 2.1. *Sia A un insieme finito, $|A| = n$. Esiste una corrispondenza biunivoca tra $\mathcal{P}(A)$ e le sequenze binarie di lunghezza n .*

DIMOSTRAZIONE. Siccome A ha n elementi possiamo assumere per semplicità che $A = \{1, 2, \dots, n\}$. Consideriamo la seguente funzione

$$F : \mathcal{P}(A) \longrightarrow \{0, 1\}^n$$

definita in questo modo: se $B \in \mathcal{P}(A)$ è un sottoinsieme di A poniamo $F(B) = (s_1, \dots, s_n)$, dove

$$s_i = \begin{cases} 1 & \text{se } i \in B \\ 0 & \text{altrimenti.} \end{cases}$$

Ad esempio, se $n = 4$ e $B = \{2, 3\}$ allora $F(B) = (0, 1, 1, 0)$.

Consideriamo ora $G : \{0, 1\}^n \rightarrow \mathcal{P}(A)$ data da

$$G((s_1, \dots, s_n)) = \{i : s_i = 1\}.$$

Ad esempio, se $S = (1, 0, 1, 0)$ allora $G(S) = \{1, 3\}$. Ci si convince facilmente che F e G sono una l'inversa dell'altra e quindi sono delle corrispondenza biunivoche. □

ESEMPIO 2.2. Se $A = \{1, 2, 3\}$ le applicazioni F e G descritte nella precedente proposizione agiscono nel modo seguente

$$\begin{array}{ll} \emptyset & \leftrightarrow (0, 0, 0) \\ \{1\} & \leftrightarrow (1, 0, 0) \\ \{2\} & \leftrightarrow (0, 1, 0) \\ \{3\} & \leftrightarrow (0, 0, 1) \\ \{1, 2\} & \leftrightarrow (1, 1, 0) \\ \{2, 3\} & \leftrightarrow (0, 1, 1) \\ \{1, 3\} & \leftrightarrow (1, 0, 1) \\ \{1, 2, 3\} & \leftrightarrow (1, 1, 1) \end{array}$$

Altri tipi di insiemi che si possono costruire da uno o più insiemi dati li incontreremo più avanti.

3. Iniziamo a contare

Ci sono alcuni principi elementari che utilizzeremo disinvoltamente senza richiamarli di volta in volta. Essi sono

- Il principio di uguaglianza: se due insiemi finiti A e B sono in corrispondenza biunivoca allora hanno la stessa cardinalità; ad esempio se voglio sapere quanti cioccolatini ho mangiato, posso contare quanti incarti ho buttato nella pattumiera.
- Il principio di somma: se due insiemi finiti A e B sono disgiunti (sai la differenza tra insiemi disgiunti e distinti?) allora $|A \cup B| = |A| + |B|$; questo principio si estende chiaramente al caso di un numero finito di insiemi. Ad esempio, se voglio sapere quanti frutti ho nella fruttiera, mi basta contare quanti frutti ho di ciascun tipo (mele, pere,...) per poi farne la somma.
- Il principio di moltiplicazione: dati due insiemi A e B si ha $|A \times B| = |A| \cdot |B|$. Ad esempio supponiamo di avere un insieme M di 3 maschi e uno F di 4 femmine. Quante coppie (etero) posso formare? Le coppie sono rappresentate in questo caso dal prodotto cartesiano $A \times B$ e sono quindi 12 per il principio di moltiplicazione. Quante sono tutte le coppie possibili (etero o omo)?

Questi principi possono essere facilmente dimostrati utilizzando la definizione di cardinalità.

Spesso in probabilità incontreremo particolari sottoinsiemi $C \subset A \times B$ del prodotto cartesiano tra A e B che soddisfano la seguente proprietà: esistono due interi positivi n, m tali che

- gli elementi $a \in A$ tali che $(a, b) \in C$ per qualche $b \in B$ sono n (sono cioè n le possibili "scelte" per la prima coordinata di un elemento $(a, b) \in C$);
- per ogni $a \in A$ come nel punto precedente gli elementi b per cui $(a, b) \in C$ sono m (fissata cioè la prima coordinata di un elemento di C le possibili "scelte" per la seconda sono m).

Diciamo in questo caso che C è un prodotto *condizionato* tra A e B di tipo n, m e possiamo verificare, utilizzando ad esempio il principio di somma, che $|C| = nm$.

Il prodotto cartesiano è l'esempio più semplice di prodotto condizionato. Un altro esempio è dato dalle coppie di elementi uguali $\{(a, a) : a \in A\}$, con $n = |A|$ e $m = 1$. O ancora i primi due numeri estratti nella tombola con $n = 90$ e $m = 89$.

La definizione di prodotto condizionato si estende naturalmente al prodotto di un numero finito di insiemi. Siano quindi A_1, \dots, A_k insiemi e $C \subseteq A_1 \times A_2 \times \dots \times A_k$. Diciamo che C è un prodotto condizionato di tipo (n_1, \dots, n_k) se

- l'elemento a_1 di un elemento di C può essere scelto in n_1 modi;
- per ogni $i < k$ e per ogni scelta di $a_1 \in A_1, \dots, a_{i-1} \in A_{i-1}$ tali che esiste un elemento di C le cui prime $i-1$ coordinate sono a_1, \dots, a_{i-1} esistono esattamente n_i elementi $a_i \in A_i$ tali che a_1, \dots, a_i sono le prime i coordinate di un elemento di C .

ESERCIZIO 2.3. Dimostrare, ad esempio per induzione, che un prodotto condizionato di tipo n_1, \dots, n_k ha esattamente $n_1 \cdot n_2 \cdot \dots \cdot n_k$ elementi.

Un esempio è costituito dalle sequenze di 5 cifre decimali aventi tutte le cifre distinte. Un altro ancora dalle sequenze di 5 cifre in cui le cifre adiacenti sono numeri consecutivi (in cui pensiamo

anche 0 e 9 come cifre consecutive). Se invece non riteniamo 0 e 9 come cifre consecutive allora non abbiamo un prodotto condizionato.

ESEMPIO 2.4. Consideriamo l'insieme

$$A = \{(a, b, c) : a, b, c \in \mathbb{N}, a \leq 10, 0 \leq b - a \leq 10, 0 \leq c - a - b \leq 10\}.$$

Si ha che A è un prodotto condizionato in $\mathbb{N} \times \mathbb{N} \times \mathbb{N}$ di tipo (11,11,11) e quindi $|A| = 1331$.

Il termine “prodotto condizionato” non compare solitamente nei libri, ma siccome ricorrerà spesso ho pensato di dare un nome a questo tipo di insiemi.

Possiamo a questo punto iniziare a contare quanti elementi contengono alcuni degli insiemi “complicati” costruiti da insieme più elementari e introdotti fino ad ora.

Il principio di moltiplicazione ci dà subito:

PROPOSIZIONE 2.5. *Le sequenze di lunghezza m in un insieme finito A sono $|A|^m$.*

DIMOSTRAZIONE. Le sequenze di lunghezza m formano l'insieme potenza cartesiana A^m che ha cardinalità $|A|^m$ per il principio di moltiplicazione. \square

COROLLARIO 2.6. *L'insieme delle parti $\mathcal{P}(A)$ di un insieme A contiene $2^{|A|}$ elementi*

DIMOSTRAZIONE. Per la Proposizione 2.1 abbiamo che $\mathcal{P}(A)$ è in corrispondenza biunivoca con $\{0, 1\}^{|A|}$. Per la Proposizione precedente abbiamo quindi che $|\mathcal{P}(A)| = 2^{|A|}$. \square

ESEMPIO 2.7. Un insieme con 3 elementi ha 8 sottoinsiemi. Un insieme con 5 elementi ha 32 sottoinsiemi. Ci sono $2^8 = 256$ byte distinti.

Altri tipi di liste particolari saranno di grande aiuto nel calcolo delle probabilità.

4. Disposizioni e permutazioni

Andiamo ora a studiare insieme di sequenze che soddisfano determinate proprietà.

DEFINIZIONE. Una sequenza (a_1, \dots, a_k) di lunghezza k in A si dice *disposizione* di lunghezza k in A se gli elementi a_1, \dots, a_k sono tutti distinti tra loro, cioè se $a_i \neq a_j$ per ogni $i \neq j$.

L'insieme delle disposizioni è un classico esempio di prodotto condizionato.

PROPOSIZIONE 2.8. *L'insieme delle disposizioni di lunghezza k in A è un prodotto condizionato di tipo $(n, n-1, \dots, n-k+1)$ in A^k , dove $n = |A|$. Di conseguenza abbiamo esattamente $n(n-1) \cdots (n-k+1)$ disposizioni di lunghezza k in un insieme con n elementi.*

DIMOSTRAZIONE. Il primo elemento a_1 è un elemento qualunque di A e quindi può essere scelto in n modi. Scelti i primi a_1, \dots, a_{i-1} distinti tra loro l'elemento a_i può essere scelto in $n-i+1$ modi e quindi il risultato segue. \square

Il numero $n(n-1) \cdots (n-k+1)$ viene talvolta chiamato *fattoriale discendente* e denotato con $(n)_k$: per ricordarne il significato basta pensare che $(n)_k$ è il prodotto di k interi consecutivi di cui il più grande è n .

ESEMPIO 2.9. Le disposizioni di lunghezza 3 nell'insieme $\{a, b, c, d, e\}$ sono $5 \cdot 4 \cdot 3 = 60$. Alcune di esse sono (a, b, d) , $(c, d, e) \dots$. Osserviamo che nelle disposizioni, essendo sequenze, si tiene conto dell'ordine degli elementi per cui la disposizione (a, b, c) è diversa dalla disposizione (b, c, a) (gli elementi vengono “disposti” in modo diverso, notiamo sempre l'uso delle parentesi).

Caso speciale di fondamentale importanza è il caso in cui $k = n$. Una disposizione di lunghezza n in un insieme A di cardinalità n contiene tutti gli elementi di A *ordinati* in tutti i modi possibili: in questo caso le disposizioni vengono chiamate *permutazioni* e sono quindi $(n)_n = n!$.

ESEMPIO 2.10. Gli anagrammi della parola “PAROLE” sono le permutazioni dell'insieme $\{P, A, R, O, L, E\}$ e sono quindi $6! = 720$.

Si noti che in questo caso abbiamo sfruttato il fatto che le lettere che compongono la parola “parole” sono tutti distinti. Quanti sono gli anagrammi della parola “PAROLA”? E quanti quelli della parola “MAMMA”? Studieremo nella seguente sezione il problema di contare il numero di anagrammi di una parola qualunque.

5. Combinazioni, sottoinsiemi di data cardinalità e anagrammi

Abbiamo già visto che un insieme di n elementi contiene 2^n sottoinsiemi sfruttando il fatto che esistono 2^n sequenze binarie di lunghezza n . Vogliamo ora fare un conto un po' più raffinato e ci chiediamo: quante sono le sequenze binarie di lunghezza n in cui il bit 0 compare esattamente a volte e il bit 1 compare esattamente b (con $a + b = n$) volte?

Anche in questo caso abbiamo una fondamentale corrispondenza biunivoca tra sequenze binarie e sottoinsiemi.

PROPOSIZIONE 2.11. *Siano $a, b > 0$ e $n = a + b$. I seguenti insiemi*

- *sequenze binarie con a volte 0 e b volte 1,*
- *sottoinsiemi di $\{1, \dots, n\}$ con a elementi,*
- *sottoinsiemi di $\{1, \dots, n\}$ con b elementi*

sono in corrispondenza biunivoca. Ciascuno di essi ha esattamente

$$\frac{(n)_a}{a!} = \frac{(n)_b}{b!} = \frac{n!}{a!b!}$$

elementi.

DIMOSTRAZIONE. Le funzioni F e G utilizzate nella dimostrazione della Proposizione 2.1 determinano una corrispondenza biunivoca tra le sequenze binarie con a volte 0 e b volte 1, e i sottoinsiemi con b elementi: è infatti evidente che se S è una sequenza binaria con b bit uguali a 1 allora $F(S)$ è un sottoinsieme con b elementi: infatti ogni bit uguale a 1 corrisponde ad un elemento e viceversa.

I sottoinsiemi con b elementi sono in corrispondenza biunivoca con i sottoinsiemi con a elementi: per ottenere questa corrispondenza biunivoca basta considerare la corrispondenza biunivoca tra questi due insiemi di sottoinsiemi data dal complementare.

Vogliamo a questo punto contare quanti sono i sottoinsiemi di $\{1, \dots, n\}$ con a elementi. Osserviamo che ogni disposizione di lunghezza a in $\{1, \dots, n\}$ individua un sottoinsieme con a elementi:

basta considerare i suoi coefficienti. Una qualunque permutazione degli elementi di questa disposizione individua lo stesso sottoinsieme e abbiamo quindi che ogni sottoinsieme è individuato da esattamente $a!$ disposizioni. Concludiamo quindi che il numero di tali disposizioni (cioè $(n)_a$) è pari a $a!$ volte il numero di sottoinsiemi e il risultato segue. \square

Facciamo un esempio le disposizioni di lunghezza 3 in $\{1, 2, 3, 4, 5\}$ sono $(5)_3 = 5 \cdot 4 \cdot 3 = 60$ mentre i sottoinsiemi con 3 elementi sono $60/3! = 10$ questi sono $\{1, 2, 3\}$, $\{1, 2, 4\}$, $\{1, 2, 5\}$, $\{1, 3, 4\}$, $\{1, 3, 5\}$, $\{1, 4, 5\}$, $\{2, 3, 4\}$, $\{2, 3, 5\}$, $\{2, 4, 5\}$, $\{3, 4, 5\}$. Ognuno di questi 10 sottoinsiemi può essere permutato in $3! = 6$ modi dando luogo alle 60 disposizioni.

Abbiamo già osservato che la scelta di un sottoinsieme con a elementi comporta automaticamente la scelta di un sottoinsieme con $b = n - a$ elementi (il suo complementare). La scelta di un sottoinsieme con a elementi è quindi equivalente alla suddivisione dell'insieme in due sottoinsiemi, il primo di cardinalità a e il secondo di cardinalità b .

DEFINIZIONE. Una combinazione di tipo (a, b) con $a + b = n$ è una coppia (ordinata) formata da due sottoinsiemi disgiunti di $\{1, \dots, n\}$, il primo di cardinalità a e il secondo di cardinalità b .

Per questa definizione i sottoinsiemi di un insieme A con k elementi si chiamano talvolta anche *combinazioni* (di lunghezza k) in A .

Il numero di sottoinsiemi costituiti da k elementi scelti da un insieme con n elementi viene anche denotato con $\binom{n}{k}$ e viene anche detto *coefficiente binomiale* in quanto interviene nella formula della potenza di un binomio:

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

Andiamo ora a generalizzare questi concetti al caso in cui abbiamo più di due sottoinsiemi coinvolti, o equivalentemente sequenze con più di due cifre. Per non appesantire troppo la trattazione ci limitiamo al caso di sequenze ternarie (e di suddivisioni in tre sottoinsiemi).

Siano quindi a, b, c tali che $a + b + c = n$.

DEFINIZIONE. Una combinazione di tipo (a, b, c) è una terna (ordinata) di sottoinsiemi di $\{1, \dots, n\}$ a due a due disgiunti A, B, C tali che $|A| = a$, $|B| = b$, $|C| = c$.

Un anagramma di tipo (a, b, c) è una sequenza ternaria in cui lo 0 compare a volte, l'1 compare b volte e il 2 compare c volte.

LEMMA 2.12. *L'insieme delle combinazioni di tipo a, b, c forma un prodotto condizionato di tipo $\left(\binom{n}{a}, \binom{n-a}{b}, 1\right)$. Di conseguenza il numero di tali combinazioni è*

$$\binom{n}{a, b, c} := \binom{n}{a} \cdot \binom{n-a}{b} = \frac{n!}{a!(n-a)!} \frac{(n-a)!}{b!(n-a-b)!} = \frac{n!}{a!b!c!}.$$

DIMOSTRAZIONE. Sia $S = \{1, \dots, n\}$. Vediamo l'insieme delle combinazioni di tipo (a, b, c) come sottoinsieme del prodotto cartesiano $\mathcal{P}(S) \times \mathcal{P}(S) \times \mathcal{P}(S)$. Il primo sottoinsieme può essere scelto in $\binom{n}{a}$ modi (è un qualunque sottoinsieme di S con a elementi). Fissato questo primo sottoinsieme, il secondo è un qualunque sottoinsieme con b elementi scelti tra i rimanenti $n - a$. Scelti i primi due il terzo è univocamente determinato dagli elementi rimanenti. \square

PROPOSIZIONE 2.13. *Esiste una corrispondenza biunivoca tra i seguenti insiemi*

- *combinazioni di tipo* (a, b, c) ,
- *anagrammi di tipo* (a, b, c) .

Ciascuno di essi ha esattamente

$$\binom{n}{a, b, c}$$

elementi.

DIMOSTRAZIONE. La dimostrazione generalizza in modo naturale quella del caso di sequenze binarie. La corrispondenza biunivoca associa ad una sequenza ternaria $S = (s_1, \dots, s_n)$ la combinazione (A, B, C) di tipo (a, b, c) data da

$$A = \{i : s_i = 0\}, \quad B = \{i : s_i = 1\}, \quad C = \{i : s_i = 2\}.$$

□

Illustriamo questa dimostrazione con un piccolo esempio per le combinazioni e gli anagrammi di tipo $(2, 1, 1)$:

$(0, 0, 1, 2) \leftrightarrow$	$(\{1, 2\}, \{3\}, \{4\})$
$(0, 0, 2, 1) \leftrightarrow$	$(\{1, 2\}, \{4\}, \{3\})$
$(0, 1, 0, 2) \leftrightarrow$	$(\{1, 3\}, \{2\}, \{4\})$
$(0, 2, 0, 1) \leftrightarrow$	$(\{1, 3\}, \{4\}, \{2\})$
$(0, 1, 2, 0) \leftrightarrow$	$(\{1, 4\}, \{2\}, \{3\})$
$(0, 2, 1, 0) \leftrightarrow$	$(\{1, 4\}, \{3\}, \{2\})$
$(1, 0, 0, 2) \leftrightarrow$	$(\{2, 3\}, \{1\}, \{4\})$
$(2, 0, 0, 1) \leftrightarrow$	$(\{2, 3\}, \{4\}, \{1\})$
$(1, 0, 2, 0) \leftrightarrow$	$(\{2, 4\}, \{1\}, \{3\})$
$(2, 0, 1, 0) \leftrightarrow$	$(\{2, 4\}, \{3\}, \{1\})$
$(1, 2, 0, 0) \leftrightarrow$	$(\{3, 4\}, \{1\}, \{2\})$
$(2, 1, 0, 0) \leftrightarrow$	$(\{3, 4\}, \{2\}, \{1\})$

Come ci si può aspettare, vale una formula analoga se vogliamo suddividere l'insieme in più di 3 sottoinsiemi. Il numero di combinazioni di tipo (a, b, c) prende anche il nome di coefficiente trinomiale in quanto questi numeri intervengono nella potenza di un trinomio:

$$(x + y + z)^n = \sum_{a, b, c \geq 0: a+b+c=n} \binom{n}{a, b, c} x^a y^b z^c.$$

Possiamo ora contare gli anagrammi di una parola data.

ESEMPIO 2.14. Determinare il numero di anagrammi della parola ATTACCA. Siccome la parola ATTACCA è costituita da tre A, due T e due C il numero di anagrammi è dato dalle combinazioni

di tipo (3,2,2) e sono quindi

$$\binom{7}{3, 2, 2} = \frac{7!}{3! 2! 2!} = 210.$$

ESERCIZIO 2.15. Determinare il numero di anagrammi della parola ATERRATA.

I coefficienti binomiali si possono calcolare ricorsivamente con il ben noto triangolo di Tartaglia utilizzando il seguente risultato.

PROPOSIZIONE 2.16 (Formula di Stifel). *Per ogni $0 < k < n$ si ha*

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

DIMOSTRAZIONE. Questa dimostrazione si può fare in modo algebrico ma preferiamo farla in modo combinatorio. Infatti abbiamo che $\binom{n-1}{k-1}$ è il numero di sottoinsiemi con k elementi e che contengono n , mentre $\binom{n-1}{k}$ è il numero di sottoinsiemi con k elementi e che NON contengono n . Il risultato segue. \square

Utilizzando la notazione con due indici per le combinazioni di tipo (a, b) la formula di Stifel qui sopra si può riscrivere nella forma

$$\binom{n}{a, b} = \binom{n-1}{a-1, b} + \binom{n-1}{a, b-1},$$

formula che può essere facilmente generalizzata alle combinazioni di tipo a, b, c . Si ha infatti:

PROPOSIZIONE 2.17. *Siano a, b, c tali che $a + b + c = n$. Allora*

$$\binom{n}{a, b, c} = \binom{n-1}{a-1, b, c} + \binom{n-1}{a, b-1, c} + \binom{n-1}{a, b, c-1}.$$

PROOF. Consideriamo le combinazioni (A, B, C) di tipo (a, b, c) . Contiamo quelle in cui $n \in A$: in questo caso possiamo prendere una combinazione (A', B, C) di tipo $(a-1, b, c)$ per poi aggiungere n al primo sottoinsieme. Similmente possiamo contare quelle in cui $n \in B$ e quelle in cui $n \in C$. La somma di questi tre contributi è proprio la formula che dobbiamo dimostrare. \square

Osserviamo anche la seguente identità

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

che nella notazione con due indici diventa semplicemente

$$\binom{n}{a, b} = \binom{n}{b, a}.$$

Questa identità si può verificare in modo algebrico, ma ancora una volta preferiamo una verifica di carattere combinatorio: basta osservare che il passaggio al complementare è una biiezione tra i sottoinsiemi con k elementi e quelli con $n-k$ elementi.

ESEMPIO 2.18. Cammini in un reticolato. Consideriamo i cammini che uniscono l'origine con il punto (m, n) , con $m, n \in \mathbb{N}$ nel piano cartesiano costituiti da una sequenza di passi unitari verso destra e verso l'alto. Quanti sono tali cammini? È evidente che ad ogni passo verso destra l'ascissa aumenta di 1 e ad ogni passo verso l'alto l'ordinata aumenta di 1. Ne segue che complessivamente

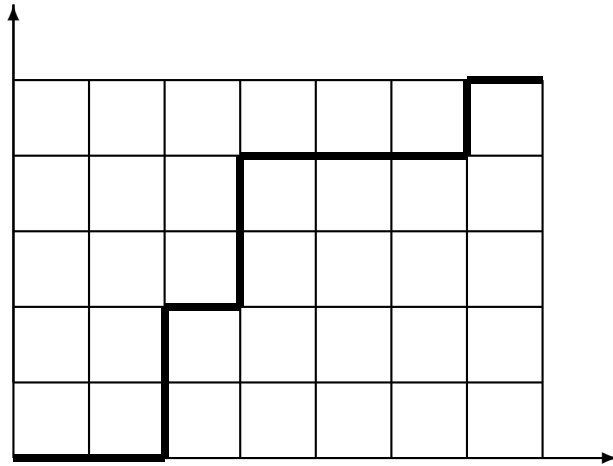


FIGURE 1. Esempio di cammino reticolare

bisogna effettuare m passi verso destra e n passi verso l'alto. Abbiamo quindi che ogni cammino corrisponde ad una sequenza binaria in cui gli 0 corrispondono ai passi verso destra e gli 1 ai passi verso l'alto in cui 0 compare esattamente m volte, e 1 compare esattamente n volte. Quindi tali cammini sono esattamente

$$\binom{m+n}{m}.$$

CHAPTER 3

Combinatoria più avanzata

Dopo aver studiato nella sezione precedente le basi della combinatoria, vogliamo ora addentrarci nello studio di alcuni strumenti un po' più sofisticati.

1. I numeri di Fibonacci

Il primo esempio che vogliamo studiare l'avete senz'altro già incontrato in altri contesti.

DEFINIZIONE. Indichiamo con F_n il numero di sottoinsiemi di $\{1, 2, \dots, n\}$ che non contengono 2 interi consecutivi. Questi numeri vengono detti di Fibonacci.

Vogliamo chiaramente un modo "facile" per calcolare i numeri di Fibonacci. Con questo scopo ricordiamo intanto la corrispondenza biunivoca tra l'insieme delle parti di $\{1, 2, \dots, n\}$ e le sequenze binarie di lunghezza n , in cui, per esempio, con $n = 6$ il sottoinsieme $\{1, 3, 6\}$ di $\{1, 2, 3, 4, 5, 6\}$ corrisponde alla sequenza 101001. Utilizzando questa corrispondenza la nostra abbiamo che F_n è anche il numero di sequenze binarie di lunghezza n in cui non compaiono due 1 consecutivi.

Si ha ad esempio $F_1 = 2$, (le sequenze binarie di lunghezza 1 in cui non ci sono due 1 consecutivi sono 2: 0 e 1) $F_2 = 3$ (qui le liste sono 3, 00, 01, 10).

Diamo ora una formula esplicita per calcolarli.

PROPOSIZIONE 3.1. *Per ogni $n \in \mathbb{N}$ si ha*

$$F_n = \sum_k \binom{n-k+1}{k}.$$

DIMOSTRAZIONE. In una sequenza binaria in cui non compaiono due 1 consecutivi, tutti gli "1" che compaiono tranne al più l'ultimo sono necessariamente seguiti da uno "0". Cancellando questi $k - 1$ "0" otteniamo una sequenza binaria di lunghezza $n - k + 1$ che ha sempre k volte "1", in quanto abbiamo cancellato solo degli "0". Ad esempio se la sequenza di partenza è 101001 otteniamo 1101. Viceversa, se io ho una qualunque sequenza di lunghezza $n - k + 1$ in cui compaiono k "1" aggiungendo uno "0" subito dopo ognuno dei primi $k - 1$ "1" ottengo una sequenza di lunghezza n in cui non compaiono due "1" consecutivi. Concludiamo che le sequenze binarie con k "1" di lunghezza n che non hanno due "1" consecutivi sono in corrispondenza biunivoca con le sequenze binarie di lunghezza $n - k + 1$ contenente esattamente k "1" e sono quindi $\binom{n-k+1}{k}$. \square

Esiste tuttavia un modo più agevole per calcolare i numeri di Fibonacci. Questo ci è fornito dal prossimo risultato.

PROPOSIZIONE 3.2. *Per ogni $n \geq 3$ si ha*

$$F_n = F_{n-1} + F_{n-2}.$$

DIMOSTRAZIONE. Chiamiamo G_n il numero di sequenze binarie di lunghezza n in cui non compaiono due 1 consecutivi e che terminano con 0 e H_n il numero di sequenze binarie di lunghezza n in cui non compaiono due 1 consecutivi e che terminano con 1. Si ha evidentemente $F_n = G_n + H_n$.

Ora, le sequenze utilizzate per G_n sono sequenze di lunghezza $n - 1$ senza 1 consecutivi a cui aggiungiamo uno 0: abbiamo quindi $G_n = F_{n-1}$. Le sequenze utilizzate per H_n devono terminare necessariamente con 01 e sono quindi formate da sequenze di lunghezza $n - 2$ senza 1 consecutivi a cui aggiungiamo 01. Abbiamo quindi $H_n = F_{n-2}$. \square

2. Il principio di inclusione-esclusione

Un'operazione elementare quale è l'unione può presentare problemi non facili da risolvere quando dobbiamo contarne gli elementi. Vorremo ora determinare la cardinalità dell'unione di due o più insiemi nel caso siano note le cardinalità delle intersezioni tra di essi. Introduciamo la funzione caratteristica $\chi_A : U \rightarrow \{0, 1\}$ di un sottoinsieme $A \subset U$. Poniamo

$$\chi_A(x) = \begin{cases} 1 & \text{se } x \in A \\ 0 & \text{se } x \notin A \end{cases}$$

per ogni $x \in U$. È chiaro che

$$|A| = \sum_{x \in U} \chi_A(x).$$

Sussiste la relazione elementare

$$\chi_{A \cup B}(x) = \chi_A(x) + \chi_B(x) - \chi_{A \cap B}(x) :$$

basta infatti mostrare che è soddisfatta da ogni $x \in U$: se $x \notin A$ e $x \notin B$ abbiamo $0 = 0 + 0 - 0$, se $x \in A$ ma $x \notin B$ abbiamo $1 = 1 + 0 - 0$, se $x \notin A$ ma $x \in B$ abbiamo $1 = 0 + 1 - 0$, e infine, se $x \in A$ e $x \in B$ abbiamo $1 = 1 + 1 - 1$. Ne segue il seguente *principio di inclusione-esclusione*:

$$|A \cup B| = |A| + |B| - |A \cap B|.$$

Questo principio ha anche una dimostrazione abbastanza intuitiva: infatti nella somma $|A| + |B|$ gli elementi che stanno sia in A che in B vengono contati *due* volte e quindi per ottenere $|A \cup B|$ dobbiamo sottrarre $|A \cap B|$. Una dimostrazione un pochino più formale la possiamo ottenere utilizzando la funzione caratteristica e la relativa interpretazione vista qui sopra:

$$\begin{aligned} |A \cup B| &= \sum_x \chi_{A \cup B}(x) \\ &= \sum_x \chi_A(x) + \sum_x \chi_B(x) - \sum_x \chi_{A \cap B}(x) \\ &= |A| + |B| - |A \cap B|. \end{aligned}$$

ESEMPIO 3.3. Quanti sono gli interi positivi minori di 100 che sono multipli di 3 o di 7? Posto $A = \{n < 100 : 3|n\}$ e $B = \{n < 100 : 7|n\}$ abbiamo $A \cap B = \{n < 100 : 21|n\}$ e quindi $|A| = 33$, $|B| = 14$ e $|A \cap B| = 4$. Di conseguenza abbiamo che i multipli di 3 o di 7 minori di 100 sono

$$|A \cup B| = |A| + |B| - |A \cap B| = 33 + 14 - 4 = 43.$$

E cosa possiamo dire nel caso di 3 insiemi? Analogamente a prima possiamo verificare che dati 3 insiemi A, B, C contenuti in un insieme U abbiamo

$$\chi_{A \cup B \cup C} = \chi_A + \chi_B + \chi_C - \chi_{A \cap B} - \chi_{A \cap C} - \chi_{B \cap C} + \chi_{A \cap B \cap C}$$

formula che può essere dimostrare facilmente caso per caso. Ad esempio se $x \in A$, $x \in B$, $x \notin C$ abbiamo $\chi_{A \cup B \cup C}(x) = 1$ e

$$\chi_A(x) + \chi_B(x) + \chi_C(x) - \chi_{A \cap B}(x) - \chi_{A \cap C}(x) - \chi_{B \cap C}(x) + \chi_{A \cap B \cap C}(x) = 1 + 1 + 0 - 1 - 0 - 0 + 0 = 1.$$

Gli altri casi possono essere verificati analogamente. Ne segue la formula

$$|A \cup B \cup C| = |A| + |B| + |C| - |A \cap B| - |A \cap C| - |B \cap C| + |A \cap B \cap C|.$$

Nel caso generale in cui abbiamo n insiemi A_1, \dots, A_n vale la formula

PROPOSIZIONE 3.4 (Principio di inclusione-esclusione). *Si ha*

$$|A_1 \cup \dots \cup A_n| = \sum_{\emptyset \neq I \subseteq \{1, \dots, n\}} (-1)^{|I|-1} \left| \bigcap_{i \in I} A_i \right|.$$

Non diamo una dimostrazione formale di questa formula, che si ottiene comunque generalizzando i casi già visti per $n = 2$ e $n = 3$ e utilizzando il principio di induzione. La sommatoria è da intendersi su tutti i sottoinsiemi di $\{1, \dots, n\}$, con esclusione del sottoinsieme vuoto. Abbiamo quindi in generale $2^n - 1$ addendi. Quindi se $n = 2, 3$ abbiamo rispettivamente 3 e 7 addendi come abbiamo già visto in precedenza.

Il simbolo di "intersecatoria" va inteso prendendo l'intersezione di tutti i sottoinsiemi con indice in I . Ad esempio, se $I = \{1, 3, 7\}$ allora

$$\bigcap_{i \in I} A_i = A_1 \cap A_3 \cap A_7.$$

Il principio di inclusione-esclusione viene spesso utilizzato per calcolare la cardinalità del complementare di un'unione. Se abbiamo A_1, \dots, A_n sottoinsiemi di un insieme U abbiamo infatti

$$\begin{aligned} |(A_1 \cup \dots \cup A_n)^C| &= |U| - |A_1 \cup \dots \cup A_n| \\ &= \left| \bigcap_{i \in \emptyset} A_i \right| - \sum_{\emptyset \neq I \subseteq \{1, \dots, n\}} (-1)^{|I|-1} \left| \bigcap_{i \in I} A_i \right| \\ &= \sum_{I \subseteq \{1, \dots, n\}} (-1)^{|I|} \left| \bigcap_{i \in I} A_i \right|, \end{aligned}$$

dove si pone per definizione $\bigcap_{i \in \emptyset} A_i = U$.

ESEMPIO 3.5. Si sa che tra 1000 persone ci sono 400 maschi, 200 bambini maschi e 300 tra bambini e bambine. Quante sono le donne adulte? In questo esempio sarà utile considerare i seguenti sottoinsiemi di $U = \{\text{tutte le persone}\}$: $A = \{\text{maschi}\}$ e $B = \{\text{bambini}\}$. L'insieme delle donne adulte è il complementare di $A \cup B$. Abbiamo quindi

$$|(A \cup B)^C| = |U| - |A \cup B| = |U| - |A| - |B| + |A \cap B| = 1000 - 400 - 300 + 200 = 500.$$

L'esempio precedente si generalizza ad ogni occasione in cui si vuole determinare la cardinalità di un'intersezione tra 2 o più insiemi, utilizzando la ben nota relazione

$$(A_1 \cap \cdots \cap A_k) = (A_1^C \cup \cdots \cup A_k^C)^C.$$

ESEMPIO 3.6. Vogliamo determinare quanti sono i numeri compresi tra 1 e 50 che non sono multipli né di 2, né di 3, né di 5. Chiamiamo A_1, A_2, A_3 rispettivamente i multipli di 2, 3, 5 compresi tra 1 e 50. Dobbiamo quindi calcolare

$$\begin{aligned} |A_1^C \cap A_2^C \cap A_3^C| &= |(A_1 \cup A_2 \cup A_3)^C| \\ &= |U| - |A_1| - |A_2| - |A_3| + |A_1 \cap A_2| + |A_1 \cap A_3| + |A_2 \cap A_3| - |A_1 \cap A_2 \cap A_3| \\ &= 50 - 25 - 18 - 10 + 8 + 5 + 3 - 1 \\ &= 12 \end{aligned}$$

E infatti questi numeri sono 1, 7, 11, 13, 17, 19, 23, 29, 31, 37, 43, 47.

Una prima applicazione importante del di inclusione-esclusione riguarda l'enumerazione delle funzioni suriettive tra due insiemi finiti. Per completezza facciamo prima il caso (molto più semplice) delle funzioni iniettive. Siano A e B due insiemi finiti di cardinalità m e n rispettivamente e senza perdere generalità assumiamo $A = \{1, \dots, m\}$ e $B = \{1, \dots, n\}$.

Le funzioni iniettive esistono se e solo se $m \leq n$. Osserviamo che ogni funzione iniettiva $f : A \rightarrow B$ definisce una disposizione in B di lunghezza m data da

$$(f(1), \dots, f(m)).$$

Questa applicazione che associa ad una funzione iniettiva una disposizione può essere facilmente invertita per cui le funzioni iniettive da A a B sono tante quante le disposizioni in B di lunghezza m , cioè $(n)_m$.

Determinare il numero di funzioni suriettive è sensibilmente più complicato. Osserviamo intanto che abbiamo necessariamente in questo caso $m \geq n$. Facciamo un piccolo esempio con $m = 5$ e $n = 3$. Una funzione iniettiva potrebbe essere data da $f(1) = 2, f(2) = 3, f(3) = 1, f(4) = 3, f(5) = 1$. Potremmo pensare di identificare questa funzione con la sequenza $(2, 3, 1, 3, 1)$ cosicché possiamo pensare alle funzioni suriettive come alle sequenze di lunghezza 5 in cui compaiono tutti i numeri tra 1 ed 3. È chiaro che anche con numeri così piccoli la risposta non è facile da ottenere. Prima di ottenere questo risultato osserviamo che il numero totale di funzioni $f = \{1, \dots, m\} \rightarrow \{1, \dots, n\}$ è n^m : abbiamo infatti n scelte per $f(1)$, n scelte per $f(2)$, \dots , n scelte per $f(m)$.

PROPOSIZIONE 3.7. Il numero di funzioni suriettive $f : \{1, \dots, m\} \rightarrow \{1, \dots, n\}$ è dato da

$$\sum_{i=0}^n (-1)^i \binom{n}{i} (n-i)^m.$$

DIMOSTRAZIONE. Poniamo $A = \{1, \dots, m\}$ e $B = \{1, \dots, n\}$. Poniamo inoltre S l'insieme delle funzioni suriettive da A a B . Per ogni $i = 1, \dots, n$ sia

$$R_i = \{f : A \rightarrow B : i \notin f(A)\}.$$

Abbiamo chiaramente che $R_1 \cup \cdots \cup R_n$ è l'insieme di tutte le funzioni da A a B che NON sono suriettive, da cui possiamo calcolare $|S|$ utilizzando il principio di inclusione-esclusione. Abbiamo

infatti per ogni $I \subseteq \{1, \dots, n\}$ che

$$\left| \bigcap_{i \in I} R_i \right| = (n - |I|)^m$$

in quanto $\bigcap_{i \in I} R_i$ sono esattamente le funzioni $f : A \rightarrow B$ le cui immagini non contengono gli elementi che stanno in I . Ne segue, per il principio di inclusione-esclusione che

$$\begin{aligned} |S| &= \sum_{I \subseteq \{1, \dots, n\}} (-1)^{|I|} \left| \bigcap_{i \in I} A_i \right| \\ &= \sum_{I \subseteq \{1, \dots, n\}} (-1)^{|I|} (n - |I|)^m \\ &= \sum_{i=0}^n (-1)^i \binom{n}{i} (n - i)^m, \end{aligned}$$

dove abbiamo sfruttato il fatto che i sottoinsiemi di $\{1, \dots, n\}$ di cardinalità i sono esattamente $\binom{n}{i}$. \square

Un'altra applicazione non banale del principio di inclusione-esclusione è data dagli scombussolementi.

DEFINIZIONE. Uno scombussolemento è una permutazione dell'insieme $\{1, 2, \dots, n\}$ in cui nessun elemento rimane al proprio posto.

Ad esempio 4213 non è uno scombussolemento perché il numero 2 è al posto 2, mentre 4123 è uno scombussolemento. Per $n = 3$ abbiamo solo due scombussolementi, cioè 312 e 231. Ma quanti sono in generale? Il principio di inclusione-esclusione ci permette ancora una volta di rispondere a questa domanda.

PROPOSIZIONE 3.8. *Il numero di scombussolementi di lunghezza n è*

$$\sum_{j=0}^{n-2} (-1)^{n-j} \binom{n}{j}.$$

DIMOSTRAZIONE. Chiamiamo S l'insieme degli scombussolementi e A_i l'insieme delle permutazioni che fissano l'elemento i . Abbiamo, come nel caso delle funzioni suriettive, che S è il complementare di $A_1 \cup \dots \cup A_n$. Si ha inoltre, per ogni $I \subseteq \{1, \dots, n\}$ che

$$\left| \bigcap_{i \in I} A_i \right| = (n - |I|)! :$$

questo segue dal fatto che le permutazioni di $\bigcap_{i \in I} A_i$ sono quelle che fissano tutti gli elementi di I e permutano i rimanenti $n - |I|$ elementi arbitrariamente. Ne segue che

$$\begin{aligned} |S| &= \sum_{I \subseteq \{1, \dots, n\}} (-1)^{|I|} \left| \bigcap_{i \in I} A_i \right| \\ &= \sum_{I \subseteq \{1, \dots, n\}} (-1)^{|I|} (n - |I|)! \\ &= \sum_{i=0}^n (-1)^i \binom{n}{i} (n - i)! \\ &= \sum_{i=0}^n (-1)^i (n)_{n-i}. \end{aligned}$$

Osservando infine che $(n)_n = (n)_{n-1}$ possiamo semplificare gli ultimi due termini in questa sommatoria e cambiando l'indice di sommatoria in $j = n - i$ otteniamo

$$|S| = \sum_{i=2}^n (-1)^i (n)_{n-i} = \sum_{j=0}^{n-2} (-1)^{n-j} (n)_j = (n)_{n-2} - (n)_{n-3} + \dots \pm (n)_0.$$

□

ESEMPIO 3.9. In quanti modi 6 persone si possono scambiare il proprio cappello in modo che nessuno riceva il proprio? La risposta è data dagli scombussolamenti su un insieme di 6 elementi e sono quindi

$$\sum_{i=0}^4 (-1)^i (6)_{6-i} = (6)_4 - (6)_3 + (6)_2 - (6)_1 + (6)_0 = 360 - 120 + 30 - 6 + 1 = 265.$$

3. Partizioni di un insieme

Dato un insieme A abbiamo più volte parlato implicitamente di partizione di A ogni qualvolta abbiamo utilizzato il principio di somma. Diamo ora la definizione formale.

DEFINIZIONE. Sia A un insieme. Un insieme di sottoinsiemi non vuoti $\{A_1, A_2, \dots\}$ (finito o infinito) di A si dice *partizione* di A se

- $A = A_1 \cup A_2 \cup \dots$.
- $A_i \cap A_j = \emptyset$ per ogni $i \neq j$.

In altre parole vogliamo che ogni elemento di A sia contenuto in esattamente uno dei sottoinsiemi A_1, A_2, \dots . I sottoinsiemi A_1, A_2, \dots si dicono *parti* o *blocchi* della partizione.

ESEMPIO 3.10. Dato $A = \{1, 2, 3, 4, 5\}$ consideriamo i sottoinsiemi $B = \{1, 2, 3\}$, $C = \{3, 4\}$, $D = \{4, 5\}$. Allora $\{B, D\}$ è una partizione di A , mentre $\{B, C, D\}$ non lo è.

ESEMPIO 3.11. Per ogni numero primo p poniamo $A_p = \{n \geq 2 : p \text{ è il più grande numero primo che divide } n\}$. Abbiamo ad esempio

$$\begin{aligned} A_2 &= \{2^a : a > 0\} \\ A_3 &= \{2^a 3^b : a \geq 0, b > 0\} \\ A_5 &= \{2^a 3^b 5^c : a, b \geq 0, c > 0\} \end{aligned}$$

Si verifica facilmente che i sottoinsiemi A_p formano una partizione di $\{2, 3, 4, \dots\}$. Notiamo che in questo esempio abbiamo un numero infinito di blocchi, ognuno dei quali contiene un numero infinito di elementi.

È evidente che se A è infinito ammetterà infinite partizioni. Ci vogliamo quindi concentrare sul caso finito.

DEFINIZIONE. Il numero di partizioni di $\{1, 2, \dots, n\}$ è detto numero di Bell e si indica con B_n .

- Se $A = \{a\}$ ha un solo elemento, allora A ammette un'unica partizione data da $\{\{a\}\}$, e quindi $B_1 = 1$
- Se $A = \{a, b\}$ ha due elementi allora A ammette due partizioni: $\{\{a, b\}\}$ e $\{\{a\}, \{b\}\}$ e quindi $B_2 = 2$.
- Se $A = \{a, b, c\}$ ha tre elementi allora A ammette 5 partizioni e queste sono: $\{\{a, b, c\}\}$, $\{\{a, b\}, \{c\}\}$, $\{\{a, c\}, \{b\}\}$, $\{\{a\}, \{b, c\}\}$, $\{\{a\}, \{b\}, \{c\}\}$ e quindi $B_3 = 5$.

Abbiamo già calcolato $B_1 = 1$, $B_2 = 2$ e $B_3 = 5$. Poniamo anche $B_0 = 1$ per convenzione. Vogliamo ora determinare una formula ricorsiva per calcolare B_n .

PROPOSIZIONE 3.12. Per ogni $n > 0$ abbiamo

$$B_n = \sum_{h=0}^{n-1} \binom{n-1}{h} B_h.$$

DIMOSTRAZIONE. Fissiamo $n > 0$ e consideriamo le partizioni di $\{1, 2, \dots, n\}$ in cui n compare in un sottoinsieme di cardinalità k . Chiamiamo P_k l'insieme di tali partizioni. Avremo quindi

$$B_n = \sum_{k=1}^n |P_k|$$

per il principio di somma. Per determinare $|P_k|$ procediamo nel modo seguente: gli altri $k-1$ elementi del sottoinsieme cui appartiene n possono essere scelti in $\binom{n-1}{k-1}$ modi. Per ognuna di tali scelte, i rimanenti $n-k$ elementi possono essere partizionati in B_{n-k} modi per cui P_k è un prodotto condizionato di tipo $\left(\binom{n-1}{k-1}, B_{n-k}\right)$. Abbiamo quindi $|P_k| = \binom{n-1}{k-1} B_{n-k}$ da cui

$$B_n = \sum_{k=1}^n \binom{n-1}{k-1} B_{n-k},$$

dove abbiamo posto per convenzione $B_0 = 1$. Sostituendo $n-k$ con h otteniamo la formula equivalente

$$B_n = \sum_{h=0}^{n-1} \binom{n-1}{h} B_h,$$

che permette di calcolare facilmente i numeri di Bell in maniera ricorsiva. \square

Questa formula ricorsiva è quella che permette di calcolare i numeri di Bell con il cosiddetto triangolo di Bell. Andiamo a vedere di cosa si tratta. Il triangolo di Bell è dato da

$$\begin{array}{cccccc} b_{00} & & & & & \\ b_{10} & b_{11} & & & & \\ b_{20} & b_{21} & b_{22} & & & \\ b_{30} & b_{31} & b_{32} & b_{33} & & \\ b_{40} & b_{41} & b_{42} & b_{43} & b_{44} & \end{array}$$

Dove

$$b_{nm} = \sum_{k=0}^m \binom{m}{k} B_{n-k}.$$

Osserviamo che sulla prima colonna compaiono i numeri di Bell: infatti $b_{n0} = \sum_{k=0}^0 \binom{0}{k} B_{n-k} = B_n$ per definizione. Inoltre anche l'ultimo elemento di ogni riga è un numero di Bell:

$$b_{nn} = \sum_{k=0}^n \binom{n}{k} B_{n-k} = B_{n+1}$$

per la Proposizione precedente, e quindi l'ultimo numero di una riga è uguale al primo della riga successiva. Usando la formula di Stifel si può inoltre verificare che ogni coefficiente che compare nel triangolo di Bell è la somma di quello alla sua sinistra e di quello sopra a quest'ultimo, cioè

$$b_{ij} = b_{i,j-1} + b_{i-1,j-1}.$$

Questo fornisce una semplice procedura per determinare il triangolo di Bell facendo delle piccole somme di volta in volta ottenendo il triangolo

$$\begin{array}{cccccc} 1 & & & & & \\ 1 & 2 & & & & \\ 2 & 3 & 5 & & & \\ 5 & 7 & 10 & 15 & & \\ 15 & 20 & 27 & 37 & 52 & \end{array}$$

Possiamo raffinare il calcolo del numero di partizioni di un insieme finito se vogliamo tener conto anche del numero di sottoinsiemi coinvolti, cioè vogliamo fissare il numero di blocchi di una partizione.

DEFINIZIONE. Chiamiamo numero di Stirling (di seconda specie) il numero $S_{n,k}$ di partizioni di $\{1, 2, \dots, n\}$ costituite da esattamente k blocchi.

Per convenzione si pone $S_{0,0} = 1$ e $S_{0,k} = 0$ per ogni $k > 0$. Abbiamo una prima formula esplicita per calcolare questi numeri

PROPOSIZIONE 3.13. *Per ogni $k \leq n$ abbiamo*

$$S_{n,k} = \sum_{i=0}^k (-1)^i \frac{(k-i)^n}{i!(k-i)!}.$$

DIMOSTRAZIONE. Per dimostrare questa formula vogliamo utilizzare un naturale legame tra i numeri di Stirling e il numero di funzioni suriettive tra due insiemi. Siano quindi $A = \{1, \dots, n\}$ e $B = \{1, \dots, k\}$. Ogni funzione suriettiva $f : A \rightarrow B$ determina una partizione di A in k sottoinsiemi: questi sottoinsiemi sono dati dalle controimmagini degli elementi di f . Ad esempio se $f : \{1, 2, 3, 4, 5\} \rightarrow \{0, 1\}$ è data da $f(1) = f(2) = f(5) = 1$ e $f(4) = f(3) = 0$, la partizione associata ad f è $\{\{1, 2, 5\}, \{3, 4\}\}$. Il numero di funzioni che determinano la stessa partizione è esattamente $k!$ (corrispondenti a tutte le possibili permutazioni dei k blocchi) e quindi abbiamo che il numero di funzioni suriettive è

$$S_{n,k}k!$$

Utilizzando la formula ottenuta precedentemente per il numero di funzioni suriettive tra due insiemi, possiamo quindi ricavare la seguente formula esplicita per i numeri di Stirling:

$$S_{n,k} = \sum_{i=0}^k (-1)^i \frac{1}{k!} \binom{k}{i} (k-i)^n = \sum_{i=0}^k (-1)^i \frac{(k-i)^n}{i!(k-i)!}.$$

□

Anche i numeri di Stirling soddisfano una formula ricorsiva che li rende facili da calcolare. Si ha infatti

PROPOSIZIONE 3.14. *Per ogni $n, k > 1$ si ha*

$$S_{n,k} = kS_{n-1,k} + S_{n-1,k-1}.$$

DIMOSTRAZIONE. Ancora una volta facciamo ricorso al principio di somma. Consideriamo le partizioni in k blocchi di cui una è $\{n\}$ e le partizioni in k blocchi in cui n compare in un sottoinsieme che contiene almeno un altro elemento. Quelli del primo tipo sono esattamente $S_{n-1,k-1}$ (gli altri $n-1$ elementi devono essere partizionati in $k-1$ blocchi). Quelli del secondo tipo si ottengono considerando una qualunque partizione di $\{1, 2, \dots, n-1\}$ in k blocchi per poi aggiungere n ad uno di questi k blocchi. Otteniamo quindi $kS_{n-1,k}$ partizioni. □

CHAPTER 4

Statistica descrittiva

Come accennato nell'introduzione la statistica descrittiva si occupa di presentare i dati raccolti in una certa indagine. Abbiamo innanzitutto di introdurre una apposita terminologia.

1. Popolazione, campione e caratteri

Si prende in considerazione un certo insieme (finito o infinito) che chiamiamo *popolazione* da analizzare. Ogni elemento della popolazione viene chiamato *unità statistica*. La proprietà che si vuole analizzare si dice *carattere*: un carattere da un punto di vista matematico è semplicemente una funzione che ha come dominio la popolazione.

ESEMPIO 4.1. Popolazione={Studenti di ISI} e Carattere="Materia preferita"
Popolazione={Studenti di ISI} e Carattere="Voto conseguito alla maturità"
Popolazione={Corsi di laurea} e Carattere="Rapporto tra iscritti e laureati"
Popolazione={Numeri interi} e Carattere="Prima lettera del loro nome"
Popolazione={Numeri interi} e Carattere="La radice quadrata"

ESEMPIO 4.2. Popolazione={Tutti i possibili lanci di una moneta} e Carattere="Testa o croce"
Popolazione={Tutti i possibili lanci di una moneta} e Carattere="Tempo necessario all'arresto"
Popolazione={Tutti i possibili lanci di una moneta} e Carattere="Numero di oscillazioni prima dell'arresto"
Popolazione={Tutti i possibili lanci di una moneta} e Carattere="Superficie finale di arresto"

Un carattere si dice *quantitativo* se assume valori numerici che rappresentano una "misura" del carattere stesso. Caratteri non quantitativi si dicono *qualitativi*. I caratteri quantitativi che ci interessano possono essere finiti, numerabili o continui, a seconda della cardinalità dell'insieme dei valori che può assumere.

I valori assunti (o che possono essere assunti) dal carattere si dicono *modalità*.

Un campione è un sottoinsieme della popolazione su cui conosciamo (o vogliamo conoscere) il valore assunto dal carattere in esame. Il campione è quindi l'oggetto dell'indagine statistica.

La frequenza (assoluta) di una certa modalità è il numero di volte che si presenta all'interno del nostro campione.

ESEMPIO 4.3. 10 studenti all'esame di MDP. Carattere="voto ottenuto allo scritto". I voti ottenuti dai 10 studenti sono 15,21,24,21,26,16,24,21,18,15. In questo caso il carattere è quantitativo. Le modalità sono 31 (tutti i voti da 0 a 30). Abbiamo ad esempio che la modalità 15 ha frequenza 2 e la modalità 21 ha frequenza 3.

Solitamente indichiamo con i simboli m_1, \dots, m_k le modalità e con i simboli n_i le rispettive frequenze.

Le frequenze relative le indichiamo invece con i simboli p_i e indicano la proporzione in cui si presenta la modalità m_i . Nell'esempio precedente la modalità 15 ha frequenza relativa 0.2 e la modalità 21 ha frequenza relativa 0.3. Osserviamo che la somma delle frequenze relative è 1.

2. Classi e istogrammi

Se le modalità sono molto numerose (o addirittura infinite) può essere conveniente raggrupparle in classi.

ESEMPIO 4.4. Voto conseguito alla maturità dagli iscritti a ISI. Non sarà conveniente considerare tutti i possibili voti da 60 a 100, ma sarà più opportuno raggrupparli in classi, ad esempio 60-69, 70-79,...,90-100

ESEMPIO 4.5. Campione di reclute in cui valutiamo il carattere altezza. Le modalità sono infinite, o comunque molto numerose, (tutte le possibili altezze) e quindi sarà senz'altro indicato suddividere i dati in classi.

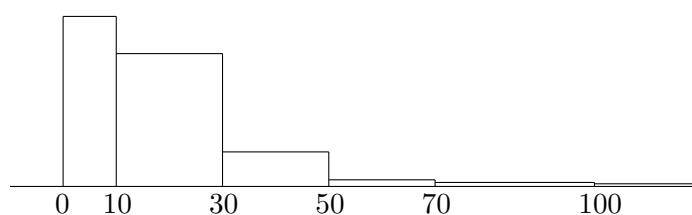
Le classi possono essere considerate come delle nuove modalità e pertanto avranno le loro frequenze assolute e relative. Una classe viene solitamente descritta dai confini, inferiore e superiore. L'ampiezza è la differenza tra il confine superiore e quello inferiore, e il valore centrale di una classe è il punto medio dei suoi confini.

L'*istogramma* è un grafico che permette di visualizzare alcune informazioni relative ad un carattere. Ad ogni modalità viene associato un rettangolo avente *area* proporzionale alla frequenza della modalità stessa. Se le modalità sono state raggruppate in classi si richiede che la base del rettangolo sia proporzionale all'ampiezza della classe, in tutti gli altri casi le basi dei rettangoli possono essere prese uguali.

ESEMPIO 4.6. Campione di 200 contribuenti suddivisi in base al reddito lordo annuo:

> 100.000 Eur	2
70.000–100.000	3
50.000–70.000	5
30.000–50.000	26
10.000–30.000	100
< 10.000	64

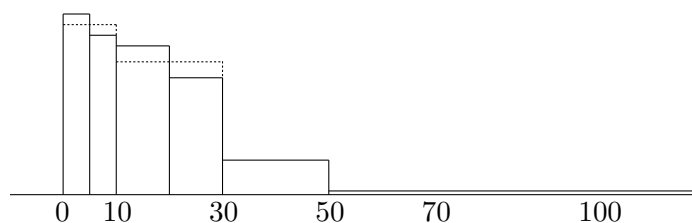
Osserviamo la prima classe ha un rettangolo più alto e una frequenza minore della seconda classe. Questo deriva dal fatto che nella prima classe c'è una maggior concentrazione di contribuenti. Osserviamo inoltre che le ultime tre classi hanno una frequenza molto bassa rispetto alle altre per cui potrebbero essere raggruppate in un'unica classe, mentre le prime due potrebbero essere ulteriormente suddivise per avere un diagramma che meglio descrive la situazione. Potremmo cioè



considerare la seguente suddivisione in classi

> 50.000 Eur	10
30.000–50.000	26
20.000–30.000	44
10.000–20.000	56
5.000–10.000	30
< 5.000	34

L'istogramma che ne risulterebbe è il seguente e rappresenta i dati raccolti in modo molto più preciso (i rettangoli relativi alle prime due classi del vecchio istogrammi sono ancora visibili tratteggiate).



3. Indici di posizione o centralità

Gli indici di posizione sono singoli valori che sintetizzano la distribuzione di un certo carattere. Ce ne sono di tanti tipi a seconda dell'occorrenza e di ciò che si vuole comunicare. L'esempio più semplice da considerare è il seguente che ha sempre senso, anche se il carattere è qualitativo.

DEFINIZIONE. La *moda* è data dalla modalità che ha la frequenza massima. Nel caso di raggruppamento in classi si parla di classe modale.

Ma l'indice di posizione di gran lunga più importante è il seguente.

DEFINIZIONE. Sia x un carattere quantitativo definito su un campione di n elementi. Chiamiamo x_1, \dots, x_n i valori assunti da x . Ad esempio, se x è l'altezza, x_1 è l'altezza del primo elemento del campione, x_2 del secondo e così via. La *media campionaria*, o più semplicemente *media*, è

$$\bar{x} = \bar{x}_n = \frac{1}{n}(x_1 + \dots + x_n).$$

Calcoliamo la media campionaria nell'Esempio 4.3. Abbiamo in questo caso

$$\bar{x} = \frac{1}{10}(15 + 21 + 24 + 21 + 26 + 16 + 24 + 21 + 18 + 15) = \frac{201}{10} = 20.1.$$

Facciamo ora qualche semplice osservazione sulle sommatorie che utilizzeremo tante volte nel corso.

- Per la proprietà commutativa della somma si ha

$$\sum_{i=1}^n (a_i + b_i) = \sum_{i=1}^n a_i + \sum_{i=1}^n b_i;$$

- per la definizione di sommatoria si ha che se c è una costante si ha

$$\sum_{i=1}^n c = nc;$$

- per la proprietà distributiva si ha

$$\sum_{i=1}^n ca_i = c \sum_{i=1}^n a_i.$$

- generalizzando la precedente, applicando ancora la proprietà distributiva abbiamo

$$\sum_{i=1}^n \sum_{j=1}^m a_i b_j = \left(\sum_{i=1}^n a_i \right) \left(\sum_{j=1}^m b_j \right).$$

Vogliamo ora studiare come si comporta la media rispetto ad un cambio di carattere del tipo $y = ax + b$. Abbiamo cioè un carattere x che assume i valori x_1, \dots, x_n e un carattere y che assume i valori y_1, \dots, y_n legati dalla relazione $y_i = ax_i + b$ per ogni $i = 1, \dots, n$. In tal caso le medie corrispondenti sono legate dalla medesima relazione: $\bar{y} = a\bar{x} + b$. Infatti, utilizzando anche le osservazioni precedenti sulle sommatorie abbiamo

$$\begin{aligned} \bar{y} &= \frac{1}{n} \sum_{i=1}^n y_i = \frac{1}{n} \sum_{i=1}^n (ax_i + b) \\ &= \frac{1}{n} \sum_{i=1}^n ax_i + \frac{1}{n} \sum_{i=1}^n b. \\ &= a\bar{x} + b \end{aligned}$$

Più in generale si può considerare un carattere $z = ax + by$ dato dalla combinazione lineare di due caratteri.

PROPOSIZIONE 4.7. *Siano x, y, z tre caratteri definiti sullo stesso campione e legati dalla relazione $z = ax + by$. Allora*

$$\bar{z} = a\bar{x} + b\bar{y}.$$

DIMOSTRAZIONE. La dimostrazione è una semplice generalizzazione del caso visto in precedenza. Supponiamo che il campione sia costituito da n elementi e chiamiamo x_i, y_i, z_i , con $i = 1, \dots, n$ i rispettivi valori che assumono i caratteri x, y, z . Utilizzando ancora le proprietà elementari delle sommatorie si ha

$$\begin{aligned}\bar{z} &= \frac{1}{n} \sum_{i=1}^n z_i = a \frac{1}{n} \sum_{i=1}^n x_i + b \frac{1}{n} \sum_{i=1}^n y_i \\ &= a\bar{x} + b\bar{y}.\end{aligned}$$

□

ESEMPIO 4.8. Siano 50, 55, 53, 44, 43, 48, 50 le temperature registrate a New York in una settimana, misurate in gradi Fahrenheit. Qual è stata la temperatura media in gradi Celsius? Se x indica la temperatura in F e y in C si ha $y = (x - 32) \frac{100}{180} = \frac{5}{9}x - \frac{160}{9}$. In questo caso abbiamo $x_1 = 50, x_2 = 55, \dots$ da cui

$$\bar{x} = \frac{1}{7}(50 + 55 + 53 + 44 + 43 + 48 + 50) = \frac{343}{7} = 49$$

e quindi $\bar{y} = (49 - 32) \frac{100}{180} = 9.44C$. Alternativamente si poteva prima trasformare tutte le temperature in gradi Celsius per poi fare la media: avremmo ottenuto lo stesso risultato, ma facendo dei conti inutili.

ESEMPIO 4.9. Si determini la media campionaria dei seguenti valori registrati in un certo campione: 180, 190, 150, 210, 250, 160, 140, 140, 210, 220. Chiamiamo x il carattere in esame. Considerando che tutti i valori sono multipli di 10 e sono tutti intorno al valore 200, per determinare la media \bar{x} può essere utile considerare un cambio di variabile $x = 10y + 200$ e quindi $y = (x - 200)/10$. I valori del carattere y così definito sono $-2, -1, -5, 1, 5, -4, -6, -6, 1, 2$. Possiamo quindi agevolmente calcolare la media di y : $\bar{y} = -15/10 = -1.5$ e quindi $\bar{x} = 10\bar{y} + 200 = 185$.

Vogliamo ora esprimere la media campionaria in funzione delle frequenze assolute e delle frequenze relative. Siano m_1, \dots, m_k le modalità, n_i e p_i le relative frequenze assolute e relative. Questo vuol dire che tra i valori x_1, \dots, x_n assunti dal carattere troviamo n_1 volte m_1 , n_2 volte m_2 e così via. Abbiamo quindi

$$x_1 + \dots + x_n = n_1 m_1 + \dots + n_k m_k$$

e quindi

$$\bar{x}_n = \frac{1}{n}(n_1 m_1 + n_2 m_2 + \dots + n_k m_k).$$

Ricordando poi che le frequenze assolute e relative sono legate dalla relazione $p_i = n_i/n$ abbiamo anche

$$\bar{x} = p_1 m_1 + p_2 m_2 + \dots + p_k m_k.$$

Se i dati sono raggruppati in classi la media campionaria viene calcolata considerando il valore centrale delle classi. Se questo valore non dovesse essere definito possiamo prendere un valore a nostro modo di vedere rappresentativo della classe. Ad esempio nella prima suddivisione in classi dell'Esempio 4.6 possiamo rappresentare la classe > 100.000 dal valore 110.000 e calcolare quindi

$$\bar{x} = \frac{1000}{200}(64 \cdot 5 + 100 \cdot 20 + 26 \cdot 40 + 5 \cdot 60 + 3 \cdot 85 + 2 \cdot 110) = 5 \cdot 4135 = 20.675$$

Utilizzando la seconda divisione in classi potremmo prendere il valore 70.000 come rappresentativo della classe > 50.000 e ottenere

$$\bar{x} = \frac{1000}{200}(34 \cdot 2.5 + 30 \cdot 7.5 + 56 \cdot 15 + 44 \cdot 25 + 26 \cdot 40 + 10 \cdot 70) = 5 \cdot 3990 = 19.950.$$

Consideriamo due campioni disgiunti costituiti da n e m elementi ciascuno e vogliamo considerare la media campionaria sull'unione di questi due campioni. Consideriamo quindi su di essi lo “stesso” carattere (o per meglio dire due caratteri aventi la stessa unità di misura). Denotiamo con x e y i caratteri sui due campioni, mentre con z il carattere sull'unione dei due campioni. Non è vero in generale che la media di z sia il valore medio delle medie di x ed y , (cioè non è vero che $\bar{z} = \frac{1}{2}(\bar{x} + \bar{y})$). Si ha infatti

$$\bar{z} = \frac{1}{n+m} \sum_{i=1}^{n+m} = \frac{1}{n+m} \left(\sum_{i=1}^n x_i + \sum_{j=1}^m y_j \right) = \frac{n\bar{x} + m\bar{y}}{n+m}.$$

ESEMPIO 4.10. In un gruppo di amici ci sono 13 maschi e 8 femmine. Chiediamo a tutti quanti fratelli hanno. Dei maschi tre ne hanno 0, sette 1, due 2 e uno 3. tra le femmine due ne hanno 0, cinque 1, e una 2. Calcolare il numero medio di fratelli dei maschi, delle femmine e complessivo. Dei maschi $14/13$ delle femmine $7/8$. Complessivamente $\frac{14+7}{21} = 1$.

Una variante della media campionaria, detta *media ponderata*, si utilizza se si vuole assegnare un'importanza (o peso) diverso ai vari valori ottenuti nel campionamento. In particolare si assegna un *peso* diverso ad ogni elemento del campione. Diciamo che se il campione è costituito da n elementi numerati da 1 a n assegniamo peso w_i all'elemento i . La media ponderata si definisce nel seguente modo:

$$\bar{x}_w = \frac{w_1x_1 + \cdots + w_nx_n}{w_1 + \cdots + w_n}.$$

ESEMPIO 4.11. Vogliamo fare la media ponderata dei voti ottenuti da uno studente fino ad ora, scegliendo come pesi il numero di CFU corrispondenti ad ogni corso. I voti ottenuti sono riportati nella seguente tabella:

Corso	CFU	Voto
ALGEBRA E GEOMETRIA	6	19
PROGRAMMAZIONE	12	28
ARCHITETTURE	12	24
ANALISI MATEMATICA	6	20
ALGORITMI	12	27
SISTEMI OPERATIVI	12	28
CALCOLO	6	22
PARADIGMI	9	24

La media ponderata è quindi

$$\frac{6 \cdot 19 + 12 \cdot 28 + 12 \cdot 24 + 6 \cdot 20 + 12 \cdot 27 + 12 \cdot 28 + 6 \cdot 22 + 9 \cdot 24}{75} = \frac{1866}{75} = 24.88.$$

4. Indici di dispersione

Gli indici di dispersione servono ad esprimere la compattezza o la rarefazione dei dati. A maggior indice di dispersione corrispondono valori assunti dal carattere più rarefatti. Ad esempio, i campioni $-1, 0, 1$ e $-100, 50, 50$ hanno entrambi media campionaria 0, ma vogliamo trovare un modo per esprimere il fatto che il primo campione è molto più compatto rispetto al secondo. L'indice di dispersione più semplice è il cosiddetto *campo di variazione* o *range*: è dato dalla differenza tra il valore più alto e il valore più basso ottenuti nel campione. Nel primo campione tale indice vale 2, mentre nel secondo vale 150.

Andiamo ora ad introdurre l'indice di dispersione più importante da conoscere. Consideriamo quindi un carattere x e denotiamo con x_i i valori che assume nel campione di riferimento.

Scegliamo un punto t sulla retta reale e calcoliamo la media dei quadrati delle distanze di t dagli x_i . Nell'esempio $-1, 0, 1$ avremmo $\frac{1}{3}((t+1)^2 + t^2 + (t-1)^2)$. In generale otteniamo la funzione

$$t \mapsto \frac{1}{n} \sum_{i=1}^n (x_i - t)^2.$$

Vogliamo determinare il valore di t per cui questa quantità è più piccola possibile: tale valore minimo sarà l'indice di dispersione che vogliamo introdurre. Ricordiamo ora che il grafico della funzione $y = ax^2 + bx + c$, con $a > 0$, è una parabola con concavità rivolta verso l'alto e che il vertice, che rappresenta un minimo assoluto, ha ascissa $-b/2a$. Applicando questo fatto abbiamo che la funzione considerata

$$(1) \quad t \mapsto \frac{1}{n} \sum_{i=1}^n (x_i - t)^2 = t^2 - 2\bar{x}t + \frac{1}{n} \sum_{i=1}^n x_i^2$$

ammette un minimo per $t = \bar{x}$ (in alternativa basta fare la derivata) e questo valore minimo si chiama *varianza campionaria*:

$$\sigma_x^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2.$$

Osserviamo quindi che la varianza campionaria è la media dei quadrati delle distanze dei valori assunti rispetto alla loro media: se i valori sono mediamente lontani dalla media campionaria avremo quindi una varianza maggiore e viceversa, se i valori assunti sono vicini alla media campionaria, avremo una varianza campionaria minore.

Lo *scarto quadratico medio* (o deviazione standard) campionario è la radice quadrata della varianza campionaria e lo indichiamo quindi con σ_x . Per il calcolo della varianza campionaria può essere conveniente utilizzare la seguente formula

PROPOSIZIONE 4.12. *Si ha*

$$\sigma_x^2 = \overline{x^2} - \bar{x}^2.$$

DIMOSTRAZIONE. Utilizzando l'equazione (1) con $t = \bar{x}$ otteniamo

$$\begin{aligned} \sigma_x^2 &= \bar{x}^2 - 2\bar{x}^2 + \frac{1}{n} \sum_{i=1}^n x_i^2 \\ &= \overline{x^2} - \bar{x}^2. \end{aligned}$$

□

Abbiamo detto come la varianza sia una misura della dispersione dei dati. In generale, per fare una grezza analisi qualitativa, ci dobbiamo aspettare che buona parte dei dati ricadano in un intervallo di ampiezza tra cinque e dieci volte σ_x .

ESERCIZIO 4.13. In una stazione meteorologica in Siberia è stato registrato lo spessore del ghiaccio in centimetri in 21 date ottenendo i seguenti dati: 53, 60, 65, 66, 78, 77, 83, 88, 90, 92, 102, 105, 106, 111, 112, 108, 110, 106, 114, 111, 110. Abbiamo quindi

$$\bar{x} = \frac{649}{7} \sim 92,71429.$$

I quadrati dei valori assunti dal carattere sono 2809, 3600, 4225, 4356, 6084, 5929, 6889, 7744, 8100, 8464, 10404, 11025, 11236, 12321, 12544, 11664, 12100, 11236, 12996, 12321, 12100 e quindi si ha

$$\overline{x^2} = \frac{188147}{21} = 8959,381.$$

Concludiamo che

$$\sigma_x^2 = 8959,381 - (92,71429)^2 = 363.$$

Osserviamo nell'esempio precedente che la scarto quadratico medio è circa 19.

Cosa succede alla varianza campionaria per i cambi di variabile lineari $y = ax + b$? Il "+b" rappresenta una traslazione dei dati e quindi ci aspettiamo che non modifichi la dispersione dei dati stessi. La moltiplicazione per a invece ci aspettiamo che "allontani" i dati se $a > 1$ e li avvicini se $0 < a < 1$. Questa osservazione è confermata dalla seguente.

PROPOSIZIONE 4.14. *Siano x ed y due caratteri definiti sullo stesso campione e legati dalla relazione $y = ax + b$ (cioè abbiamo $y_i = ax_i + b$ per ogni i). Allora*

$$\sigma_y^2 = a^2 \sigma_x^2.$$

DIMOSTRAZIONE. Utilizziamo la formula della Proposizione 4.12 per calcolare la varianza di y e le proprietà della media della Proposizione 4.7. Abbiamo:

$$\begin{aligned} \sigma_y^2 &= \overline{y^2} - \bar{y}^2 = \overline{(a^2x^2 + b^2 + 2abx)} - \overline{ax + b}^2 \\ &= a^2\overline{x^2} + b^2 + 2ab\bar{x} - (a\bar{x} + b)^2 \\ &= a^2(\overline{x^2} - \bar{x}^2) = a^2\sigma_x^2. \end{aligned}$$

□

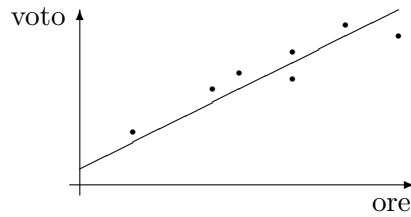
5. Correlazione

In questa sezione studiamo il legame che ci può essere tra due caratteri (quantitativi) definiti sullo stesso campione. Studieremo in particolare quando esiste una sorta di proporzionalità diretta (correlazione positiva) o inversa (correlazione negativa) tra di essi. Studiamo questo problema tenendo a mente il seguente esempio guida.

In un campione di 7 studenti che fanno l'esame di MDP valutiamo il carattere x = "numero di ore dedicate alla preparazione dell'esame" e y = "voto ottenuto all'esame". I risultati ottenuti sono riportati nella tabella a sinistra.

ore	voto	ore ²	voto ²	ore · voto
10	10	100	100	100
60	28	3600	784	1680
50	30	2500	900	1500
40	25	1600	625	1000
40	20	1600	400	800
30	21	900	441	630
25	18	625	324	450

Il *diagramma a dispersione* si ottiene riportando in un piano cartesiano questi valori, rappresentati dai puntini nella seguente figura.



In questo caso abbiamo deciso di riportare le ore nelle ascisse e il voto nelle ordinate perchè vogliamo studiare come il numero di ore di studio influenzi il voto ottenuto. Se volessimo studiare come il voto ottenuto influenzi il numero di ore studiate avremmo fatto la scelta opposta.

Lo scopo è quello di determinare una retta che approssimi nel migliore dei modi possibili i dati e prima di procedere dobbiamo dare un senso preciso a questa affermazione.

DEFINIZIONE. La *retta ai minimi quadrati* è la retta di equazione $y = ax + b$ che minimizza la media (o equivalentemente la somma) dei quadrati degli errori che si commettono stimando i valori y_i del secondo carattere con $ax_i + b$.

Il nostro prossimo scopo è quello di determinare i parametri a della retta ai minimi quadrati in funzione dei dati a disposizione sui due caratteri. Considerando una retta generica di equazione $y = ax + b$ dipendente dai parametri a e b si vuole cioè minimizzare la quantità

$$S(a, b) = \frac{1}{n} \sum_{i=1}^n (y_i - ax_i - b)^2.$$

Facciamo un po' di manipolazione algebrica per studiare questa quantità. Aggiungendo e togliendo $a\bar{x}$ e \bar{y} all'interno della parentesi e poi sviluppando il quadrato del trinomio si arriva a

$$\begin{aligned} S(a, b) &= \frac{1}{n} \sum \left((y_i - \bar{y}) + a(-x_i + \bar{x}) + (\bar{y} - a\bar{x} - b) \right)^2 \\ &= \frac{1}{n} \sum (y_i - \bar{y})^2 + \frac{1}{n} a^2 \sum (-x_i + \bar{x})^2 + (\bar{y} - a\bar{x} - b)^2 \\ &\quad + \frac{1}{n} 2a \sum (y_i - \bar{y})(-x_i + \bar{x}) + \frac{1}{n} 2(\bar{y} - a\bar{x} - b) \sum (y_i - \bar{y}) + \frac{1}{n} 2(\bar{y} - a\bar{x} - b) \sum (-x_i + \bar{x}) \end{aligned}$$

Osservando che $\sum(x_i - \bar{x}) = 0$ per ogni carattere concludiamo che le ultime due sommatorie si annullano. Otteniamo quindi

$$S(a, b) = \sigma_y^2 + a^2 \sigma_x^2 - 2a\sigma_{x,y} + (\bar{y} - b - a\bar{x})^2$$

dove abbiamo posto

$$\sigma_{x,y} = \frac{1}{n} \sum (x_i - \bar{x})(y_i - \bar{y}).$$

Quest'ultima quantità si dice *covarianza campionaria* dei caratteri x ed y . Osserviamo ora che il parametro b non compare nei primi tre termini dell'espressione ottenuta per $S(a, b)$ e che il quarto termine è un quadrato. Abbiamo quindi che per ogni valore di a fissato il minimo lo otteniamo ponendo il quarto termine uguale a 0. Deve quindi essere $b = -a\bar{x} + \bar{y}$. E osserviamo che questo fatto implica che la retta ai minimi quadrati passa per il punto (\bar{x}, \bar{y}) . Imponendo questa condizione rimane una funzione di secondo grado in a che assume il suo valore minimo per $a = \frac{\sigma_{x,y}}{\sigma_x^2}$. Abbiamo quindi che il segno della covarianza ci fornisce il segno del coefficiente angolare della retta ai minimi quadrati.

Nell'esempio in corso calcoliamo $\bar{x} = \frac{255}{7} = 36,49$, $\bar{y} = \frac{152}{7} = 21,71$, $\bar{x}^2 = \frac{10925}{7} = 1560,71$, $\bar{y}^2 = \frac{3574}{7} = 510,571$. Di conseguenza calcoliamo le varianze campionarie (che valori ci aspettiamo?)

$$\begin{aligned}\sigma_x^2 &= \bar{x}^2 - \bar{x}^2 = \frac{10925}{7} - \frac{65025}{49} = \frac{11450}{49} = 233,67 \\ \sigma_y^2 &= \bar{y}^2 - \bar{y}^2 = \frac{3574}{7} - \frac{23104}{49} = \frac{1914}{49} = 39,06.\end{aligned}$$

Per calcolare la covarianza campionaria dimostriamo prima una formula più semplice.

LEMMA 4.15. *Si ha*

$$\sigma_{x,y} = \overline{xy} - \bar{x}\bar{y}.$$

DIMOSTRAZIONE. Andiamo a calcolare

$$\begin{aligned}\sigma_{x,y} &= \frac{1}{n} \sum (x_i - \bar{x})(y_i - \bar{y}) \\ &= \frac{1}{n} \sum (x_i y_i - \bar{y} x_i - \bar{x} y_i + \bar{x} \bar{y}) \\ &= \overline{xy} - \bar{y}\bar{x} - \bar{x}\bar{y} + \bar{x}\bar{y} \\ &= \overline{xy} - \bar{x}\bar{y}\end{aligned}$$

e quindi la covarianza campionaria è effettivamente la media campionaria del prodotto meno il prodotto delle medie campionarie. □

La media campionaria del prodotto nel nostro esempio è $\overline{xy} = \frac{6160}{7} = 880$ e quindi $\sigma_{x,y} = 880 - \frac{255}{7} \frac{152}{7} = \frac{4360}{49} = 88,98$ Abbiamo quindi

$$\begin{aligned}a &= \frac{\sigma_{x,y}}{\sigma_x^2} = \frac{\frac{4360}{49}}{\frac{11450}{49}} = \frac{436}{1145} = 0,38 \\ b &= -a\bar{x} + \bar{y} = -\frac{436}{1145} \frac{255}{7} + \frac{152}{7} = -\frac{436}{229} \frac{51}{7} + \frac{152}{7} = \frac{12623}{1603} = 7,87\end{aligned}$$

Facciamo qualche osservazione qualitativa sulla covarianza. Se la covarianza è positiva allora i suoi addendi $(x_i - \bar{x})(y_i - \bar{y})$ saranno più facilmente positivi e quindi ci aspettiamo che se x_i è più

grande della media \bar{x} anche y_i sarà più grande della media \bar{y} . Similmente ci aspettiamo che se x_i è più piccolo della media \bar{x} anche y_i sarà più piccolo della media \bar{y} . Diciamo in questo caso che i caratteri sono positivamente o direttamente correlati.

Se invece la covarianza è negativa allora i suoi addendi $(x_i - \bar{x})(y_i - \bar{y})$ saranno più facilmente negativi e quindi ci aspettiamo che se x_i è più grande della media \bar{x} allora y_i sarà più piccolo della media \bar{y} . Similmente ci aspettiamo che se x_i è più piccolo della media \bar{x} anche y_i sarà più grande della media \bar{y} . Diciamo in questo caso che i caratteri sono negativamente o inversamente correlati.

Se la covarianza è nulla i due caratteri si dicono incorrelati. Ciò non vuol dire che non esista alcuna dipendenza tra le due. Ad esempio se i valori si dispongono lungo una parabola la covarianza campionaria sarà nulla.

La covarianza è sensibile ai cambiamenti di scala e per questo si tende ad utilizzare altri tipi di coefficienti per misurare la correlazione tra 2 variabili. Il coefficiente di correlazione è dato dalla formula

$$\rho_{x,y} = \frac{\sigma_{x,y}}{\sigma_x \sigma_y}$$

e si verifica facilmente che tale coefficiente è invariante per cambi di variabile di primo grado.

Diamo ora un'interpretazione intuitiva del coefficiente di correlazione. Riprendiamo pertanto la media dei quadrati degli errori $Err = S(a, b)$ con i valori ottenuti $a = \frac{\sigma_{x,y}}{\sigma_x^2}$ e $b = \bar{y} - a\bar{x}$. Otteniamo

$$\begin{aligned} Err &= \sigma_y^2 + \frac{\sigma_{x,y}^2}{\sigma_x^2} - 2 \frac{\sigma_{x,y}^2}{\sigma_x^2} \\ &= \sigma_y^2 - \frac{\sigma_{x,y}^2}{\sigma_x^2} \\ &= \sigma_y^2 (1 - \rho_{x,y}^2). \end{aligned}$$

Da questa formula deduciamo che il coefficiente di correlazione è un numero (puro, cioè senza unità di misura) sempre compreso tra -1 e 1 (perché Err è necessariamente non negativo) e che se il coefficiente di correlazione vale ± 1 allora $Err = 0$ e quindi i dati si dispongono lungo una retta.

Maggiore è il valore assoluto dell'indice di correlazione, migliore è l'approssimazione dei dati che si ottiene tramite la retta ai minimi quadrati.

6. Altri indici di posizione

Andiamo ora ad introdurre altri indici di posizione che si possono incontrare in statistica e che è utile conoscere.

DEFINIZIONE. I *quantili* di ordine n suddividono i valori assunti dal carattere in n gruppi equinumerosi. In particolare, l' i -esimo quantile di ordine n è tale che gli elementi del campione su cui il carattere assume un valore minore di esso sono (circa) gli $\frac{i}{n}$ del totale.

Consideriamo un carattere che assume i seguenti valori: 2,5,3,3,7,8,4,2,2,1,1,1,2,1,3. Una volta riordinati diventano 1,1,1,1,2,2,2,3,3,3,4,5,7,8.

- L'unico quantile di ordine 2 si dice *mediana*: questo suddivide i dati in due parti equinumerose. Nell'esempio precedente vale 2. Solitamente si definisce la mediana come il valore centrale se il campione ha cardinalità dispari e come la media dei due valori centrali se il campione ha cardinalità pari.

- I quantili di ordine 4 si dicono *quartili*: suddividono i dati in quattro parti equinumerose. Chiaramente il secondo quartile è proprio la mediana. Nell'esempio precedente si può scegliere come primo quartile 1.5 e come terzo quartile 3.5. La differenza tra il terzo ed il primo quartile si dice *scarto interquartile*: è un indice di dispersione solitamente più utilizzato e più significativo del campo di variazione.
- I quantili di ordine 10 si dicono *decili* e quelli di ordine 100 si dicono *percentili*. Ad esempio il 90mo percentile è tale che gli elementi del campione che assumono valore minore di esso costituiscono il 90% del totale.

È talvolta opportuno trasformare i dati in nostro possesso per calcolare una media più significativa per il problema che ci interessa. In particolare, se A e B sono due intervalli (eventualmente illimitati), i nostri dati si trovano in A e $f: A \rightarrow B$ è una trasformazione invertibile allora la media associata ad f è

$$\bar{x}_f = f^{-1}\left(\frac{1}{n}(f(x_1) + \cdots + f(x_n))\right),$$

cioè trasformiamo i dati tramite la f , calcoliamo la media e poi applichiamo l'inversa di f . Gli esempi più importanti sono il caso $f(x) = \log x$ (media geometrica), $f(x) = x^2$ (media quadratica), e $f(x) = 1/x$ (media armonica).

- La media geometrica. Supponiamo di avere delle azioni in borsa per 4 anni che hanno registrato queste variazioni annue: +10%, -20%, +24%, -14%. Qual è stata la variazione media?

Per rispondere a questa domanda abbiamo bisogno di convertire le variazioni percentuali in coefficienti moltiplicativi. Ad esempio +10% corrisponde al fattore 1.10, mentre -14% corrisponde al fattore 0.86. Abbiamo quindi in conclusione una variazione complessiva data da un fattore $1.10 \cdot 0.80 \cdot 1.24 \cdot 0.86 = 0.9384 = (0.9842)^4$. La variazione complessiva è del -6,16% (corrispondente al fattore 0.9384) e quindi una variazione media annua del -1,58% (corrispondente al fattore 0.9384). La media da utilizzare in questo caso è la geometrica

$$(x_1 \cdots x_n)^{1/n}$$

tra le percentuali espresse come fattori moltiplicativi, in questo caso cioè tra i numeri 1.10, 0.80, 1.24, 0.86.

- La media quadratica. Supponiamo di avere 3 tubi di diametro 1, 2, 3 centimetri e li vogliamo sostituire con 3 tubi uguali in modo che la portata complessiva rimanga invariata. In questo caso vogliamo che la sezione complessiva dei 3 tubi rimanga la stessa. Se d è il diametro che stiamo cercando vogliamo cioè

$$3\pi(d/2)^2 = \pi(1/2)^2 + \pi(2/2)^2 + \pi(3/2)^2$$

da cui

$$d = \sqrt{\frac{1^2 + 2^2 + 3^2}{3}} = 2.16.$$

Abbiamo fatto in questo caso la media quadratica

$$\sqrt{\frac{1}{n}(x_1^2 + \cdots + x_n^2)}.$$

- Facciamo una passeggiata in montagna. All'andata si va a 3 km/h. Al ritorno a 7 km/h. Qual è la velocità media? Detto s lo spazio percorso in km sia all'andata che al ritorno abbiamo che all'andata abbiamo impiegato $t_1 = s/3$ ore mentre al ritorno $t_2 = s/7$ ore. La velocità media complessiva è quindi

$$\frac{2s}{s/3 + s/7} = \frac{1}{\frac{1/3 + 1/7}{2}} = 4.2 km/h.$$

Abbiamo fatto in questo caso la media armonica

$$\left(\frac{x_1^{-1} + \dots + x_n^{-1}}{n} \right)^{-1}.$$

CHAPTER 5

Spazi di probabilità

In questo capitolo introduciamo le nozioni di base della teoria della probabilità.

1. Fenomeni deterministici e casuali: risultati ed eventi

Un fenomeno è un qualunque accadimento che porta ad un certo *risultato*, o *esito*, appartenente ad un determinato insieme. Il fenomeno si dice *deterministico* se il risultato può essere previsto o calcolato (determinato) con esattezza prima dell'effettuazione del fenomeno. Ad esempio se lasciamo un grave ad una quota di 1 metro sappiamo calcolare con precisione l'istante in cui giungerà a terra. O se lanciamo una palla con una data angolazione ad una certa velocità sappiamo prevedere il punto in cui toccherà il terreno.

Un fenomeno si dice *casuale* (o *aleatorio*) se non è possibile prevedere il risultato. Ad esempio, quale faccia uscirà nel lancio di un dado? O quanto finisce una certa partita di calcio? O in che anno mi si romperà la lavatrice? O quanto sarà alto un bimbo nato oggi tra un anno?

L'insieme di tutti i possibili risultati di un fenomeno si indica solitamente con Ω . Nei casi precedenti abbiamo $\Omega = \{1, 2, \dots, 6\}$, $\{(0, 0), (0, 1), (1, 0), \dots\}$, $\{2017, 2018, \dots\}$, $[50, 200]$.

DEFINIZIONE. Un *evento* è un sottoinsieme di Ω . In altre parole, un evento è un insieme di possibili risultati di un fenomeno aleatorio.

Negli esempi precedenti possiamo avere $E = \{2, 4, 6\}$, $\{(x, y) : x > y\}$, $\{x : x > 2025\}$, $\{x : x < 60\}$. In realtà vedremo più avanti che nella teoria più generale cui ci interesseremo solo *alcuni* sottoinsiemi di Ω potranno essere chiamati eventi.

Sugli eventi, essendo insiemi, possiamo considerare le comuni operazioni di intersezione, unione e complementare. Ad esempio, nel lancio di 2 dadi l'evento “esce almeno un 1” (l'evento cioè costituito da tutti i risultati (i, j) in cui $i = 1$ oppure $j = 1$) si vede facilmente essere unione dei due eventi $E_1 =$ “il primo dado dà 1” ed $E_2 =$ “il secondo dado dà 1”. L'evento “escono due 1” ne è chiaramente l'intersezione. L'evento “non esce neanche un 1” è invece il complementare di quest'ultimo.

DEFINIZIONE. Una *valutazione di probabilità* è una funzione che associa ad ogni evento un numero tanto più grande quanto più riteniamo che tale evento possa accadere. Denotiamo con $P(E)$ tale valutazione dell'evento E .

Vorremo certe proprietà. Ad esempio, se A e B sono eventi disgiunti ci piacerebbe $P(A \cup B) = P(A) + P(B)$. Nel caso della partita di calcio la probabilità che la squadra di casa non perda dovrà essere la somma della probabilità che vinca e di quella che pareggi. Nella prossima sezione introduciamo gli spazi di probabilità in modo formale andando ad iniziare il loro studio sistematico.

2. Spazi di probabilità

Sia dato un insieme Ω (di possibili risultati di un fenomeno aleatorio). Gli eventi saranno per noi i sottoinsiemi di Ω di cui vorremo e potremo calcolare le probabilità.

Pertanto, NON HA SENSO calcolare la probabilità di un sottoinsieme di Ω che non sia un evento. E' chiaro quindi che ci piacerebbe che unioni, intersezioni e complementari di eventi siano ancora eventi. Formalizziamo questo desiderio dopo aver introdotto la seguente notazione.

Se E_1, E_2, \dots (finiti o infiniti) sono eventi diciamo che formano una *successione* di eventi. In generale se scriviamo un ultimo elemento, cioè scriviamo E_1, E_2, \dots, E_n stiamo imponendo che la successione sia finita. Se invece non scriviamo l'ultimo elemento, cioè scriviamo E_1, E_2, \dots richiediamo che la successione sia infinita. In certi casi una successione finita può essere vista come una successione infinita, ad esempio aggiungendo infinite volte l'insieme vuoto.

In modo del tutto analogo considereremo successioni di numeri, di funzioni o di altri oggetti.

DEFINIZIONE. Sia \mathcal{A} una collezione di sottoinsiemi di Ω . Diciamo che \mathcal{A} è una *famiglia coerente di eventi* su Ω se:

- (1) $\emptyset, \Omega \in \mathcal{A}$;
- (2) $E^C \in \mathcal{A}$ se $E \in \mathcal{A}$;
- (3) Se E_1, E_2, \dots sono in \mathcal{A} allora anche la loro unione e la loro intersezione stanno in \mathcal{A} .

Se $E \in \mathcal{A}$ diciamo che E è un evento.

ESEMPIO 5.1. dato $\Omega = \{1, 2, 3, 4, 5, 6\}$, si considerino le seguenti collezioni di sottoinsiemi.

$$\mathcal{A}_1 = \{\emptyset, \{1, 2, 3, 4, 5, 6\}, \{1, 3\}, \{2, 4, 5, 6\}\}$$

e

$$\mathcal{A}_2 = \{\emptyset, \{1, 2, 3, 4, 5, 6\}, \{1, 3\}, \{2, 4, 5, 6\}, \{1, 3, 5\}, \{2, 4, 6\}\}.$$

\mathcal{A}_1 è una famiglia coerente di eventi mentre \mathcal{A}_2 non lo è (non è chiusa rispetto all'intersezione).

Una famiglia coerente di eventi \mathcal{A} viene solitamente chiamata una σ -algebra. Avremmo potuto scegliere anche solo unioni e intersezioni finite anzichè numerabili nella definizione qui sopra, ma si preferisce utilizzare questo tipo di definizione.

Nel caso in cui Ω è finito oppure numerabile (ad esempio $\Omega = \mathbb{Z}$) sarà spesso naturale prendere come \mathcal{A} la famiglia di *tutti* i sottoinsiemi di Ω , che chiaramente è una famiglia coerente di eventi.

DEFINIZIONE. Data una famiglia coerente di eventi \mathcal{A} una (valutazione di) probabilità su \mathcal{A} è una funzione

$$P : \mathcal{A} \rightarrow \mathbb{R}_{\geq 0}$$

che soddisfa le seguenti proprietà

- (1) $P(\Omega) = 1$;
- (2) data una successione di eventi a due a due disgiunti A_1, A_2, \dots si ha

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i).$$

La proprietà (2) qui sopra vale chiaramente anche per unioni finite, ma non viene richiesta nel caso di unioni infinite non numerabili. Potrà benissimo accadere che, se $\Omega = \mathbb{R}$, allora $P(\{x\}) = 0$ per ogni $x \in \mathbb{R}$. Ed anzi vedremo che capiterà piuttosto spesso...

DEFINIZIONE. Uno spazio di probabilità è una terna (Ω, \mathcal{A}, P) , dove Ω è un insieme, \mathcal{A} è una famiglia coerente di eventi su Ω e P una probabilità su \mathcal{A} .

Lo studio di un fenomeno aleatorio si suddivide in 2 parti. La prima, lasciata all'interesse, alla sensibilità ed all'esperienza dello studioso, consiste nel determinare uno spazio di probabilità che permette di studiare il fenomeno. Questa fase, detta modellizzazione, non ha un'unica soluzione, ed anzi non c'è neanche un giudizio universale che dica se un modello è migliore di un altro. Anche se poi nella maggior parte dei casi si converrà che un modello meglio si adatta di un altro nello studio di un fenomeno aleatorio.

La seconda fase, invece, consiste nello studio del modello, cioè dello spazio di probabilità costituito nella prima fase: qui non c'è più ambiguità e discrezionalità e i risultati sono univocamente determinati. La seconda fase, in altre parole, è matematica. Facciamo un primo esempio.

ESEMPIO 5.2. Abbiamo tre urne. La prima contiene un certo numero imprecisato di palline numerate 1 o 2, mentre sia la seconda che la terza contengono - anche qui un numero imprecisato - di palline numerate 3 o 4. Scegliamo un'urna a caso ed estraiamo una pallina. Vogliamo modellizzare questo fenomeno aleatorio con uno spazio di probabilità.

I risultati possibili sono 1,2,3,4 per cui abbiamo $\Omega = \{1, 2, 3, 4\}$. Il risultato $\{1\}$ tuttavia non potrà essere considerato come un evento, perchè non abbiamo informazioni per determinarne la probabilità. La cosa più naturale da fare è scegliere

$$\mathcal{A} = \{\Omega, \emptyset, \{1, 2\}, \{3, 4\}\},$$

ponendo poi

$$P(\{1, 2\}) = 1/3, \quad P(\{3, 4\}) = 2/3.$$

Questa è la modellizzazione di questo semplice fenomeno aleatorio.

3. La probabilità uniforme

Prendiamo il caso in cui Ω è un insieme finito e supponiamo che per ragioni di simmetria tutti i risultati siano equiprobabili. In questo caso si sceglie \mathcal{A} come l'insieme di tutti i possibili sottoinsiemi di Ω e parliamo di *spazio di probabilità uniforme*.

PROPOSIZIONE 5.3. Se (Ω, \mathcal{A}, P) è uno spazio di probabilità uniforme ed $A \subseteq \Omega$ è un evento allora

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

DIMOSTRAZIONE. Infatti, sia $\Omega = \{r_1, \dots, r_n\}$ e supponiamo che $P(\{r_1\}) = \dots = P(\{r_n\}) = p$; allora, per la proprietà di additività della probabilità su eventi disgiunti, abbiamo

$$1 = P(\Omega) = P\left(\bigcup_{i=1}^n \{r_i\}\right) = \sum_{i=1}^n P(\{r_i\}) = np$$

da cui segue che $p = 1/n$. Se l'evento A contiene k elementi, $A = \{r_{i_1}, \dots, r_{i_k}\}$ allora similmente abbiamo

$$P(A) = P\left(\bigcup_{j=1}^k \{r_{i_j}\}\right) = \sum_{i=1}^k P(\{r_{i_j}\}) = k \cdot \frac{1}{n} = \frac{|A|}{|\Omega|}.$$

□

Abbiamo cioè che, se tutti i risultati possibili sono equiprobabili, allora la probabilità di un evento A è data dal rapporto tra il numero di risultati "favorevoli" (cioè $|A|$) e il numero di risultati possibili (cioè $|\Omega|$).

Il calcolo della probabilità di un evento si riduce quindi ad un problema di combinatoria, più precisamente al calcolo delle cardinalità di certi insiemi finiti.

- ESEMPIO 5.4. (1) Qual è la probabilità di lanciare un dado ed ottenere un numero pari?
 (2) E qual è la probabilità di indovinare un terno al lotto?
 (3) Qual è la probabilità che esca un numero fissato al lotto?

Rispondiamo con ordine alle re domande

- (1) In questo caso abbiamo $\Omega = \{1, 2, 3, 4, 5, 6\}$ con probabilità uniforme. Se $A = \{2, 4, 6\}$ è l'evento costituito dai risultati pari abbiamo facilmente

$$P(A) = \frac{|A|}{|\Omega|} = \frac{3}{6} = 50\%.$$

Niente di sorprendente, vero?

- (2) Qui abbiamo bisogno di fare un po' di attenzione in più. Chi è Ω in questo caso? Ricordiamo che nel lotto vengono estratti 5 numeri tra 1 e 90, fare terno vuol dire che fra i 5 numeri estratti ci sono i 3 su cui abbiamo scommesso. L'errore tipico che si fa a questo punto è quello di scegliere $\Omega = \{1, 2, \dots, 90\}$. Ma in questo caso il fenomeno aleatorio consiste nell'estrazione di 5 numeri, non di 1 solo! I numeri vengono estratti con un certo ordine, ma ai fini del nostro problema l'ordine non è importante per cui possiamo prendere come Ω l'insieme di tutti i sottoinsiemi di $\{1, 2, \dots, 90\}$ costituiti da esattamente 5 elementi:

$$\Omega = \{\{1, 2, 3, 4, 5\}, \{1, 2, 3, 4, 6\}, \dots, \{86, 87, 88, 89, 90\}\}$$

per cui

$$|\Omega| = \binom{90}{5} = \frac{90 \cdot 89 \cdot 88 \cdot 87 \cdot 86}{5!} = 43.949.268$$

Quali sono i risultati favorevoli? Sono tutti i sottoinsiemi di 5 elementi che contengono i 3 numeri su cui ho scommesso. Supponendo di aver scommesso sui numeri 1,2,3 per comodità abbiamo

$$A = \{\{1, 2, 3, 4, 5\}, \{1, 2, 3, 4, 6\}, \dots, \{1, 2, 3, 89, 90\}\}$$

Per ottenere gli elementi di A abbiamo quindi bisogno di aggiungere ad 1,2,3 altri 2 numeri scelti tra gli altri 87; abbiamo quindi

$$|A| = \binom{87}{2} = \frac{87 \cdot 86}{2} = 3.741$$

Concludiamo che

$$P(A) = \frac{|A|}{|\Omega|} = \frac{1}{11.748}.$$

- (3) A questo punto questo dovrebbe sembrare più semplice. Lo spazio Ω è quello del punto precedente. L'evento E di cui vogliamo calcolare la probabilità è quello che contiene il numero su cui abbiamo puntato e abbiamo quindi, analogamente a prima,

$$|E| = \binom{89}{4} = 2.441.626$$

e quindi

$$P(E) = \frac{|E|}{|\Omega|} = \frac{1}{18}.$$

C'è un altro modo più semplice di calcolare questa probabilità, tenendo conto dell'ordine con cui i numeri vengono estratti. Chiamiamo A_1 l'evento dato da "il mio numero è uscito come primo estratto", A_2 l'evento dato da "il mio numero è uscito come secondo estratto" e analogamente definiamo A_3, A_4, A_5 . Per ragioni di simmetria abbiamo

$$P(A_1) = \frac{1}{90}$$

perché il primo estratto è uno dei possibili 90 valori. Chiaramente gli eventi A_1, A_2, \dots, A_5 sono disgiunti perché il mio numero non può essere estratto più di una volta e quindi per l'additività della probabilità su eventi disgiunti abbiamo

$$P(E) = P(A_1) + \dots + P(A_5) = \frac{1}{90} + \dots + \frac{1}{90} = \frac{1}{18}.$$

ESEMPIO 5.5. Soffermiamoci sulla soluzione dell'ultimo punto dell'esempio precedente. Supponiamo di essere interessati solo al secondo estratto al lotto. In questo caso i risultati possibili sono i numeri da 1 a 90 e quindi $\Omega = \{1, 2, \dots, 90\}$ e per ragioni di simmetria non c'è motivo perché un numero sia più probabile di un altro. Siamo quindi in presenza di probabilità uniforme e quindi tale probabilità è $1/90$.

ESEMPIO 5.6. Vediamo ora il caso in cui i possibili risultati sono infiniti, ad esempio $\Omega = \mathbb{N}$. Consideriamo il fenomeno "scelta di un numero a caso", senza soffermarci troppo su come questo numero venga selezionato. Qual è la probabilità di estrarre il 396?

La domanda posta così non ha senso perché non abbiamo specificato lo spazio di probabilità. La domanda presuppone che l'insieme $\{396\}$ sia un evento. Per uniformità tutti i numeri sono eventi. Ancora per uniformità hanno tutti la stessa probabilità. Se questa probabilità fosse 0 per la proprietà di additività avremmo che $P(\Omega) = 0$. Se questa probabilità fosse strettamente maggiore di zero avremmo $P(\Omega) = +\infty$. In ogni caso contraddiciamo la definizione di probabilità. Ne deduciamo che la domanda non può avere senso in ogni caso!

Per modellizzare questo fenomeno abbiamo quindi bisogno di scegliere solo determinati eventi e decidere una funzione probabilità che sia per noi coerente con l'uniformità. Ad esempio potremmo decidere di considerare gli eventi costituiti dalle progressioni aritmetiche insieme a tutti gli altri che si ottengono da essi facendo unioni, intersezioni e passaggi al complementare. Ad esempio i multipli di 2 formerebbero un esempio o anche i multipli di 3. E saremmo tentati di dire che la probabilità di scegliere un multiplo di 2 è $1/2$ e la probabilità di scegliere un multiplo di 3 è $1/3$. Questo è

legittimo, perchè nella modellizzazione uno fa come vuole. Ma se i numeri interi li scrivessimo in questo ordine 0,1,3,2,6,4,9,5,12,7,15,8,18,10,21,11,24,13,27,14....? Se li guardiamo a 4 a 4 abbiamo sempre 2 numeri pari e 2 multipli di 3... e non vorremmo che lo spazio di probabilità del fenomeno “scelta di un numero a caso” dipendesse dall’ordine con cui li scriviamo....

Altri problemi nella scelta di una σ -algebra si registrano se Ω è continuo, ad esempio nel fenomeno dato dal tempo di rottura di un certo apparecchio. Abbiamo bisogno di tutti gli intervalli, che comunque non bastano, e non possiamo prendere tutti i sottoinsiemi come eventi perché otterremmo una σ -algebra troppo complessa da poter essere esaminata. Ma ci torneremo più in là.

Cominciamo a fare qualche osservazione utile al calcolo della probabilità di un evento. Dati due eventi qualunque A e B si ha $B = (B \cap A) \cup (B \cap A^C)$ dove l’unione è disgiunta. Ne segue

$$(2) \quad P(B) = P(B \cap A) + P(B \cap A^C)$$

Queta formula dice semplicemente che il calcolo di $P(B)$ può essere diviso in due sottoeventi: quello formato dai risultati che stanno anche in A e quello formato dai risultati che non stanno in A .

ESEMPIO 5.7. Abbiamo una scatola con due palline rosse e una nera e una scatola con due palline nere e due rosse. Estruendo una pallina a caso qual è la probabilità di estrarre una pallina rossa? Ancora non sappiamo calcolare questa probabilità ma sappiamo semplificarla in 2 problemi più semplici. In questo caso lo spazio dei risultati Ω è formato dalle sette palline. Consideriamo l’evento R costituito dalle palline rosse, l’evento A_1 costituito dalle palline della prima scatola e A_2 l’evento costituito dalle palline della seconda scatola. Abbiamo

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

Possiamo cioè effermare che la probabilità di estrarre una pallina rossa è la somma tra la probabilità di estrarre una rossa dalla prima scatola e la probabilità di estrarre una rossa dalla seconda scatola.

Questo è un esempio della seguente costruzione che vale molto più in generale. Sia $\{A_1, A_2, \dots\}$ una partizione di Ω (cioè, lo ricordiamo, $A_i \cap A_j = \emptyset$ per ogni $i \neq j$ e $\cup A_i = \Omega$). Se gli A_i sono anche eventi diremo che la successione A_1, A_2, \dots è una partizione in eventi di Ω .

Sia quindi A_1, A_2, \dots una partizione in eventi di Ω ; abbiamo $B = \cup_{i=1}^{\infty} (B \cap A_i)$ e quindi abbiamo

$$(3) \quad P(B) = \sum_{i=1}^{\infty} P(B \cap A_i)$$

Dalla proprietà precedente (2) deduciamo che

- $P(A^C) = 1 - P(A)$ ponendo $B = \Omega$;
- $P(A) \leq P(B)$ se $A \subseteq B$.

In particolare ricordiamo anche che il complementare di un’unione è uguale all’intersezione dei complementari e quindi

$$P(\cup A_i) = 1 - P(\cap A_i^C).$$

ESEMPIO 5.8. Calcolare la probabilità di ottenere almeno un 6 lanciando 2 dadi. Poniamo A_i l’evento “esce 6 al lancio i ”. La probabilità da calcolare è quindi $P(A_1 \cup A_2)$. Tramite la formula precedente possiamo facilmente calcolare la probabilità dell’evento complementare, osservando che $|A_1^C \cap A_2^C| = 25$ e quindi $P(A_1 \cup A_2) = 1 - \frac{25}{36} = \frac{11}{36}$.

Abbiamo qui visto un modo per calcolare la probabilità di un'unione. Vediamone un altro. Osserviamo che si ha sempre

$$A \cup B = A \cup (B \setminus A),$$

dove la seconda unione è disgiunta. Ne segue, utilizzando (2)

$$P(A \cup B) = P(A) + P(B \setminus A) = P(A) + P(B \cap A^C) = P(A) + P(B) - P(A \cap B),$$

che ricorda il principio di inclusione-esclusione visto nel calcolo combinatorio. Sostanzialmente dobbiamo sottrarre $P(A \cap B)$ perché questa quantità è stata calcolata 2 volte (una in $P(A)$ e una in $P(B)$.) Rivediamo l'esempio precedente con questa formula. Abbiamo

$$P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2) = \frac{1}{6} + \frac{1}{6} - \frac{1}{36} = \frac{11}{36}.$$

Questa formula si generalizza al caso dell'unione di 3 eventi e più in generale al caso di k eventi in modo del tutto analogo al principio di inclusione-esclusione.

4. La probabilità condizionale.

La probabilità condizionale è la probabilità che si ottiene se per qualche motivo sappiamo già, o vogliamo imporre, che il risultato del fenomeno aleatorio che stiamo studiando stia in un certo evento A . Un esempio potrebbe essere il seguente: abbiamo studiato prima la probabilità di ottenere un terno al lotto. Potremmo chiederci: qual è la probabilità di ottenere un terno sapendo che il primo estratto è uno dei 3 numeri che ho giocato? Questo è un tipico esempio di probabilità condizionale: se indichiamo con E l'evento "fare terno" e con A l'evento "il primo estratto è uno dei tre numeri giocati", la probabilità condizionale verrà indicata con $P(E|A)$.

Affrontiamo questo problema da un punto di vista più formale. Dato uno spazio (Ω, \mathcal{A}, P) e fissato un evento A tale che $P(A) > 0$ (che sarà l'evento che fornisce la condizione) formiamo un nuovo spazio di probabilità.

PROPOSIZIONE 5.9. *La terna $(\Omega, \mathcal{A}, P')$ con*

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

è uno spazio di probabilità.

DIMOSTRAZIONE. Verifichiamo che P' è effettivamente una probabilità.

- $P'(\Omega) = 1$ ovvio;
- Siano B_i disgiunti. Allora

$$P'(B_1 \cup B_2 \cup \dots) = \frac{P(A \cap (B_1 \cup B_2 \cup \dots))}{P(A)} = \frac{P((A \cap B_1) \cup (A \cap B_2) \cup \dots)}{P(A)} = P'(B_1) + P'(B_2) + \dots.$$

□

Come anticipato, la probabilità $P'(B)$ si dice probabilità condizionale di B rispetto ad A , o più semplicemente probabilità di B dato A , e si indicherà $P(B|A)$.

Cerchiamo ora di capire perché questa definizione fornisce effettivamente una risposta al quesito che ci eravamo posti: determinare la probabilità che si verifichi un evento B sapendo che se ne verifica un altro A . Se rappresentiamo gli eventi con dei diagrammi di Venn possiamo fare in

modo che la superficie corrispondente ad un certo evento sia proporzionale alla sua probabilità. In questo modo molte delle formule viste prima e che vedremo in seguito hanno un'immediata interpretazione grafico-geometrica. Detta $S(A)$ la superficie dell'evento A , siccome $P(\Omega) = 1$ abbiamo che $P(A) = S(A)/S(\Omega)$; nel caso della probabilità condizionale la quantità

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

rappresenta quindi la proporzione della superficie di A che sta anche in B . Stiamo cioè considerando il nuovo spazio di probabilità dato osservando ciò che capita esclusivamente all'interno di A . In altre parole $P(B|A)$ rappresenta la probabilità che accada l'evento B sapendo che il risultato appartiene all'evento A .



Possiamo ora completare l'esempio 5.7. Proseguendo con la stessa notazione abbiamo

$$P(R) = P(R \cap A_1) + P(R \cap A_2) = P(R|A_1)P(A_1) + P(R|A_2)P(A_2) = \frac{2}{3} \cdot \frac{1}{2} + \frac{2}{4} \cdot \frac{1}{2} = \frac{7}{12}.$$

La probabilità condizionale è molto importante anche nella fase di modellizzazione, in quanto talvolta si conoscono delle probabilità condizionali, ma non quelle assolute.

Riprendiamo anche l'esempio del lotto. Sia quindi B l'evento "esce il terno $\{1, 2, 3\}$ " e A l'evento "il primo estratto è 1, 2 o 3". Determinare $P(B|A)$. Abbiamo già calcolato nell'Esempio 5.4 $P(B) = \frac{1}{11.748} = 0,000085121$, mentre chiaramente $P(A) = \frac{1}{30}$. Abbiamo bisogno di calcolare $P(A \cap B)$: consideriamo ora i risultati delle estrazioni "ordinati"; i risultati favorevoli sono

$$\binom{4}{2} \cdot 3! \cdot 87 \cdot 86 = 269352$$

mentre i risultati possibili sono $(90)_5$ per cui

$$P(A \cap B) = \frac{269352}{(90)_5} = \frac{1}{19624}$$

e quindi

$$P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{30}{19624} = 0,0015$$

che, come era lecito aspettarsi, è sensibilmente più grande di $P(B)$.

5. La formula delle probabilità totali e la formula di Bayes

La formula delle probabilità totali viene utilizzata quando ci viene naturale (o ci rendiamo conto che è conveniente) suddividere il calcolo della probabilità di un evento in "casi", corrispondenti ad una partizione di Ω . Abbiamo infatti:

se A_1, \dots, A_n è una partizione in eventi di Ω si ha per ogni evento B la seguente formula delle probabilità totali:

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

Questa formula, che abbiamo già implicitamente utilizzato per risolvere il problema dell'esempio 5.7, si ottiene facilmente dalla eq. (3) insieme alla definizione di probabilità condizionale.

ESEMPIO 5.10. estraiamo n palline senza rimpiazzo da un'urna contenente b palline bianche e r palline rosse;

- qual è la probabilità che la prima pallina sia bianca?
- qual è la probabilità che la seconda pallina sia bianca?
- qual è la probabilità che l' n -esima pallina sia bianca?

In questo caso consideriamo Ω l'insieme di tutte le n -ple ordinate di palline distinte scelte tra le $b + r$ nell'urna (tutte le possibili estrazioni). Su Ω abbiamo la probabilità uniforme in quanto le n -ple sono tutte equiprobabili. Sia B_i l'evento "l' i -esima pallina estratta è bianca". Abbiamo facilmente $P(B_1) = \frac{b}{b+r}$. Per calcolare B_2 utilizziamo la formula delle probabilità totali rispetto alla partizione $\{B_1, B_1^C\}$. Abbiamo:

$$P(B_2) = P(B_1)P(B_2|B_1) + P(B_1^C)P(B_2|B_1^C) = \frac{b}{b+r} \frac{b-1}{b+r-1} + \frac{r}{b+r} \frac{b}{b+r-1} = \frac{b}{b+r},$$

ed è quindi uguale a $P(B_1)$. Osserviamo che in effetti si ha $P(B_i) = \frac{b}{b+r}$ per ogni i . Infatti, lo scambio della prima e della i -esima pallina estratta determina una biiezione tra i risultati in B_1 e i risultati in B_i : abbiamo quindi che le n -ple di palline in cui la prima è bianca sono tante quante le n -ple di palline in cui la i -esima è bianca. Ne segue che $P(B_i) = P(B_1) = \frac{b}{b+r}$ per ogni i .

ESERCIZIO 5.11. Il 40% degli abitanti di una certa regione è costituito da fumatori, il rimanente 60% da non fumatori. Si sa che il 25% dei fumatori e il 7% dei non fumatori soffrono di una certa malattia. Qual è la probabilità che una persona scelta a caso sia un fumatore? Poniamo Ω l'insieme di tutti gli abitanti, M l'insieme dei malati, F dei fumatori e N dei non fumatori. È chiaro che $\{F, N\}$ è una partizione di Ω per cui, usando la formula delle probabilità totali otteniamo

$$P(M) = P(M|F)P(F) + P(M|N)P(N) = 0,25 \cdot 0,40 + 0,07 \cdot 0,60 = 0,142 = 14.2\%.$$

Molto utile risulta anche la formula di Bayes: essa permette di calcolare una probabilità condizionale $P(B|A)$ quando ci sembra più conveniente calcolare la probabilità condizionale $P(A|B)$. Infatti, osserviamo che

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

da cui deduciamo la formula, detta di Bayes, $P(A|B) = \frac{P(B|A)P(A)}{P(B)}$.

ESERCIZIO 5.12. Qual è la probabilità che un malato scelto a caso nella regione dell'esercizio 5.11 sia un fumatore? Dobbiamo calcolare $P(F|M)$ e per farlo utilizziamo la formula di Bayes

$$P(F|M) = \frac{P(M|F)P(F)}{P(M)} = \frac{0,25 \cdot 0,40}{0,142} = 0,704 = 70.4\%$$

ESERCIZIO 5.13. Avendo 3 mobili con 2 cassetti, il primo con 2 monete d'oro, il secondo un oro e un argento il terzo 2 argenti. Aprendo a caso un cassetto si trova oro. Qual è la probabilità che nell'altro cassetto ci sia ancora oro? La risposta è $2/3$, ma prova ad arrivarci senza ulteriori indizi.

Altro esempio simile al precedente: un'urna contiene 2 carte. Una ha entrambi i lati neri, l'altro un bianco e un nero. Si prende una carta e si guarda una faccia: è nera. Qual è la probabilità che anche l'altra faccia sia nera? Sia A_i l'evento estrazione della carta i e N e B esposizione di una faccia bianca e nera. Dobbiamo calcolare

$$P(A_1|N) = \frac{P(N|A_1)P(A_1)}{P(N)} = \frac{1 \cdot \frac{1}{2}}{\frac{3}{4}} = \frac{2}{3}.$$

ESEMPIO 5.14. Abbiamo 3 urne. La i -esima ha 1 pallina rossa e i palline bianche. Sapendo di aver estratto una pallina bianca, qual è la probabilità di averla estratta dalla terza urna? Questo è un esempio tipico di applicazione combinata della formula delle probabilità totalie della formula di Bayes. Sia A_i l'evento dato dalle palline dell'urna i e B l'evento dato dalle palline bianche. Dobbiamo calcolare $P(A_3|B)$ e chiaramente sappiamo calcolare $P(B|A_3) = 3/4$ per utilizzare la formula di Bayes abbiamo quindi di $P(A_3) = 1/3$ e di $P(B)$ per calcolare quest'ultima usiamo la formula delle probabilità totali

$$P(B) = P(A_1)P(B|A_1) + P(A_2)P(B|A_2) + P(A_3)P(B|A_3) = \frac{1}{3} \cdot \frac{1}{2} + \frac{1}{3} \cdot \frac{2}{3} + \frac{1}{3} \cdot \frac{3}{4} = \frac{23}{36}$$

e quindi

$$P(A_3|B) = \frac{P(A_3)P(B|A_3)}{P(B)} = \frac{1}{3} \cdot \frac{3}{4} \cdot \frac{36}{23} = \frac{9}{23}.$$

ESEMPIO 5.15. Un'urna contiene 3 palline bianche e 5 nere. Ne estraiamo 2 con rimpiazzo. Qual è la probabilità di estrarre 2 palline dello stesso colore? E la probabilità di estrarre almeno una pallina bianca? Consideriamo gli eventi A_i = "bianca all'estrazione i ".

ESEMPIO 5.16. Un'urna contiene 4 palline bianche e 3 rosse. Se ne estraggono 2 con rimpiazzo. E' più probabile che siano dello stesso colore o diverse? Sia B_i l'evento "pallina bianca all'estrazione i " e similmente definiamo R_i . L'evento E = "estrazione di 2 palline uguali è quindi dato da

$$E = (B_1 \cap B_2) \cup (R_1 \cap R_2).$$

Siccome $B_1 \cap B_2$ e $R_1 \cap R_2$ sono chiaramente disgiunti abbiamo

$$P(E) = P(B_1 \cap B_2) + P(R_1 \cap R_2).$$

Sfruttiamo ora il fatto che B_1 e B_2 sono indipendenti (e similmente R_1 e R_2) e concludiamo che

$$P(E) = P(B_1)P(B_2) + P(R_1)P(R_2) = \frac{4}{7} \cdot \frac{4}{7} + \frac{3}{7} \cdot \frac{3}{7} = \frac{25}{49}.$$

E' quindi leggermente più probabile ottenere 2 palline uguali che diverse.

6. Eventi indipendenti

Due eventi A e B si dicono indipendenti il fatto che uno dei due accade non cambia la probabilità che anche l'altro accada, cioè $P(B) = P(B|A)$ e $P(A) = P(A|B)$; riprendendo la definizione di probabilità condizionale abbiamo che quest due equazioni sono equivalenti alla condizione $P(A \cap B) = P(A) \cdot P(B)$. Similmente definiamo n eventi indipendenti. Guardando alla probabilità condizionale questo ci dice che $P(B) = P(B|A)$ (se $P(A) \neq 0$) e $P(A) = P(A|B)$ (se $P(B) \neq 0$). Sapere o imporre che due eventi sono indipendenti è fondamentale nella fase di modellizzazione.

ESEMPIO 5.17. Lanciando un dado diciamo che otteniamo un successo se esce un numero ≥ 5 . Lanciando 8 volte il dado, qual è la probabilità di ottenere una sequenza 00101001, dove lo 0 indica un insuccesso e 1 un successo? Siccome le prove sono indipendenti tra loro abbiamo che tale probabilità è

$$\frac{2}{3} \frac{2}{3} \frac{1}{3} \frac{2}{3} \frac{1}{3} \frac{2}{3} \frac{2}{3} \frac{1}{3} = \left(\frac{1}{3}\right)^5 \left(\frac{2}{3}\right)^3.$$

Osserviamo che tale probabilità non dipende dall'ordine con cui si susseguono i successi e gli insuccessi, ma solo da quanti successi e da quanti insuccessi si devono verificare. Ad esempio, la probabilità di ottenere una sequenza 00000111 sarebbe stata la stessa.

CHAPTER 6

Variabili aleatorie discrete

1. Variabili aleatorie

Ci troviamo spesso nella situazione di volere o dovere considerare delle quantità che sono funzioni del risultato di un fenomeno aleatorio (o che possono essere il risultato stesso).

ESEMPIO 6.1. Faccio 3 puntate alla roulette e mi propongo di studiare la quantità $X = \text{vincita}$. In particolare vorrei sapere qual è la probabilità che la vincita sia maggiore dei soldi puntati.

Una domanda più sottile cui risponderemo più avanti è: "mediamente quanto mi devo aspettare di vincere (o perdere) se non sono particolarmente fortunato o sfortunato?".

Vediamo un esempio concreto.

ESEMPIO 6.2. Supponiamo che il nostro fenomeno aleatorio sia costituito da sei lanci di un dado. In questo caso è naturale considerare come risultati tutte le sequenze di lunghezza 6 in $\{1, 2, 3, 4, 5, 6\}$ abbiamo quindi 6^6 possibili risultati. Consideriamo le quantità aleatorie, che dipendono dal risultato $X =$ numero di volte in cui abbiamo ottenuto 6, e $Y =$ numero di volte in cui un lancio ha dato un risultato maggiore del precedente. La variabile X assume valori da 0 a 6 e potremmo chiederci qual è la probabilità che X sia almeno 2. La variabile Y assume valori da 0 a 6 e potremmo chiederci qual è la probabilità che $Y = 2$. Impareremo più avanti a rispondere a queste domande (ma se vuoi puoi provarci da subito!)

Diamo ora la definizione formale di variabile aleatoria in modo da poterne sviluppare la relativa teoria.

DEFINIZIONE. Sia (Ω, \mathcal{A}, P) uno spazio di probabilità. Una funzione $X : \Omega \rightarrow \mathbb{R}$ è una variabile aleatoria se il sottoinsieme di Ω dato da

$$(4) \quad \{\omega \in \Omega : X(\omega) \leq a\}$$

è un evento per ogni $a \in \mathbb{R}$.

ESEMPIO 6.3. Si consideri $\Omega = \mathbb{N}_{>0}$ e $\mathcal{A} = \{\mathbb{N}, \emptyset, 2\mathbb{N}, 2\mathbb{N}+1\}$. In questo caso la funzione identità $X(n) = n$ non è una variabile aleatoria. Infatti se scelgo ad esempio $a = 2$ abbiamo

$$\{n \in \mathbb{N}_{>0} : X(n) \leq 2\} = \{1, 2\}$$

che non è un evento (perché gli eventi sono solo $\mathbb{N}, \emptyset, 2\mathbb{N}, 2\mathbb{N}+1$). Se invece consideriamo la funzione $Y(n) = (-1)^n$ abbiamo

$$\{n \in \mathbb{N}_{>0} : Y(n) \leq a\} = \begin{cases} \emptyset & \text{se } a < -1 \\ 2\mathbb{N} + 1 & \text{se } -1 \leq a < 1 \\ \mathbb{N} & \text{se } a \geq 1 \end{cases}$$

Abbiamo quindi un evento per ogni $a \in \mathbb{R}$ e di conseguenza Y è una variabile aleatoria secondo la definizione data qui sopra.

La scelta del simbolo \leq non è importante nella definizione di variabile aleatoria, come mostra il prossimo risultato.

PROPOSIZIONE 6.4. *Sia X una variabile aleatoria. Allora per ogni $a \in \mathbb{R}$ abbiamo*

- a) $\{\omega \in \Omega : X(\omega) > a\}$ è un evento;
- b) $\{\omega \in \Omega : X(\omega) < a\}$ è un evento;
- c) $\{\omega \in \Omega : X(\omega) \geq a\}$ è un evento;
- d) $\{\omega \in \Omega : X(\omega) = a\}$ è un evento;
- e) $\{\omega \in \Omega : X(\omega) \neq a\}$ è un evento;

DIMOSTRAZIONE. L'insieme in a) è il complementare di (4) è quindi è un evento.

L'insieme in b) è un evento perché possiamo scriverlo come unione numerabile di eventi:

$$\{\omega \in \Omega : X(\omega) < a\} = \bigcup_{i=1}^n \{\omega \in \Omega : X(\omega) \leq a - 1/n\}.$$

L'insieme in c) è il complementare di quello in a), l'insieme in d) è l'intersezione tra quello in (4) e quello in c), l'insieme in e) è il complementare di quello in d). Siccome sappiamo che la famiglia degli eventi è chiusa per complementare, intersezioni e unioni numerabile il risultato segue. \square

Se X è una variabile aleatoria l'insieme $\{\omega \in \Omega : X(\omega) \leq t\}$ è quindi un evento per ogni t per definizione e possiamo quindi calcolarne la probabilità. Saremo quindi spesso interessati a considerare la funzione $F_X : \mathbb{R} \rightarrow \mathbb{R}$

$$F_X(t) = P(\{\omega : X(\omega) \leq t\}).$$

che chiameremo *funzione di ripartizione* di X . D'ora in avanti considereremo quasi sempre variabili aleatorie in termini della loro funzione di ripartizione (o equivalentemente della loro *densità* come vedremo più avanti), talvolta dimenticandoci dello spazio di probabilità su cui sono definite, perché questo sarà influente.

Per semplificare la notazione scriveremo ad esempio $\{X \leq t\}$ per indicare l'evento $\{\omega \in \Omega : X(\omega) \leq t\}$, e similmente scriveremo $P(X \leq t)$ anziché $P(\{\omega \in \Omega : X(\omega) \leq t\})$. Notazioni analoghe valgono sostituendo il simbolo \leq con uno dei simboli $<, >, \geq, =, \neq$.

2. Variabili discrete

Una variabile aleatoria X si dice *discreta* se i valori che assume sono finiti oppure numerabili. Nel caso siano numerabili molto spesso potremo pensare che siano semplicemente i numeri interi, ma non vogliamo escludere anche casi in cui abbiamo dei valori non interi.

ESEMPIO 6.5. Consideriamo il fenomeno aleatorio dato dal normale utilizzo di uno smartphone fino alla rottura. Possiamo considerare le seguenti funzioni: T =tempo trascorso; X =numero di telefonate effettuate; Y la variabile che vale 1 sono state effettuate mediamente più di 5 telefonate al giorno, e vale 0 altrimenti.

La variabile T non è una variabile discreta: il tempo trascorso è un qualsiasi numero reale. Tratteremo il caso di variabili continue in un altro capitolo;

La variabile X può assumere un qualunque valore intero non negativo: è quindi una variabile discreta;

La variabile Y può assumere solo i valori 0 e 1: è una variabile finita e quindi anche discreta.

Introduciamo ora il concetto di densità di una variabile discreta.

DEFINIZIONE. Data una variabile aleatoria (discreta) X la funzione $p_X : \mathbb{R} \rightarrow \mathbb{R}$ data da $p_X(h) = P(X = h)$ si dice *densità discreta* (concreta) della variabile aleatoria X .

Osserviamo che la funzione p_X soddisfa le seguenti proprietà.

- (1) $p_X(h) \neq 0$ per una quantità al più numerabile di valori h ;
- (2) $p_X(h) \geq 0$ per ogni $h \in \mathbb{R}$;
- (3) $\sum_{h \in \mathbb{R}} p_X(h) = 1$;

Abbiamo preferito usare la lettera h anziché la lettera t proprio per ricordare il fatto che la variabile è discreta e che quindi, solitamente, assume solo valori interi. Osserviamo anche che la somma che appare in (3) è una serie numerica in quanto, per (1) i termini diversi da zero sono al più numerabili. È inoltre ben definita in quanto i termini sono tutti non negativi grazie alla proprietà (2).

Le proprietà della densità discreta di una variabile X suggeriscono la seguente definizione

DEFINIZIONE. Una funzione $p : \mathbb{R} \rightarrow \mathbb{R}$ si dice densità discreta (astratta) se soddisfa le seguenti proprietà:

- (1) $p(h) \neq 0$ per una quantità al più numerabile di valori h ;
- (2) $p(h) \geq 0$ per ogni $h \in \mathbb{R}$;
- (3) $\sum_{h \in \mathbb{R}} p(h) = 1$;

ESEMPIO 6.6. Consideriamo la funzione $p : \mathbb{R} \rightarrow \mathbb{R}$ data da:

$$p(h) = \begin{cases} 1/2 & \text{se } h = -1 \\ 1/4 & \text{se } h = 1, 3 \\ 0 & \text{altrimenti} \end{cases}$$

Soddisfa le 3 proprietà qui sopra e quindi è una densità discreta. Consideriamo ora la funzione

$$p'(h) = \begin{cases} 1/2^h & \text{se } h = 1, 2, 3, \dots \\ 0 & \text{altrimenti} \end{cases}$$

Anche p' soddisfa le proprietà di densità astratta: la (1) e la (2) sono evidenti per la (3) abbiamo

$$\sum_{h \in \mathbb{R}} p'(h) = \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \dots = 1.$$

Nelle prossime sezioni partiremo talvolta da un fenomeno aleatorio concreto e quindi da una variabile aleatoria con la sua densità concreta, altre volte partiremo da una densità astratta per cercare poi di capire in quali situazioni concrete si potrà applicare.

3. La densità uniforme

Come già nel caso degli spazi di probabilità il caso più semplice di variabile aleatoria è quella uniforme.

Sia $A = \{x_1, \dots, x_n\}$; una variabile X che assume i valori in A tutti con la stessa probabilità $\frac{1}{n}$ si dice *variabile uniforme* su A . Scriviamo in questo caso $X \sim U(A)$ oppure $X \sim U(x_1, \dots, x_n)$. La densità è quindi

$$p_X(h) = \begin{cases} \frac{1}{n} & \text{se } h \in \{x_1, x_2, \dots, x_n\} \\ 0 & \text{altrimenti.} \end{cases}$$

ESEMPIO 6.7. Nel lancio di un dado sia $\Omega = \{1, 2, 3, 4, 5, 6\}$ l'insieme dei possibili risultati. Allora la funzione $X(n) = n$ per ogni $n \in \Omega$ è una variabile aleatoria uniforme $X \sim U(1, 2, 3, 4, 5, 6)$.

Cosa possiamo dire della funzione $Y(\omega) = \cos(n\pi)$? Si ha che anche Y è uniforme, ed in particolare $Y \sim U(-1, 1)$. La variabile $Z = \cos(n\pi/2)$ invece non è uniforme: abbiamo in questo caso

$$p_Z(h) = \begin{cases} 1/2 & h=0 \\ 1/6 & h=-1 \\ 1/3 & h=1 \\ 0 & \text{altrimenti.} \end{cases}$$

4. La funzione caratteristica e la densità di Bernoulli

Se da un lato la variabile uniforme risulta semplice perchè la sua densità è costante sui valori che assume, un altro esempio di variabile elementare è dato dalle variabili di Bernoulli perchè queste possono assumere solo i valori 0 e 1.

Se A è un evento la sua funzione caratteristica χ_A data da

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

è una variabile aleatoria. Infatti, verifichiamo che soddisfa la definizione, cioè che $\{\chi_A \leq a\}$ è un evento per ogni $a \in \mathbb{R}$:

$$\{\chi_A \leq a\} = \begin{cases} \emptyset & \text{se } a < 0 \\ A^C & \text{se } 0 \leq a < 1 \\ \Omega & \text{se } a > 1. \end{cases}$$

La sua densità è data da

$$p(h) = \begin{cases} P(A) & \text{se } h = 1 \\ 1 - P(A) & \text{se } h = 0 \\ 0 & \text{se } h \neq 0, 1 \end{cases}$$

DEFINIZIONE. Una variabile aleatoria X che assume solo i valori 0 e 1 si dice variabile di Bernoulli.

Se la probabilità che assuma il valore 1 è p (e quindi la probabilità che assuma il valore 0 è $1 - p$) scriviamo $X \sim B(1, p)$. Possiamo quindi concludere che per ogni evento A la variabile χ_A è una variabile di Bernoulli

$$\chi_A \sim B(1, P(A)).$$

5. La densità binomiale

Una naturale generalizzazione delle variabili di Bernoulli è data dalle variabili binomiali. Prima di introdurle nel dettaglio vediamo come sempre qualche esempio.

ESEMPIO 6.8. Si consideri il fenomeno aleatorio dato da 3 lanci di una moneta. Sia X la variabile che conta quante volte è uscita testa. Determinare la densità di X . La variabile X può assumere i valori 0, 1, 2, 3. Dire $X = 3$ è equivalente a richiedere che escano 3 teste nei 3 lanci: siccome i risultati dei 3 lanci sono indipendenti l'uno dall'altro abbiamo $P(X = 3) = p_X(3) = (1/2)^3 = 1/8$. Dire $X = 2$ è equivalente a richiedere che escano 2 teste nei 3 lanci e questo si può verificare in 3 modi diversi, TTC, TCT o CTT, ognuno dei quali ha probabilità $1/8$. Possiamo analogamente vedere anche gli altri casi e concludere che la densità di X è

$$p_X(h) = \begin{cases} 1/8 & \text{se } h=0,3 \\ 3/8 & \text{se } h=1,2 \\ 0 & \text{altrimenti.} \end{cases}$$

ESEMPIO 6.9. Supponiamo di andare al casinò con 15 Euro e puntiamo 3 volte 5 Euro alla roulette sul rosso. Consideriamo la variabile aleatoria X data dai soldi che abbiamo in tasca alla fine delle 3 puntate. La probabilità di ottenere un numero rosso è di $\frac{18}{37}$ ad ogni tentativo. Determinare la densità di X .

Il problema è molto simile a quello precedente. La variabile X in questo caso può assumere i valori 30, 20, 10, 0 a seconda di quanti rossi sono usciti. Abbiamo

$$\begin{aligned} P(X = 30) &= \left(\frac{18}{37}\right)^3 = 0,115 \\ P(X = 20) &= 3 \cdot \left(\frac{18}{37}\right)^2 \cdot \frac{19}{37} = 0,365 \\ P(X = 10) &= 3 \cdot \left(\frac{18}{37}\right) \cdot \left(\frac{19}{37}\right)^2 = 0,385 \\ P(X = 0) &= \left(\frac{19}{37}\right)^3 = 0,135. \end{aligned}$$

Consideriamo un fenomeno aleatorio e un certo evento ad esso associato. Per semplicità diciamo che se l'evento si realizza otteniamo un "successo", viceversa che abbiamo ottenuto un "insuccesso". Negli esempi precedenti i successi erano rappresentati dal "testa" e "rosso" rispettivamente. Uno schema successo-insuccesso consiste nel ripetere questo fenomeno un certo fissato numero di volte, dette anche prove o tentativi, in cui siamo interessati unicamente al numero di successi ottenuti nelle varie prove.

Supponiamo ora che le prove siano indipendenti l'una dall'altra, cioè vengono effettuate tutte nelle medesime condizioni, indipendentemente dal risultato delle prove precedenti: parliamo in

questo caso di schema successo-insuccesso con ripetizione o a prove indipendenti. Consideriamo quindi la variabile X data dal numero di successi in n tentativi. Abbiamo già visto nell'Esempio 5.17 del Cap. 2 e nella risoluzione degli Esempi 6.8 e 6.9 che, nel caso in cui le prove siano indipendenti l'una dall'altra, la probabilità di ottenere una certa prefissata sequenza di successi e insuccessi dipende solo dal numero di questi ultimi e non dall'ordine con cui vogliamo che si succedano. La probabilità di ottenere una sequenza fissata in cui abbiamo k successi e $n - k$ insuccessi è data da $p^k(1 - p)^{n-k}$, dove p è la probabilità che ogni singolo tentativo abbia successo. Siccome i modi in cui possiamo scegliere le k posizioni per i successi sono date da $\binom{n}{k}$ otteniamo per la variabile X la densità data da

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

Scriveremo in questo caso $X \sim B(n, p)$ e diciamo che X è una variabile binomiale.

ESEMPIO 6.10. Un giocatore di basket realizza mediamente l'85% dei tiri liberi. Determinare la probabilità che realizzi esattamente 85 dei prossimi 100 tiri liberi. Supponiamo (anche se non è vero!) che ogni tiro libero sia indipendente dagli altri e che questi siano effettuati sempre nelle medesime condizioni. Possiamo quindi assumere che la variabile X = numero di tiri liberi realizzati nei prossimi 100 is aun avariabile binomiale $X \sim B(100, 0.85)$. Abbiamo quindi

$$P(X = 85) = \binom{100}{85} 0.85^{85} \cdot 0.15^{15} = 0,111.$$

6. La densità ipergeometrica

Un altro tipo di schema successo-insuccesso è il cosiddetto schema successo-insuccesso *senza ripetizione*. In questo caso il fenomeno aleatorio consiste nell'estrazione di n palline, senza rimpiazzare di volta in volta la pallina estratta, da una scatola contenente b palline bianche e r palline rosse. Diciamo che la prova (estrazione) dà successo se viene estratta una pallina bianca e consideriamo la variabile X data dal numero di successi ottenuti, cioè dal numero di palline bianche estratte.

Osserviamo che X può assumere tutti i valori tra $\max(0, n - r)$ e $\min(n, b)$: infatti se $n > r$ siamo sicuri di pescare almeno $n - r$ palline bianche e chiaramente il numero di bianche estratte non può superare né il numero estrazioni né il numero di palline bianche. Lo spazio Ω è dato da tutti i possibili sottoinsiemi di n oggetti (le palline estratte) scelti da un insieme di $b + r$ oggetti (le palline disponibili), con probabilità uniforme. L'evento " $X = k$ " è dato quindi da tutti i sottoinsiemi delle $b + r$ palline costituiti da k palline bianche e $n - k$ palline rosse. Le k bianche le posso scegliere in $\binom{b}{k}$ modi. Le $n - k$ rosse in $\binom{r}{n-k}$ modi. Abbiamo quindi che la densità di X è data da

$$p_X(k) = \begin{cases} \frac{\binom{b}{k} \binom{r}{n-k}}{\binom{b+r}{n}} & \text{se } k = \max(0, n - r), \dots, \min(n, b) \\ 0 & \text{altrimenti.} \end{cases}$$

Scriviamo in questo caso $X \sim H(n; b, r)$ e diciamo che X è una variabile ipergeometrica.

I prossimi due esercizi vanno visti in parallelo.

ESERCIZIO 6.11. Bulloni prodotti da una fabbrica sono difettosi con probabilità del 20% e vengono commercializzati in confezioni da 3. Qual è la probabilità che in una scatola ci sia al più un bullone difettoso? 0.896

ESERCIZIO 6.12. Un'urna contiene 8 palline rosse e 2 bianche. Ne estraiamo 3 senza rimpiazzo. Qual è la probabilità di estrarne al più una bianca? In questo caso, a differenza del precedente abbiamo che la variabile X = numero di bianche estratte è una ipergeometrica $X \sim H(3; 2, 8)$ per cui

$$P(X \leq 1) = P(X = 0) + P(X = 1) = \frac{\binom{2}{0} \cdot \binom{8}{3}}{\binom{10}{3}} + \frac{\binom{2}{1} \cdot \binom{8}{2}}{\binom{10}{3}} = \frac{14}{15} = 0.933.$$

7. La densità geometrica

In uno schema successo-insuccesso, anziché al numero di successi ottenuti su n tentativi, potremmo essere interessati al numero di tentativi che devo effettuare prima di ottenere un successo. Vediamo qualche esempio.

ESEMPIO 6.13. Consideriamo un'urna con 2 palline bianche e 3 rosse. Le estraggo senza rimpiazzo finché non trovo una pallina rossa. Considero la variabile aleatoria X = numero di estrazioni per trovare una pallina rossa. X è una variabile finita in quanto assume solo i valori 1, 2, 3. La sua densità si determina facilmente ed è data da

$$p_X(k) = \begin{cases} \frac{6}{10} & \text{se } k = 1 \\ \frac{3}{10} & \text{se } k = 2 \\ \frac{1}{10} & \text{se } k = 3 \\ 0 & \text{altrimenti.} \end{cases}$$

ESEMPIO 6.14. Consideriamo ora lo stesso tipo di fenomeno in cui però rimpiazziamo di volta in volta in volta la pallina estratta. La variabile X diventa numerabile e si ha

$$p(k) = \begin{cases} \left(\frac{2}{5}\right)^{k-1} \frac{3}{5} & \text{se } k = 1, 2, 3, \dots \\ 0 & \text{altrimenti.} \end{cases}$$

Infatti richiedere $X = k$ vuol dire aver ottenuto $k - 1$ insuccessi nei primi $k - 1$ tentativi (ognuno con probabilità $2/5$) per poi finalmente ottenere un successo al k -esimo tentativo (con probabilità $3/5$).

Andiamo a formalizzare in generale quanto osservato in questo ultimo esempio. Si consideri uno schema successo-insuccesso con ripetizioni e prendiamo in esame la variabile T data dal tempo di primo successo. In altre parole T è il numero di prove effettuate per ottenere il primo successo. Si può considerare ad esempio il lancio di un dado in cui il successo è dato dal 6. Se chiamiamo X_i le variabili $B(1, p)$ date da $X_i = 1$ se l' i -esimo tentativo dà successo e $X_i = 0$ altrimenti abbiamo

$$\begin{aligned} P(T = k) &= P(X_1 = X_2 = \dots = X_{k-1} = 0 \text{ e } X_k = 1) \\ &= P(X_1 = 0)P(X_2 = 0) \dots P(X_{k-1} = 0)P(X_k = 1) \\ &= (1 - p)^{k-1}p. \end{aligned}$$

La densità di T è quindi

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

È questa quella che si dice la *densità geometrica modificata* e scriveremo in questo caso $T \sim \tilde{G}(p)$ ed è il primo esempio di variabili discrete che vediamo e che possono assumere una quantità infinita di valori. La densità geometrica standard si può vedere come una densità geometrica modificata in cui è stato effettuato uno shift di 1. Più precisamente, una variabile geometrica standard la possiamo vedere come il numero di insuccessi ottenuti in uno schema successo-insuccesso prima di ottenere un successo. In particolare abbiamo che tale densità è data da

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

Se X è una variabile di densità geometrica standard di parametro p scriveremo $X \sim G(p)$. Osserviamo che questa è effettivamente una densità discreta astratta in quanto, ricordando la somma della serie geometrica

$$\sum_{k=0}^{\infty} a^k = \lim_{n \rightarrow \infty} (1 + a + \dots + a^n) = \lim_{n \rightarrow \infty} \frac{1 - a^{n+1}}{1 - a} = \frac{1}{1 - a}$$

se $0 \leq a < 1$, abbiamo che

$$\sum_{k=0}^{\infty} p(k) = \sum_{k=0}^{\infty} p(1-p)^k = p \sum_{k=0}^{\infty} (1-p)^k = p \cdot \frac{1}{p} = 1.$$

Abbiamo quindi che se $T \sim \tilde{G}(p)$ è una variabile di densità geometrica modificata, allora $T - 1 \sim G(p)$ è una variabile geometrica standard.

ESEMPIO 6.15. Si lancino simultaneamente due dadi, uno rosso e uno blu. Ripetiamo questo esperimento finché non otteniamo due 6. Poniamo X = primo lancio in cui il dado rosso ha dato 6, Y = primo lancio in cui il dado blu ha dato 6 e Z = numero di lanci complessivi. a) Determinare la densità delle tre variabili X , Y e Z . b) Determinare inoltre la probabilità di ottenere 6 con il dado rosso prima di ottenere 6 con il dado blu. a) Per costruzione abbiamo che X e Y sono variabili geometriche modificate di parametro $1/6$, $X, Y \sim \tilde{G}(1/6)$. Anche Z è una geometrica modificata: $Z \sim \tilde{G}(1/36)$: abbiamo sempre prove indipendenti, e la probabilità di avere successo (in questo caso due 6) è proprio $1/36$. b) Questo è decisamente più complicato! Ma vediamo di arrivarci. Abbiamo

$$1 = P(X < Y) + P(X > Y) + P(X = Y)$$

e chiaramente per simmetria del problema $P(X < Y) = P(X > Y)$ da cui

$$P(X < Y) = \frac{1 - P(X = Y)}{2}.$$

Ci basterà calcolare quindi $P(X = Y)$:

$$P(X = Y) = \sum_{k=1}^{\infty} P(X = Y = k) = \sum_{k=1}^{\infty} \left(\frac{25}{36}\right)^{k-1} \cdot \frac{1}{36} = \frac{1}{1 - 25/36} \cdot \frac{1}{36} = \frac{1}{11},$$

dove abbiamo ancora utilizzato la somma di una serie geometrica, e quindi

$$P(X < Y) = \frac{5}{11}.$$

Sia X una variabile aleatoria che assume solo valori interi non negativi. Diciamo che X gode della proprietà di *mancanza di memoria* se per ogni $k, m \in \mathbb{N}$ si ha

$$P(X \geq k + m | X \geq k) = P(X \geq m).$$

Nel caso in cui X rappresenta il tempo di attesa affinché un certo evento si verifichi la proprietà di mancanza di memoria dice che l'aver già aspettato un certo tempo k non influisce in nessun modo sulla probabilità che il tempo che dobbiamo ancora aspettare sia almeno m .

Se X è una variabile geometrica questa gode della proprietà di “mancanza di memoria”. Infatti, osserviamo subito che

$$P(X \geq k) = \sum_{i=k}^{\infty} p(1-p)^i = p(1-p)^k \sum_{i=0}^{\infty} (1-p)^i = (1-p)^k.$$

In alternativa, e più semplicemente, si poteva pensare al fatto che X rappresenta il numero di insuccessi prima di ottenere un successo e quindi l'evento “ $X \geq k$ ” consiste nell'aver ottenuto k insuccessi nei primi k tentativi, e questo ha chiaramente probabilità $(1-p)^k$. Abbiamo ora

$$\begin{aligned} P(X \geq k + m | X \geq k) &= \frac{P(X \geq k + m)}{P(X \geq k)} \\ &= \frac{(1-p)^{k+m}}{(1-p)^k} \\ &= (1-p)^m \\ &= P(X \geq m). \end{aligned}$$

La proprietà di mancanza di memoria caratterizza la densità geometrica nel senso del seguente risultato.

PROPOSIZIONE 6.16. *Sia X una variabile aleatoria che assume solo valori interi non negativi e che soddisfa la proprietà di mancanza di memoria $P(X \geq k + m | X \geq k) = P(X \geq m)$. Allora X è una variabile geometrica.*

DIMOSTRAZIONE. (non fatta) Infatti osserviamo intanto che se X gode della proprietà di mancanza di memoria allora si ha anche

$$P(X = k + m | X \geq k) = P(X = m).$$

Per convincersi di questo fatto è sufficiente pensare che

$$P(X = m) = P(X \geq m) - P(X \geq m + 1)$$

e similmente

$$P(X = k + m | X \geq k) = P(X \geq k + m | X \geq k) - P(X \geq k + m + 1 | X \geq k)$$

per poi usare la proprietà di mancanza di memoria.

Posto quindi $p = P(X = 0)$ procediamo per induzione e supponiamo che $P(X = i) = p(1-p)^i$ per ogni $i < k$. Abbiamo quindi

$$P(X \geq k) = 1 - P(X < k) = 1 - \sum_{i=0}^{k-1} p(1-p)^i = 1 - p \frac{1 - (1-p)^k}{1 - (1-p)} = (1-p)^k$$

e di conseguenza, siccome la proprietà di mancanza di memoria per $m = 0$ ci dà

$$P(X = k | X \geq k) = \frac{P(X = k)}{P(X \geq k)} = P(X = 0)$$

concludiamo che

$$P(X = k) = P(X = 0)P(X \geq k) = p(1 - p)^k.$$

□

8. La densità di Poisson

Se consideriamo una variabile binomiale $X \sim B(n, p)$ in cui il numero n di tentativi è molto grande e il parametro p molto piccolo fare i conti con la densità binomiale risulta essere computazionalmente dispendioso. Facciamo un esempio:

ESEMPIO 6.17. Supponiamo che in Italia nascano 1 milione di bambini ogni anno e che ci sia una malattia rara che colpisce un bambino su 200mila. Qual è la probabilità che quest'anno nascano più di 5 bambini affetti da questa malattia? La variabile $X =$ "numero di bambini nati con questa malattia quest'anno" è una variabile binomiale $X = B(10^6, 1/200000)$. Tuttavia se proviamo ad effettuare il calcolo $P(X > 5)$, anche se abbiamo una formula esatta per effettuarlo, è inevitabilmente molto complesso.

Per risolvere questo tipo di problema introduciamo la seguente

DEFINIZIONE. La densità (astratta) di Poisson è data da

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

dove λ è un parametro reale strettamente positivo.

Il fatto che questa sia effettivamente una densità viene dallo sviluppo di Taylor $e^x = \sum_{k \geq 0} \frac{x^k}{k!}$, per $x = \lambda$. Infatti

$$\sum_{k=0}^{\infty} e^{-\lambda} \frac{\lambda^k}{k!} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^{-\lambda} e^{\lambda} = 1.$$

Se X è una variabile aleatoria che ha come densità la densità astratta di Poisson di parametro λ scriviamo $X \sim P(\lambda)$. Ricordando anche l'altro limite notevole

$$\left(1 + \frac{x}{n}\right)^n \longrightarrow e^x,$$

per $n \rightarrow +\infty$, vogliamo verificare che una variabile di Poisson approssima sotto certe ipotesi una variabile binomiale. Abbiamo infatti

PROPOSIZIONE 6.18. Sia $\lambda > 0$. Allora per $n \rightarrow +\infty$ la densità binomiale $B(n, \lambda/n)$ tende alla densità di Poisson $P(\lambda)$.

DIMOSTRAZIONE. Sia $X \sim B(n, \lambda/n)$ e $p(k)$ la sua densità. Abbiamo densità di X è data da

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

□

In conclusione abbiamo che una variabile $B(n, p)$ con n molto grande e p piccolo può essere approssimata con una variabile di Poisson $P(np)$ di parametro $\lambda = np$. In statistica si adotta l'approssimazione della distribuzione binomiale tramite la distribuzione di Poisson quando $n > 20$ e $p < 1/20$, o preferibilmente quando $n > 100$ e $np < 10$.

Come già anticipato i casi in cui è utile ed indicato utilizzare una variabile di Poisson sono quelli in cui abbiamo un gran numero di tentativi, ognuno dei quali dà successo con probabilità molto bassa.

Riprendiamo l'Esempio 6.17: in questo caso $np = 5$ per cui possiamo pensare alla variabile X come ad una variabile di Poisson di parametro 5. Otteniamo

$$\begin{aligned}
 P(X > 5) &= 1 - P(X = 0) - P(X = 1) - P(X = 2) - P(X = 3) - P(X = 4) \\
 &= 1 - e^{-5} \left(1 + 5 + \frac{25}{2} + \frac{125}{6} + \frac{625}{24} + \frac{3125}{120}\right) = 0,384.
 \end{aligned}$$

ESEMPIO 6.19. Numero di telefonate ad un call center di una certa azienda: ogni cliente lo possiamo interpretare come un tentativo che dà successo se chiama in un determinato giorno. É chiaro che in questo caso i tentativi sono tanti e la probabilità che uno di essi dia successo è molto bassa.

Un altro esempio di variabile di Poisson potrebbe essere ad esempio il numero di auto gialle che passano in una certa strada in un giorno.

CHAPTER 7

Variabili discrete multidimensionali

1. Densità congiunte e marginali

Siamo spesso interessati a considerare diverse variabili aleatorie definite sullo stesso spazio di probabilità. Se abbiamo m variabili aleatorie discrete X_1, \dots, X_m possiamo considerare la funzione

$$\underline{X} = (X_1, \dots, X_m) : \Omega \rightarrow \mathbb{R}^m.$$

Questa si dice *variabile aleatoria m dimensionale*. Osserviamo che se $\underline{x} = (x_1, \dots, x_m) \in \mathbb{R}^m$, allora

$$\{\underline{X} = \underline{x}\} = \{X_1 = x_1\} \cap \dots \cap \{X_m = x_m\}.$$

e deduciamo quindi che $\{\underline{X} = \underline{x}\}$ è un evento per ogni $\underline{x} \in \mathbb{R}^m$.

DEFINIZIONE. La densità di una variabile m -dimensionale \underline{X} è la funzione $p_{\underline{X}} : \mathbb{R}^m \rightarrow \mathbb{R}_{\geq 0}$ data da $p(\underline{x}) = P(\underline{X} = \underline{x})$. Questa densità si dice anche *densità congiunta* delle variabili X_1, \dots, X_m .

ESEMPIO 7.1. Ho una scatola con 2 palline rosse, 2 bianche e 2 verdi. Estraggo 2 palline e considero la variabile X_1 = palline rosse estratte e X_2 = palline bianche estratte. Determinare la densità congiunta delle variabili X_1 ed X_2 . Si ha

$$\begin{aligned} p(0, 0) &= \frac{1}{\binom{6}{2}} = \frac{1}{15}; \\ p(1, 1) &= \frac{4}{15}; \\ p(1, 0) &= \frac{4}{15} = p(0, 1); \\ p(2, 0) &= \frac{1}{15} = p(0, 2). \end{aligned}$$

Osserviamo che effettivamente $\sum_{\underline{x} \in \mathbb{R}^2} p(\underline{x}) = 1$.

Quando si ha a che fare con m variabili aleatorie X_1, \dots, X_m le singole densità di queste ultime si dicono *densità marginali*, per distinguerle dalla densità congiunta. Le indichiamo solitamente con p_1, \dots, p_m .

ESEMPIO 7.2. Estraiamo con e senza rimpiazzo due palline da un'urna contenente 2 palline bianche e 3 rosse. Nel caso con rimpiazzo poniamo

$$X_i = \begin{cases} 1 & \text{se l}'i\text{-esima estratta è bianca} \\ 0 & \text{se l}'i\text{-esima estratta è rossa} \end{cases}$$

per $i = 1, 2$, e nel caso senza rimpiazzo poniamo similmente

$$Y_i = \begin{cases} 1 & \text{se l'i-esima estratta è bianca} \\ 0 & \text{se l'i-esima estratta è rossa} \end{cases}.$$

Determinare le densità marginali delle variabili X_1, X_2, Y_1, Y_2 e le densità congiunte di $\underline{X} = (X_1, X_2)$ e di $\underline{Y} = (Y_1, Y_2)$. Le variabili X_1, X_2, Y_1, Y_2 sono chiaramente tutte variabili di Bernoulli in quanto assumono solo i valori 0 e 1. Sappiamo inoltre (si veda anche l'Esempio 5.10 del Capitolo 2) che il parametro p è dato da $2/5$ per tutte e quattro le variabili e quindi abbiamo che le densità marginali sono tutte $B(1, 2/5)$. Nonostante ciò le densità congiunte sono distinte. Abbiamo infatti

$$\begin{aligned} p_{\underline{X}}(0, 0) &= \frac{9}{25}; \\ p_{\underline{X}}(0, 1) &= \frac{6}{25}; \\ p_{\underline{X}}(1, 0) &= \frac{6}{25}; \\ p_{\underline{X}}(1, 1) &= \frac{4}{25}. \end{aligned}$$

e

$$\begin{aligned} p_{\underline{Y}}(0, 0) &= \frac{3}{10}; \\ p_{\underline{Y}}(0, 1) &= \frac{3}{10}; \\ p_{\underline{Y}}(1, 0) &= \frac{3}{10}; \\ p_{\underline{Y}}(1, 1) &= \frac{1}{10}. \end{aligned}$$

Questo esempio mostra come le densità marginali non determinino univocamente la densità congiunta. È invece vero il contrario: vediamo per semplicità il caso $m = 2$.

PROPOSIZIONE 7.3. *Sia $p(h, k)$ la densità congiunta delle variabili X, Y , e siano p_X e p_Y le rispettive densità marginali. Allora*

$$p_X(h) = \sum_{k \in \mathbb{R}} p(h, k)$$

per ogni $h \in \mathbb{R}$ e analogamente

$$p_Y(k) = \sum_{h \in \mathbb{R}} p(h, k)$$

per ogni $k \in \mathbb{R}$.

DIMOSTRAZIONE. Sia $h \in \mathbb{R}$. Allora

$$\begin{aligned} p_X(h) &= P(X = h) \\ &= P\left(\bigcup_{k \in \mathbb{R}} (X = h, Y = k)\right) \\ &= \sum_{k \in \mathbb{R}} P(X = h, Y = k) \\ &= \sum_{k \in \mathbb{R}} p(h, k). \end{aligned}$$

□

In questa dimostrazione abbiamo usato la simbologia $\bigcup_{k \in \mathbb{R}}$ e $\sum_{k \in \mathbb{R}}$ facendo un piccolo abuso di notazione. Anche se l'indice di unione e di somma è non numerabile gli insiemi che si considerano nel primo caso sono non vuoti solo per una quantità numerabile di indici; così come i numeri che si considerano nel secondo caso sono diversi da zero solo per una quantità numerabile di indici. Gli oggetti che si considerano sono quindi in ogni caso sempre numerabili e le operazioni sono quindi sempre lecite.

2. Variabili aleatorie indipendenti

Il concetto di indipendenza tra eventi visto nel Capitolo 2 si generalizza in modo naturale all'indipendenza tra variabili aleatorie. Una variabile X sarà indipendente da una variabile Y se la condizione che Y assuma un certo valore non va a modificare la probabilità che X ne assuma un altro. Formalmente, abbiamo

DEFINIZIONE. Date n variabili aleatorie X_1, \dots, X_n , diciamo che esse sono indipendenti se per ogni $h_1, \dots, h_n \in \mathbb{R}$ gli eventi $\{X_1 \leq h_1\}, \dots, \{X_n \leq h_n\}$ sono indipendenti. In altre parole le variabili X_1, \dots, X_n sono indipendenti se vale la relazione

$$P(X_1 \leq h_1, \dots, X_n \leq h_n) = P(X_1 \leq h_1) \cdots P(X_n \leq h_n).$$

Nel caso in cui le variabili X_1, \dots, X_n sono discrete il simbolo \leq nella definizione di variabili indipendenti può essere sostituito con il simbolo $=$

PROPOSIZIONE 7.4. *Siano X_1, \dots, X_n variabili aleatorie discrete. Esse sono indipendenti se e solo se per ogni h_1, \dots, h_n si ha*

$$P(X_1 = h_1, \dots, X_n = h_n) = P(X_1 = h_1) \cdots P(X_n = h_n),$$

o, detto in altri termini, se la densità congiunta è il prodotto delle marginali:

$$p_{X_1, \dots, X_n}(h_1, \dots, h_n) = p_{X_1}(h_1) \cdots p_{X_n}(h_n).$$

DIMOSTRAZIONE. (non fatta) Vediamo la dimostrazione nel caso in cui abbiamo solo due variabili X e Y ...

□

Come conseguenza di questo risultato abbiamo che se le nostre variabili sono indipendenti possiamo calcolare la densità congiunta se conosciamo le marginali. Questo fatto possiamo verificarlo nell'Esempio 7.2 nel caso delle estrazioni con rimpiazzo. Nel caso senza rimpiazzo le variabili Y_1

e Y_2 sono dipendenti. Lo vediamo nel caso generale in cui abbiamo b palline bianche ed r palline rosse: si ha

$$P(X_1 = 1, X_2 = 1) = \frac{b}{b+r} \frac{b-1}{b+r-1} \neq P(X_1 = 1)P(X_2 = 1) = \frac{b}{b+r} \frac{b}{b+r}.$$

e quindi la densità congiunta non è il prodotto delle marginali.

Osserviamo che per mostrare che due variabili X e Y sono indipendenti dobbiamo mostrare che $p(h, k) = p_X(h)p_Y(k)$ per ogni $h, k \in \mathbb{R}$, mentre per mostrare che sono dipendenti è sufficiente verificare che *esistono* due valori $h, k \in \mathbb{R}$ tali che $p(h, k) \neq p_X(h)p_Y(k)$.

3. Trasformazioni di variabili aleatorie

Come sempre accade in matematica, anche con le variabili aleatorie saremo interessati alla costruzione di una o più nuove variabili aleatorie partendo da variabili aleatorie note. Il primo esempio di tale costruzione che vogliamo studiare riguarda il massimo e il minimo.

Intanto ricordiamo che

DEFINIZIONE. Data una qualunque variabile aleatoria X (non necessariamente discreta) la sua funzione di ripartizione è data da

$$\begin{aligned} F_X : \mathbb{R} &\rightarrow \mathbb{R} \\ t &\mapsto P(X \leq t). \end{aligned}$$

ESEMPIO 7.5. Lanciamo simultaneamente una moneta e un dado fino ad ottenere almeno una volta testa e almeno una volta 6. Qual è la probabilità che occorranza k lanci?

Si sta studiando una variabile di tipo $Z = \max(S, T)$, dove S e T sono variabili geometriche modificate indipendenti di parametri rispettivamente $p = \frac{1}{2}$ e $q = \frac{1}{6}$, cioè $S \sim \tilde{G}(\frac{1}{2})$ e $T \sim \tilde{G}(\frac{1}{6})$. Usiamo il seguente risultato

LEMMA 7.6. *Siano S, T due variabili aleatorie indipendenti e $X = \max(S, T)$. Allora*

$$F_Z(t) = F_S(t)F_T(t).$$

DIMOSTRAZIONE. La funzione di ripartizione di Z è data da

$$F_Z(t) = P(Z \leq t) = P(\max(S, T) \leq t) = P(S \leq t, T \leq t) = P(S \leq t)P(T \leq t) = F_S(t)F_T(t),$$

dove nella terza uguaglianza abbiamo sfruttato il fatto che richiedere che il massimo di due numeri sia $\leq t$ è come chiedere che siano entrambi $\leq t$, e nella quarta uguaglianza si è sfruttato che S e T sono indipendenti. \square

Ricordiamo ora che la funzione di ripartizione di una variabile $X \sim \tilde{G}(p)$ è data da

$$F_X(k) = 1 - (1 - p)^k$$

e di conseguenza, nell'esempio che stiamo studiando

$$F_Z(k) = (1 - (1 - p)^k)(1 - (1 - q)^k)$$

da cui

$$\begin{aligned} P(Z = k) &= F_Z(k) - F_Z(k-1) \\ &= (1 - (1-p)^k)(1 - (1-q)^k) - (1 - (1-p)^{k-1})(1 - (1-q)^{k-1}) \\ &= p(1-p)^{k-1} + q(1-q)^{k-1} + (pq - p - q)(1-p)^{k-1}(1-q)^{k-1}. \end{aligned}$$

Sostituendo a questo punto $p = 1/2$ e $q = 1/6$ otteniamo

$$P(Z = k) = \frac{1}{2^k} + \frac{5^{k-1}}{6^k} - \frac{7}{12} \frac{5^{k-1}}{12^{k-1}}.$$

Se invece si è interessati al minimo (nell'esempio lancio la moneta e il dado finché otteniamo un 6 o una testa), si considera la variabile $W = \min(S, T)$ e sarà più conveniente calcolare $1 - F_W(k) = P(W > k)$. Abbiamo infatti in questo caso

$$1 - F_W(k) = P(W > k) = P(\min(S, T) > k) = P(S > k, T > k) = P(S > k)P(T > k) = (1 - F_S(k))(1 - F_T(k)),$$

dove abbiamo sfruttato che richiedere che il minimo di due numeri sia $> k$ è come richiedere che siano entrambi $> k$. Il calcolo fatto finora vale sempre, purché le variabili S e T siano indipendenti. Nel nostro caso utilizziamo adesso il fatto che S e T sono geometriche modificate. Otteniamo

$$1 - F_W(k) = (1-p)^k(1-q)^k = (1-p-q+pq)^k.$$

Ne deduciamo che la funzione di ripartizione di W è data da $F_W(k) = 1 - (1-p-q+pq)^k$ e quindi concludiamo che W è ancora una variabile geometrica modificata di parametro $p+q-pq$, $W \sim \tilde{G}(p+q-pq)$. Nel caso particolare abbiamo quindi $W \sim \tilde{G}(7/12)$.

Vediamo ora più in generale come possiamo ottenere la densità di una variabile aleatoria ottenuta trasformando altre variabili di cui conosciamo la densità congiunta.

Facciamo anche qui un esempio. Consideriamo una variabile X di densità

$$p_X(h) = \begin{cases} \frac{1}{4} & h = 1 \\ \frac{1}{3} & h = -1 \\ \frac{1}{4} & h = 2 \\ \frac{1}{6} & h = -3 \\ 0 & \text{altrimenti.} \end{cases}$$

ed una nuova variabile data da $\Phi(X) = X^2$. Che valori assume la variabile $Y = \Phi(X)$? Qual è la sua densità? La variabile Y assume i valori 1, 4, 9 e la densità è data da $p_Y(1) = \frac{1}{4} + \frac{1}{3} = \frac{7}{12}$, $p_Y(4) = \frac{1}{4}$ e $p_Y(9) = \frac{1}{6}$.

In generale, se $\Phi : \mathbb{R}^m \rightarrow \mathbb{R}$, allora $\Phi(\underline{X})$ è una variabile aleatoria discreta. Inoltre, dato $k \in \mathbb{R}$ si ha

$$\{\Phi(\underline{X}) = k\} = \bigcup_{\underline{x}: \Phi(\underline{x})=k} \{\underline{X} = \underline{x}\},$$

e quindi questo è un evento perché unione numerabile di eventi. Ne segue che $\Phi(\underline{X})$ è una variabile aleatoria. Abbiamo inoltre la seguente interpretazione per la densità della variabile $\Phi(\underline{X})$:

$$p_{\Phi(\underline{X})}(k) = \sum_{\underline{h}: \Phi(\underline{h})=k} p_{\underline{X}}(\underline{h}).$$

Se trasformiamo due variabili indipendenti queste rimangono, come intuibile, indipendenti. Vediamo il caso di dimensione 1 (non fatto a lezione). Siano X e Y due variabili aleatorie indipendenti e $\Phi, \Psi : \mathbb{R} \rightarrow \mathbb{R}$.

$$\begin{aligned} P(\Phi(X) = k_1, \Psi(Y) = k_2) &= \sum_{\substack{h_1: \Phi(h_1)=k_1 \\ h_2: \Psi(h_2)=k_2}} p_X(h_1)p_Y(h_2) \\ &= \sum_{h_1: \Phi(h_1)=k_1} p_X(h_1) \sum_{h_2: \Psi(h_2)=k_2} p_Y(h_2) \\ &= P(\Phi(X) = k_1)P(\Psi(Y) = k_2). \end{aligned}$$

Vediamo ora alcuni esempi su come possano essere “manipolate” le densità di variabili multidimensionali per risolvere problemi concreti.

ESEMPIO 7.7. Si lanciano simultaneamente una moneta e un dado ripetutamente. Qual è la probabilità che venga testa prima che venga 6? Detto S il primo lancio che dà testa e T il primo lancio che dà 6 vogliamo determinare

$$P(S < T) = \sum_{h < k} p(h, k),$$

dove p è la densità congiunta delle variabili S e T . Siccome le variabili $S \sim \tilde{G}(\frac{1}{2})$ e $T \sim \tilde{G}(\frac{1}{6})$ sono indipendenti abbiamo

$$p(h, k) = \frac{1}{2} \left(\frac{1}{2}\right)^{h-1} \frac{1}{6} \left(\frac{5}{6}\right)^{k-1},$$

se h ed k sono entrambi interi positivi, e $p(h, k) = 0$ altrimenti. Abbiamo quindi

$$\begin{aligned} P(S < T) &= \sum_{h=1}^{\infty} \sum_{k=h+1}^{\infty} p(h, k) \\ &= \frac{1}{5} \sum_{h=1}^{\infty} \left(\frac{1}{2}\right)^h \sum_{k=h+1}^{\infty} \left(\frac{5}{6}\right)^k \\ &= \sum_{h=1}^{\infty} \left(\frac{5}{12}\right)^h \\ &= \frac{5}{12} \frac{1}{1 - \frac{5}{12}} \\ &= \frac{5}{7}, \end{aligned}$$

dove abbiamo sfruttato ancora la somma della serie geometrica, cioè che per ogni $0 < a < 1$, $n \in \mathbb{N}$, si ha

$$\sum_{k=n}^{\infty} a^k = \frac{a^n}{1-a}.$$

In questo esempio abbiamo implicitamente usato la trasformazione di variabili $\Phi(S, T) = S - T$ e studiato l'evento $\{\Phi(S, T) > 0\}$.

Abbiamo visto che

$$P(\Phi(X) = k) = \sum_{\underline{h}: \Phi(\underline{h})=k} p(\underline{h}).$$

In particolare, se X ed Y sono due variabili che hanno la stessa densità, anche $\Phi(X)$ e $\Phi(Y)$ hanno la stessa densità. E ribadiamo ancora una volta: quello che importa di una variabile è la sua densità. E quindi se due variabili, anche se definite su spazi completamente diversi, hanno la stessa densità, vanno considerate “come se fossero la stessa cosa”.

Se ad esempio consideriamo n variabili U_1, \dots, U_n indipendenti, tutte di tipo $B(1, p)$, cosa possiamo dire della variabile somma $U_1 + \dots + U_n$? La variabile n -dimensionale (U_1, \dots, U_n) ha la stessa densità di (X_1, \dots, X_n) in uno schema successo-insuccesso con rimpiazzo. Si ha quindi necessariamente che $U_1 + \dots + U_n$ ha densità binomiale $B(n, p)$, perché tale è la densità di $X_1 + \dots + X_n$. Più in generale se ho due variabili indipendenti di densità $B(n_1, p)$ e $B(n_2, p)$ si ha che la loro somma è anche somma di $n_1 + n_2$ variabili indipendenti, tutte di densità $B(1, p)$ e quindi è una variabile di densità $B(n_1 + n_2, p)$.

ESEMPIO 7.8. Luca lancia un dado 5 volte e vince ogni volta che esce almeno 5. Poi estrae con rimpiazzo per 10 volte una pallina da un'urna che ne ha una bianca e 2 rosse e vince ogni volta che estrae la rossa. Qual è la probabilità che vinca meno di 3 volte?

La trasformazione più naturale da effettuare quando si hanno a disposizione due variabili aleatorie è la loro somma. Vediamo cosa accade se queste due variabili sono di Poisson.

Consideriamo due variabili $X \sim P(\lambda)$ e $Y \sim P(\mu)$ di Poisson di parametri rispettivamente λ e μ , indipendenti. Queste le possiamo vedere come approssimate da due variabili binomiali $B(n, \lambda/n)$ e $B(m, \mu/m)$, con n ed m abbastanza grandi. Inoltre possiamo sceglierli in modo che $\frac{m}{n}$ sia circa uguale a $\frac{\mu}{\lambda}$. Ne segue che $p = \lambda/n = \mu/m$. E quindi $X + Y$ è approssimata da $B(n + m, p)$. E a sua volta questa è approssimata da una variabile di Poisson di parametro

$$(n + m)p = np + mp = n\frac{\lambda}{n} + m\frac{\mu}{m} = \lambda + \mu.$$

Ne segue che $X + Y$ è approssimata da una variabile di Poisson di parametro $\lambda + \mu$. In realtà questo calcolo può essere reso più preciso e dimostrare che $X + Y \sim P(\lambda + \mu)$, cioè è esattamente una variabile di Poisson di parametro $\lambda + \mu$.

ESERCIZIO 7.9. Le quantità di telefonate ricevute ogni minuto da due call-center hanno densità di Poisson di densità $\lambda = 2$ e $\mu = 3$. Qual è la probabilità che complessivamente ricevano meno di 4 telefonate in un minuto? Stiamo considerando la somma di due variabili di Poisson che possiamo ritenere indipendenti e che avrà quindi densità di Poisson di parametro 5. Ne segue

$$P(X + Y < 4) = e^{-5} \left(\frac{5^0}{0!} + \frac{5^1}{1!} + \frac{5^2}{2!} + \frac{5^3}{3!} \right) = 26.5\%.$$

CHAPTER 8

Il valore atteso e la varianza

In questo capitolo introduciamo e studiamo due quantità, il valore atteso e la varianza, che generalizzano in ambito aleatorio la media e la varianza che abbiamo studiato in statistica descrittiva.

1. Il valore atteso

Iniziamo con un esempio e consideriamo una variabile $X \sim B(2, 1/3)$. Scriviamo la sua densità e proviamo a definire una media. Viene naturale definirla come media pesata dei valori che assume, scegliendo come pesi le rispettive probabilità, cioè come

$$(2/3)^2 \cdot 0 + (4/9) \cdot 1 + (1/9) \cdot 2 = 2/3.$$

In generale poniamo

DEFINIZIONE. Sia X una variabile aleatoria discreta di densità $p(h)$. La media, o valore atteso, o speranza matematica di X è data da

$$E[X] = \sum_{h \in \mathbb{R}} h p(h).$$

Osserviamo che ricorda la media pesata di un carattere. Per definire la media di una variabile aleatoria discreta è in realtà necessaria una condizione di assoluta convergenza che spesso trascureremo. Questa condizione serve a garantire che la media sia effettivamente ben definita e che non dipenda dall'ordine con cui consideriamo i valori che la variabile assume. Nel caso in cui la variabile sia finita chiaramente tali problemi non sussistono. Anche nel caso in cui la variabile assume solo valori nonnegativi, e quindi essenzialmente in tutti i tipi di variabili visti in questo corso, tale problema non sussiste.

ESEMPIO 8.1. Se X è una variabile di tipo $B(1, p)$ si ha

$$E[X] = 0 \cdot (1 - p) + 1 \cdot p = p.$$

ESEMPIO 8.2. Calcoliamo esplicitamente anche la media di una variabile $B(2, p)$. Si ha

$$E[X] = 0 \cdot (1 - p)^2 + 1 \cdot 2p(1 - p) + 2 \cdot p^2 = 2p.$$

Guardando questi due esempi verrebbe da pensare che la media di una variabile $B(n, p)$ sia np . Il calcolo diretto è però impegnativo. Vediamo allora qualche risultato di carattere generale che ci servirà anche per calcolare la media di una $B(n, p)$. Il primo obiettivo sarà quello di capire come calcolare la media di una variabile aleatoria ottenuta trasformando variabile aleatorie note,

ESEMPIO 8.3. Si consideri una variabile X di densità

$$p_X(h) = \begin{cases} \frac{1}{4} & \text{se } h = 1, -1, 2 \\ \frac{1}{8} & \text{se } h = -2, 3 \\ 0 & \text{altrimenti.} \end{cases}$$

Vogliamo determinare la media di X e di X^2 . La media di X è data da

$$E[X] = \frac{1}{4}(1 - 1 + 2) + \frac{1}{8}(-2 + 3) = \frac{5}{8}.$$

Per determinare $E[X^2]$ calcoliamo intanto la densità

$$p_{X^2}(h) = \begin{cases} \frac{1}{2} & \text{se } h = 1 \\ \frac{3}{8} & \text{se } h = 4 \\ \frac{1}{8} & \text{se } h = 9 \\ 0 & \text{altrimenti.} \end{cases}$$

dove abbiamo utilizzato i risultati sulla densità di una variabile trasformata, da cui possiamo ottenere

$$E[X^2] = \sum_{h \in \mathbb{R}} h \cdot p_{X^2}(h) = 1 \cdot \frac{1}{2} + 4 \cdot \frac{3}{8} + 9 \cdot \frac{1}{8} = \frac{25}{8}.$$

Potremmo essere tentati di calcolare $E[X^2]$ anche nel seguente modo

$$\sum_{h \in \mathbb{R}} h^2 p_X(h) = 1^2 \frac{1}{4} + (-1)^2 \frac{1}{4} + 2^2 \frac{1}{4} + (-2)^2 \frac{1}{8} + 3^2 \frac{1}{8} = \frac{25}{8} :$$

avremmo ottenuto lo stesso risultato. Il prossimo teorema ci assicura che possiamo utilizzare questa ultima formula anche in un contesto molto più generale.

TEOREMA 8.4. Sia $\underline{X} = (X_1, \dots, X_m)$ una variabile aleatoria m -dimensionale discreta e $\Phi := \mathbb{R}^m \rightarrow \mathbb{R}$ una qualunque trasformazione. Poniamo $Z = \Phi(\underline{X})$. Allora

$$E[Z] = \sum_{\underline{h} \in \mathbb{R}^m} \Phi(\underline{h}) p(\underline{h}),$$

dove $p(\underline{h})$ indica la densità di \underline{X} .

DIMOSTRAZIONE. Ricordando la formula per la densità di una variabile trasformata abbiamo

$$\begin{aligned} E[Z] &= \sum_{k \in \mathbb{R}} k P(Z = k) \\ &= \sum_{k \in \mathbb{R}} k \sum_{\underline{h}: \Phi(\underline{h})=k} p(\underline{h}) \\ &= \sum_{\underline{h} \in \mathbb{R}^m} \Phi(\underline{h}) p(\underline{h}). \end{aligned}$$

□

PROPOSIZIONE 8.5. La funzione E è lineare, cioè $E[X + Y] = E[X] + E[Y]$, e $E[cX] = cE[X]$, dove X e Y sono variabili aleatorie e $c \in \mathbb{R}$.

DIMOSTRAZIONE. Sfruttiamo 2 volte il teorema precedente e otteniamo

$$\begin{aligned}
 E[cX] &= \sum_{x \in \mathbb{R}} (cx)p(x) = cE[X]; \\
 E[X + Y] &= \sum_{(h,k) \in \mathbb{R}^2} (h+k)p(h,k) \\
 &= \sum_{(h,k) \in \mathbb{R}^2} hp(h,k) + \sum_{(h,k) \in \mathbb{R}^2} kp(h,k) \\
 &= \sum_{h \in \mathbb{R}} h \sum_{k \in \mathbb{R}} p(h,k) + \sum_{k \in \mathbb{R}} k \sum_{h \in \mathbb{R}} p(h,k) \\
 &= \sum_{h \in \mathbb{R}} hp_X(h) + \sum_{k \in \mathbb{R}} kp_Y(k),
 \end{aligned}$$

dove abbiamo utilizzato la formula che esprime le densità marginali in funzione della densità congiunta. \square

ESEMPIO 8.6. Si hanno a disposizione 6 vaschette di gelato in 6 gusti. I primi 3 gusti, cioccolato, fragola e pistacchio, sono scelti mediamente da 2 persone su 9 e gli altri gusti da 1 persona su 9. Offriamo un gelato ciascuno ai nostri 10 amici. Si consideri la seguente variabile aleatoria

$$X_i = \begin{cases} 1 & \text{se l}'i\text{-esimo gusto è scelto da qualcuno} \\ 0 & \text{altrimenti.} \end{cases}$$

(a) Qual'è la densità delle X_i ?

Detta $p_i(x)$ la densità della variabile X_i si ha, se $i = 1, 2, 3$

$$p_i(x) = \begin{cases} 0.92 & \text{se } x = 1 \\ 0.08 & \text{se } x = 0 \\ 0 & \text{altrimenti} \end{cases},$$

mentre se $i = 4, 5, 6$

$$p_i(x) = \begin{cases} 0.69 & \text{se } x = 1 \\ 0.31 & \text{se } x = 0 \\ 0 & \text{altrimenti} \end{cases}.$$

(b) Le variabili X_i sono indipendenti?

Le variabili X_i sono dipendenti in quanto, ad esempio, la probabilità

$$P(X_1 = X_2 = X_3 = X_4 = X_5 = X_6 = 0) = 0$$

che nessun gusto venga scelto è chiaramente nulla, mentre le probabilità

$$P(X_i = 0)$$

che il gusto i non sia scelto è chiaramente diversa da zero.

(c) Quante vaschette mi devo aspettare di utilizzare (mediamente)?

Il numero di vaschette utilizzate è dato dalla variabile $X = X_1 + X_2 + \dots + X_6$, la cui densità è molto difficile da calcolare. Tuttavia, la sua media si calcola facilmente grazie alla linearità. Si ha

$$E[X] = E[X_1] + E[X_2] + E[X_3] + E[X_4] + E[X_5] + E[X_6] = 3 \cdot 0.92 + 3 \cdot 0.69 = 2.76 + 2.07 = 4.83.$$

Diciamo quindi che mediamente almeno una vaschetta mi rimane chiusa.

La Proposizione 8.5 come abbiamo appena visto in questo esempio è molto utile negli esercizi, ma ci permette anche di calcolare facilmente la media di variabili binomiali, ipergeometriche e di Poisson.

Consideriamo quindi una variabile binomiale $X \sim B(n, p)$. Siccome tale variabile è somma di n variabili $B(1, p)$, la sua media sarà la somma delle medie di queste ultime. Avremo quindi $E[X] = np$.

In modo del tutto analogo, se X è una variabile ipergeometrica $H(n, b, r)$ sappiamo che X è somma di n variabili di densità $B(1, \frac{b}{b+r})$. La sua media sarà dunque $n \frac{b}{b+r}$.

Proviamo ora a calcolare la media di una variabile con densità di Poisson $P(\lambda)$. Sappiamo che tale variabile è approssimata da una $B(n, \lambda/n)$ che ha media λ . Ne segue (in modo non rigoroso) che se $X \sim P(\lambda)$ allora $E[X] = \lambda$. In alternativa si può effettuare il calcolo esplicito

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

Per calcolare la media di una variabile geometrica modificata abbiamo bisogno di qualche strumento in più. Lanciando un dado ci potremmo aspettare che mediamente dobbiamo effettuare sei lanci per ottenere un 6. Questo è confermato nel seguente risultato.

PROPOSIZIONE 8.7. *Se $X \sim \tilde{G}(p)$ allora $E[X] = \frac{1}{p}$.*

DIMOSTRAZIONE. Abbiamo bisogno della seguente osservazione. Sia

$$f(x) = \sum_{k=1}^{\infty} x^k.$$

Dallo studio fatto sulle serie geometriche sappiamo che $f(x) = \frac{x}{1-x}$ (se $0 < x < 1$). La sua derivata la possiamo calcolare in due modi diversi, considerando le due espressioni di $f(x)$ (e ricordando certe proprietà delle serie geometriche...). Otteniamo

$$f'(x) = \sum_{k=1}^{\infty} k x^{k-1} = \frac{1}{(1-x)^2}.$$

Calcoliamo ora la media della variabile X . Otteniamo

$$\begin{aligned} E[X] &= \sum_{k=1}^{\infty} k p (1-p)^{k-1} \\ &= p f'(1-p) \\ &= p \frac{1}{(1-(1-p))^2} \\ &= \frac{1}{p}. \end{aligned}$$

□

Abbiamo visto che la media di una somma è la somma delle medie. Cosa si può dire del prodotto di due variabili? Vediamo due esempi.

Nello schema successo-insuccesso con rimpiazzo consideriamo le variabili X_1 e X_2 , entrambe di densità $B(1, p)$. La variabile prodotto $X_1 X_2$ può assumere ancora solo i valori 0 e 1 e quindi è una variabile $B(1, p')$ dove $p' = P(X_1 X_2 = 1)$. Si ha

$$p' = P(X_1 X_2 = 1) = P(X_1 = 1, X_2 = 1) = P(X_1 = 1)P(X_2 = 1) = p^2.$$

Ne segue che $E[X_1 X_2] = p^2 = E[X_1]E[X_2]$. Vediamo cosa capita nello schema successo-insuccesso senza rimpiazzo. Le variabili X_1 e X_2 hanno densità $B(1, \frac{b}{b+r})$. Il loro prodotto è ancora una variabile di Bernoulli $B(1, q)$. Determiniamo q :

$$q = P(X_1 X_2 = 1) = P(X_1 = 1, X_2 = 1) = \frac{b}{b+r} \frac{b-1}{b+r-1}.$$

Ne segue che $E[X_1 X_2] = \frac{b}{b+r} \frac{b-1}{b+r-1} \neq (\frac{b}{b+r})^2 = E[X_1]E[X_2]$. Queste osservazioni sono coerenti con la seguente.

PROPOSIZIONE 8.8. *Se X ed Y sono variabili aleatorie indipendenti, allora*

$$E[XY] = E[X]E[Y].$$

DIMOSTRAZIONE. Sfruttiamo il Teorema 8.4 per calcolare $E[XY]$:

$$\begin{aligned} E[XY] &= \sum_{(h,k) \in \mathbb{R}^2} hk p(h, k) \\ &= \sum_{(h,k) \in \mathbb{R}^2} hk p_X(h) p_Y(k) \\ &= \sum_{h \in \mathbb{R}} h p_X(h) \sum_{k \in \mathbb{R}} k p_Y(k) \\ &= E[X]E[Y]. \end{aligned}$$

□

Sfortunatamente il viceversa non è vero. Può ben accadere che due variabili X ed Y siano dipendenti e che $E[XY] = E[X]E[Y]$, come mostrato nel prossimo esempio.

ESEMPIO 8.9. Si consideri la variabile bidimensionale (X, Y) di densità

$$p(h, k) = \begin{cases} 1/4 & \text{se } (h, k) = (1, 2) \\ 1/4 & \text{se } (h, k) = (-1, 2) \\ 1/2 & \text{se } (h, k) = (0, 0) \\ 0 & \text{altrimenti} \end{cases}$$

In questo caso si ha $E[X] = E[XY] = 0$ e $E[Y] = 1$ ed in particolare $E[X]E[Y] = E[XY]$. Tuttavia X e Y sono dipendenti in quanto, ad esempio, $p_X(1) = 1/4$, $p_Y(2) = 1/2$ e $p(1, 2) = 1/4$.

ESERCIZIO 8.10. Si lancia un dado ripetutamente e sia X la variabile aleatoria data da $X = i$ se otteniamo 6 per la seconda volta al lancio i . Stabilire la media di X . Sia X_1 il numero di lanci necessari ad ottenere 6 la prima volta e X_2 il numero di lanci necessari ad ottenere 6 per la seconda

volta, dopo aver ottenuto 6 per la prima volta. Abbiamo quindi $X = X_1 + X_2$ e che X_1 e X_2 sono entrambe $\tilde{G}(1/6)$. Concludiamo che

$$E[X] = E[X_1] + E[X_2] = 6 + 6 = 12.$$

2. Varianza

Se X è una variabile aleatoria discreta definiamo la varianza di X ricordando la definizione di varianza di un carattere in statistica descrittiva

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

abbiamo anche in questo caso la "media del quadrato della differenza rispetto alla media". Scriveremo talvolta σ_X^2 al posto di $\text{Var}(X)$. Valgono le stesse osservazioni fatte nel capitolo sulla statistica descrittiva riguardo la varianza come misura di dispersione. Vedremo più in là un risultato, la disuguaglianza di Chebyshev, che ci dà la probabilità che X si discosti per una certa quantità dalla sua media: tale probabilità è proporzionale alla sua varianza.

Come in statistica descrittiva vale la seguente formula per il calcolo

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

La dimostrazione è sostanzialmente la stessa: basta sfruttare la linearità della media

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

Possiamo quindi facilmente calcolare la varianza di una variabile di Bernoulli $X \sim B(1, p)$ (osservando che in tal caso $X^2 = X$)

$$\text{Var}(X) = p - p^2 = p(1 - p).$$

Come possiamo fare per calcolare la varianza di una variabile $B(n, p)$? È ancora vero che la varianza è lineare? La risposta è no, ma qualche utile proprietà la possiamo ancora dimostrare. Ad esempio si verifica facilmente che $\text{Var}(aX) = a^2 \text{Var}(X)$. Molto più importante è il seguente risultato.

PROPOSIZIONE 8.11. *Date due variabili aleatorie si ha*

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y),$$

dove $\text{Cov}(X, Y) = E[XY] - E[X]E[Y]$. In particolare, se X ed Y sono indipendenti abbiamo $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$.

DIMOSTRAZIONE.

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

□

Possiamo quindi calcolare la varianza di una variabile binomiale $X \sim B(n, p)$, in quanto essa è somma di n variabili $B(1, p)$ indipendenti:

$$\text{Var}(X) = np(1 - p).$$

Similmente, considerando che una variabile di Poisson $X \sim P(\lambda)$ di parametro λ è limite di una successione di variabili $B(n, \lambda/n)$ aventi varianza $n \frac{\lambda}{n} (1 - \frac{\lambda}{n})$ verifichiamo facilmente che $\text{Var}(X) = \lambda$. Deduciamo pertanto che il parametro λ in una variabile di Poisson rappresenta sia la media che la varianza della variabile stessa.

ESEMPIO 8.12. In uno schema successo-insuccesso senza rimpiazzo si considerino le variabili di Bernoulli X_1 e X_2 relative alle prime 2 estrazioni. Determinare $\text{Cov}(X_1, X_2)$.

Sappiamo che X_1 e X_2 sono variabili di Bernoulli $B(1, \frac{b}{b+r})$. Per determinare la covarianza abbiamo bisogno di calcolare la media di $X_1 X_2$. Tale variabile è ancora di Bernoulli di parametro $\frac{b}{b+r} \frac{b-1}{b+r-1}$ e questa è quindi anche la sua media. Di conseguenza

$$\text{Cov}(X_1, X_2) = E[X_1 X_2] - E[X_1]E[X_2] = \frac{b}{b+r} \frac{b-1}{b+r-1} - \frac{b}{b+r} \frac{b}{b+r} = \frac{-br}{(b+r)^2(b+r-1)}.$$

Sappiamo già che se la covarianza di X ed Y è nulla non è necessariamente vero che le variabili X ed Y siano indipendenti (si veda l'esempio 8.9). E per esercizio potrebbe essere utile calcolare a questo punto la varianza di $X + Y$ dell'esempio 8.9. Come per i caratteri definiamo anche qui il coefficiente di correlazione come

$$\rho_{X,Y} = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y},$$

dove $\sigma_X = \sqrt{\text{Var}(X)}$ è lo scarto quadratico medio di X .

ESEMPIO 8.13. In uno schema successo-insuccesso senza rimpiazzo si considerino le variabili di Bernoulli X_1 e X_2 relative alle prime 2 estrazioni. Determinare il coefficiente di correlazione.

Sappiamo che X_1 e X_2 sono variabili di Bernoulli $B(1, \frac{b}{b+r})$ e quindi $\sigma_{X_1} = \sigma_{X_2} = \sqrt{\frac{br}{(b+r)^2}}$. Dall'esempio 8.12 sappiamo anche che

$$\text{Cov}(X_1, X_2) = \frac{-br}{(b+r)^2(b+r-1)}$$

e quindi

$$\rho_{X_1, X_2} = \frac{\text{Cov}(X_1, X_2)}{\sigma_{X_1} \sigma_{X_2}} = \frac{\frac{-br}{(b+r)^2(b+r-1)}}{\frac{br}{(b+r)^2}} = -\frac{1}{b+r-1}.$$

Osserviamo che per grandi valori di $b+r$ tale coefficiente tende a 0. Questo si giustifica anche pensando al fatto che se il numero di palline è elevato le variabili X_1 e X_2 sono "quasi" indipendenti.

3. La disuguaglianza di Chebyshev e la legge dei grandi numeri

Come abbiamo anticipato la disuguaglianza di Chebyshev esprime in modo quantitativo come la varianza sia un indice di dispersione nelle variabili aleatorie. Per noi sarà anche uno strumento fondamentale per dimostrare la legge dei grandi numeri.

PROPOSIZIONE 8.14 (Disuguaglianza di Chebyshev). *Sia X una variabile aleatoria discreta. Allora preso comunque $\varepsilon > 0$ si ha*

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

DIMOSTRAZIONE. Si consideri l'evento

$$A = \{|X - E[X]| > \varepsilon\}.$$

Ricordiamo la funzione caratteristica di A , cioè

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

e che $\chi_A \sim B(1, P(A))$. In particolare $E[\chi_A] = P(A)$.

Confrontiamo ora le due variabili aleatorie $Y = \varepsilon^2 \chi_A$, e $Z = (X - E[X])^2$. Osserviamo che $Z(\omega) \geq Y(\omega)$ per ogni $\omega \in \Omega$. Infatti, se $\omega \in A$ abbiamo $Y(\omega) = \varepsilon^2$, mentre $Z(\omega) > \varepsilon^2$ per definizione di A . Se $\omega \notin A$ abbiamo $Y(\omega) = 0$, mentre $Z(\omega) \geq 0$ perché Z , essendo un quadrato, non è mai negativa. Deduciamo che $E[Z] \geq E[Y]$. Ma $E[Z] = \text{Var}(X)$, mentre $E[Y] = \varepsilon^2 P(A) = \varepsilon^2 P(|X - E[X]| > \varepsilon)$. Il risultato segue. \square

ESEMPIO 8.15. Stimare, utilizzando la disuguaglianza di Chebyshev, la probabilità che una variabile $X \sim B(10, 1/2)$ soddisfi $|X - 5| \geq 3$. Abbiamo

$$P(|X - 5| \geq 3) = P(|X - 5| > 2.99) \leq \frac{10/4}{2.99^2} = 0.28.$$

Cerchiamo ora di dare un senso preciso ad un'espressione, la legge dei grandi numeri, oramai molto diffusa, talvolta a sproposito, anche nel linguaggio comune.

Se lanciamo una moneta n volte e otteniamo k volte testa è difficile ottenere $\frac{k}{n} = \frac{1}{2}$. Però l'intuizione ci suggerisce che se n è molto grande (e la moneta non è truccata) il rapporto k/n non debba discostarsi troppo da $1/2$. Abbiamo a che fare in questo caso con una sequenza di variabili X_1, X_2, \dots tutte di densità $B(1, 1/2)$ e stiamo valutando la variabile

$$\bar{X}_n = \frac{1}{n}(X_1 + \dots + X_n).$$

La variabile \bar{X}_n viene talvolta detta media campionaria, ma non è da confondere con la media campionaria definita in statistica descrittiva. Il rapporto k/n di cui sopra non è altri che il valore di \bar{X}_n e la legge dei grandi numeri permette di dare un senso preciso alla nostra intuizione, cioè che il valore di \bar{X}_n debba essere vicino al valore $1/2$.

Abbiamo per questo bisogno di un concetto di convergenza per le variabili aleatorie.

DEFINIZIONE. Diciamo che una successione di variabili aleatorie X_1, X_2, \dots converge in probabilità alla costante a se per ogni ε fissato si ha

$$\lim_{n \rightarrow \infty} P(|X_n - a| > \varepsilon) = 0.$$

TEOREMA 8.16 (Legge dei grandi numeri). *Sia X_n , $n \in \mathbb{N}$ una successione di variabili aleatorie discrete indipendenti ed aventi tutte la stessa densità (con media μ e varianza σ^2). Allora posto*

$$\bar{X}_n = \frac{1}{n}(X_1 + X_2 + \dots + X_n),$$

la successione \bar{X}_n converge in probabilità alla variabile aleatoria costante μ .

DIMOSTRAZIONE. Osserviamo che $E[\bar{X}_n] = \mu$ e $\text{Var}(\bar{X}_n) = \frac{\sigma^2}{n}$. Applichiamo la disuguaglianza di Chebyshev alla variabile \bar{X}_n , ottenendo

$$P(|\bar{X}_n - \mu| > \varepsilon) \leq \frac{\text{Var}(\bar{X}_n)}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2} \rightarrow 0.$$

□

Un risultato analogo vale anche per la varianza (dimostrazione non fatta in aula)

PROPOSIZIONE 8.17 (Legge dei grandi numeri per la varianza). *Stesse ipotesi del Teorema 8.16. Poniamo*

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

Allora la successione $\bar{\sigma}_n^2$ converge in probabilità alla variabile costante σ^2 .

DIMOSTRAZIONE. (non fatta) Per semplicità di notazione poniamo $Y_i = X_i^2$ per ogni i e definiamo \bar{Y}_n similmente a prima. Come per la varianza di un carattere o di una variabile vale la formula

$$\bar{\sigma}_n^2 = \bar{Y}_n - \bar{X}_n^2,$$

cioè la varianza campionaria è la media dei quadrati meno il quadrato della media. Applicando il Teorema 8.16 abbiamo che \bar{Y}_n converge in probabilità a $E[Y_1] = E[X_1^2]$, mentre \bar{X}_n^2 converge in probabilità a $E[X_1]^2$. Ne segue che $\bar{\sigma}_n^2$ converge in probabilità a $E[X_1^2] - E[X_1]^2 = \sigma^2$. Abbiamo usato in questa proposizione alcune proprietà della convergenze in probabilità che non abbiamo dimostrato. □

Considerando il problema di un'indagine statistica in cui X_i rappresenta il valore assunto da un carattere sull' i -esimo elemento di un campione di n elementi, allora \bar{X}_n non è altri che la media del campione (e per questo si è soliti chiamarla *media campionaria* anche se in realtà nel caso generale non abbiamo né un'indagine statistica né un campione). Similmente $\bar{\sigma}_n^2$ si dice *varianza campionaria*.

CHAPTER 9

Variabili aleatorie continue

In questo capitolo introduciamo le basi delle variabili aleatorie continue. I risultati che otterremo saranno decisamente più limitati perchè gli strumenti matematici che servirebbero in questo contesto sono più avanzati.

1. Funzione di ripartizione e densità di variabili continue

Iniziamo dando la definizione

DEFINIZIONE. Una variabile aleatoria $X : \Omega \rightarrow \mathbb{R}$ si dice *continua* in un intervallo I se la funzione di ripartizione

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

è continua per ogni $t \in I$. Una variabile aleatoria X si dice continua se è continua su tutto \mathbb{R} .

Osserviamo e ricordiamo che se F è la funzione di ripartizione di X e $a < b$ allora abbiamo sempre

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

se $a \leq b$. Usiamo questa osservazione per mostrare che mostrare il seguente (sorprendente) risultato.

LEMMA 9.1. *Se una variabile aleatoria X è continua in un intervallo aperto I , allora*

$$P(X = t) = 0$$

per ogni $t \in I$.

DIMOSTRAZIONE. Si ha

$$0 \leq P(X = t) \leq P(t - \frac{1}{n} < X \leq t) = F(t) - F(t - \frac{1}{n}),$$

e quest'ultima quantità tende a 0 per la continuità di F . □

Ne segue che se X è continua allora gli eventi $\{X < t\}$ e $\{X \leq t\}$ hanno la stessa probabilità.

La densità nel senso delle variabili aleatorie discrete è quindi identicamente nulla per il Lemma 9.1.

Nel contesto delle variabili continue utilizziamo quindi un nuovo concetto di densità, che indichiamo solitamente con la lettera f anziché la lettera p , e che chiameremo densità continua, anziché densità discreta.

DEFINIZIONE. Sia X una variabile aleatoria continua su \mathbb{R} . Una funzione $f : \mathbb{R} \rightarrow \mathbb{R}$ integrabile si dice *densità continua* per X se

$$P(a \leq X \leq b) = \int_a^b f(s) ds.$$

La stessa formula continua a valere se si sostituisce a con $-\infty$ o b con $+\infty$ (o entrambi). La densità di una variabile aleatoria continua non è univocamente determinata: cambiare il valore di f in un numero finito di punti non altera il valore di un suo integrale. Possiamo però sempre assumere che, se $f : \mathbb{R} \rightarrow \mathbb{R}$ è la densità di una variabile aleatoria X , allora

$$(1) \quad f(s) \geq 0 \text{ per ogni } s \in \mathbb{R}$$

oltre alla evidente proprietà

$$(2) \quad \int_{-\infty}^{+\infty} f(s) ds = 1.$$

Analogamente a quanto fatto nel caso delle variabili discrete, una funzione che soddisfa queste 2 proprietà verrà detta una *densità continua (astratta)*. Se conosciamo la densità continua f di una variabile aleatoria X possiamo calcolare la funzione di ripartizione F calcolando un integrale

$$F(t) = \int_{-\infty}^t f(s) ds.$$

Come possiamo determinare la densità se conosciamo la funzione di ripartizione? Osserviamo che

$$F(b) - F(a) = \int_a^b f(s) ds.$$

Il teorema fondamentale del calcolo integrale ci dice che tale formula vale se F è derivabile e $F'(t) = f(t)$. Abbiamo quindi il seguente fatto: se una variabile continua X ha funzione di ripartizione $F(t)$ derivabile (tranne al più in un numero finito di punti), allora $f(t) = F'(t)$ è la densità di X .

Osserviamo inoltre che la funzione di ripartizione F di una variabile continua X soddisfa le seguenti proprietà:

- (1) F è continua su \mathbb{R} ;
- (2) F è debolmente crescente;
- (3) $\lim_{t \rightarrow -\infty} F(t) = 0$ e $\lim_{t \rightarrow +\infty} F(t) = 1$.

Una funzione che soddisfa queste condizioni viene anche detta una funzione di ripartizione astratta.

La nozione di indipendenza tra variabili rimane la stessa, cioè diremo che 2 variabili X ed Y sono indipendenti se presi comunque t, u

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

2. La media e la varianza

In generale la media di una variabile continua di densità f si definisce tramite

$$E[X] = \int_{-\infty}^{+\infty} s f(s) ds.$$

Come siamo oramai abituati a fare, definiamo la varianza in termini della media tramite

$$\text{Var}(X) = E[(X - E[X])^2],$$

e si può verificare anche in questo caso che vale la formula

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

Le proprietà che conosciamo per la media e la varianza di una variabile discreta continuano a valere anche per variabili continue. In particolare, abbiamo che la media è lineare cioè

$$E(aX + bY) = aE[X] + bE[Y],$$

quali che siano le variabili continue X ed Y e i coefficienti $a, b \in \mathbb{R}$.

La varianza soddisfa la proprietà quadratica

$$\text{Var}(aX) = a^2 \text{Var}(X).$$

Se X ed Y sono indipendenti valgono le formule

$$E[XY] = E[X]E[Y]$$

e

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

Infine, se $\phi : \mathbb{R} \rightarrow \mathbb{R}$ è una funzione continua ed X è una variabile di densità f allora

$$E[\phi(X)] = \int_{-\infty}^{+\infty} \phi(s) f(s) ds.$$

3. Densità uniforme

La variabile continua più semplice da studiare è quella che può assumere solo valori in un certo intervallo limitato e in modo "uniforme".

Cominciamo con un esempio. Si consideri la variabile X avente funzione di ripartizione

$$F_X(t) = \begin{cases} 0 & \text{se } t \leq 0 \\ t & \text{se } 0 \leq t \leq 1 \\ 1 & \text{se } t \geq 1 \end{cases}$$

Osserviamo che $F_X(t)$ soddisfa le condizioni di una funzione di ripartizione astratta e che è derivabile ovunque tranne in 0 e in 1. Allora X ammette densità data dalla derivata di F_X , cioè

$$f(s) = \begin{cases} 1 & \text{se } 0 < s < 1 \\ 0 & \text{altrimenti.} \end{cases}$$

I valori della densità f nei punti di non derivabilità di F non sono influenti e quindi possono anche non essere definiti. In generale tendiamo a definirli in modo che la densità sia continua a destra o a sinistra. In questo caso potevamo definire indifferentemente $f(0) = 0$ o $f(0) = 1$. Questa variabile si dice *uniforme* nell'intervallo $[0, 1]$.

Vediamo più in generale una variabile uniforme nell'intervallo $[a, b]$. È definita in modo che la sua densità sia costante nell'intervallo $[a, b]$ e sia nulla altrove. Questo implica che la probabilità che X ricada in un certo sottointervallo I di $[a, b]$ sia proporzionale all'ampiezza dell'intervallo. Infatti se la densità è costante uguale a k e $I = [c, d]$ abbiamo

$$P(X \in [c, d]) = \int_c^d k ds = k(d - c)$$

(che altri non è che l'area del rettangolo di base I e altezza k , vedi Figura 1)

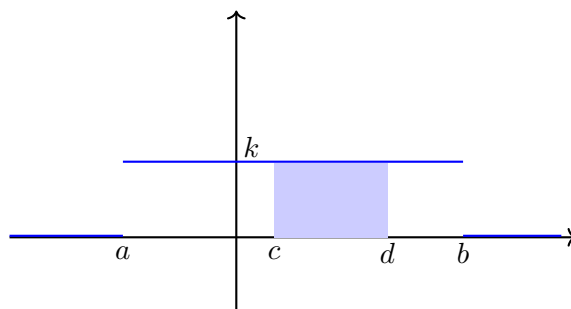


FIGURE 1. Densità uniforme.

Siccome l'area totale racchiusa tra il grafico della densità e l'asse x deve essere 1 si ha che necessariamente abbiamo che la densità uniforme nell'intervallo $[a, b]$ è data da

$$f(s) = \begin{cases} \frac{1}{b-a} & s \in [a, b] \\ 0 & \text{altrimenti.} \end{cases}$$

La sua funzione di ripartizione la ricaviamo (facilmente?) per integrazione della densità: otteniamo

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

si veda la Figura 2.

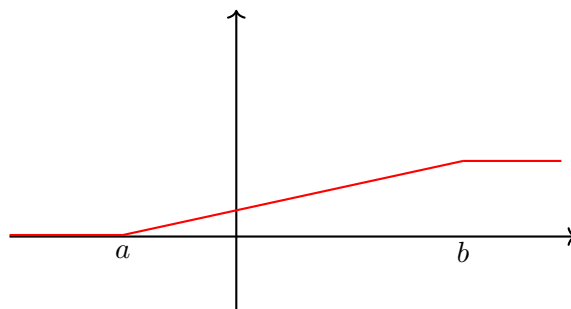


FIGURE 2. Funzione di ripartizione uniforme.

Determiniamo ora la media di una variabile uniforme nell'intervallo $[a, b]$. In tal caso si ha

$$E[X] = \int_a^b s \frac{1}{b-a} ds = \frac{1}{b-a} \left(\frac{b^2}{2} - \frac{a^2}{2} \right) = \frac{a+b}{2}.$$

La media è quindi, come ci si aspettava, il punto medio dell'intervallo $[a, b]$. Per determinare la varianza calcoliamo la media di X^2 . In tal caso calcoliamo

$$E[X^2] = \int_a^b s^2 \frac{1}{b-a} ds = \frac{1}{b-a} \frac{b^3 - a^3}{3} = \frac{a^2 + ab + b^2}{3}$$

e quindi

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

4. Densità esponenziale

Fissiamo un parametro a e ricordiamo che una primitiva della funzione e^{-at} è data da $-\frac{1}{a}e^{-at}$ da cui abbiamo che:

$$(5) \quad \int_0^{+\infty} e^{-at} dt = \frac{1}{a}.$$

PROPOSIZIONE 9.2. *La funzione*

$$f(s) = \begin{cases} ae^{-as} & \text{se } s \geq 0 \\ 0 & \text{altrimenti} \end{cases},$$

dove a è un parametro reale positivo, è una densità continua astratta.

PROOF. Si ha chiaramente $f(s) \geq 0$ per ogni $s \in \mathbb{R}$. Quanto all'integrale

$$\int_{-\infty}^{+\infty} f(s) ds = \int_0^{\infty} ae^{-as} ds = 1$$

grazie a (5). □

Una variabile X che ammette tale densità si dice *esponenziale* di parametro a e scriveremo $X \sim \text{Exp}(a)$.

Lo stesso calcolo mostra che la funzione di ripartizione di una variabile $X \sim \text{Exp}(a)$ è data da

$$F_X(t) = \begin{cases} 0 & \text{se } t \leq 0 \\ 1 - e^{-at} & \text{se } t \geq 0 \end{cases}$$

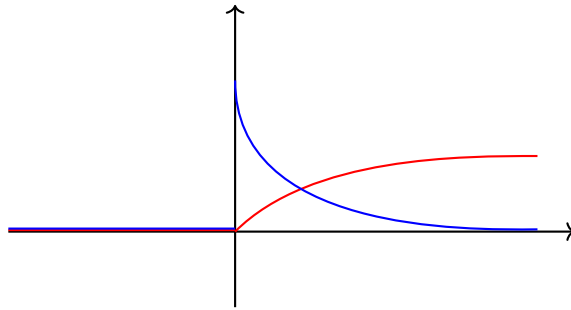


FIGURE 3. Densità (in blu) e funzione di ripartizione (in rosso) esponenziale.

Una variabile esponenziale soddisfa la proprietà di mancanza di memoria. Infatti

$$P(X \geq t+u | X \geq t) = \frac{P(X \geq t+u)}{P(X \geq t)} = \frac{e^{-a(t+u)}}{e^{-at}} = e^{-au} = P(X \geq u).$$

Per questo motivo una variabile esponenziale non è adatta a modellizzare una variabile di tipo "tempo di rottura". Infatti è chiaro che il fatto che un apparecchio non si sia rotto per un

certo tempo influenza la probabilità che si possa rompere in un certo periodo, a causa dell'usura dell'apparecchio stesso. Alcuni fenomeni naturali possono invece essere ben modellizzati tramite una variabile esponenziale, tipo il "tempo di decadimento di una particella radioattiva" o il tempo di eruzione di un certo vulcano.

Determiniamo ora media e varianza di una variabile $X \sim \text{Exp}(a)$. Si ha

$$\begin{aligned} E[X] &= \int_0^\infty sae^{-as} ds \\ &= \left[-se^{-as} \right]_0^\infty + \int_0^\infty e^{-as} ds \\ &= \frac{1}{a} \end{aligned}$$

dove abbiamo utilizzato (5). Mentre

$$\begin{aligned} E[X^2] &= \int_0^\infty s^2 ae^{-as} ds \\ &= \left[-s^2 e^{-as} \right]_0^\infty + \int_0^\infty 2se^{-as} ds \\ &= 0 + \frac{2}{a} \int_0^\infty sae^{-as} ds \\ &= \frac{2}{a} \frac{1}{a} \\ &= \frac{2}{a^2} \end{aligned}$$

dove abbiamo sfruttato anche il calcolo precedente per determinare $E[X]$. Concludiamo che

$$\text{Var}(X) = E[X^2] - E[X]^2 = \frac{2}{a^2} - \frac{1}{a^2} = \frac{1}{a^2}.$$

5. Trasformazioni di variabili

Anche nel caso di variabili continue possiamo considerare alcune (semplici) trasformazione che portatno a considerare nuove variabili. Se X è una variabile aleatoria continua e

$$\phi : \mathbb{R} \rightarrow \mathbb{R},$$

non è detto che $\phi(X)$ sia ancora una variabile aleatoria. Però lo è se ϕ è continua tranne al più in un numero finito di punti. In tal caso come si può fare per calcolare la densità della variabile trasformata $\phi(X)$? Sia f la densità di X e G la funzione di ripartizione di $\phi(X)$. Allora

$$G(t) = P(\phi(X) \leq t) = P(X \in \phi^{-1}((-\infty, t])) = \int_{\phi^{-1}((-\infty, t])} f(s) ds$$

Se si è fortunati si può esprimere quest'ultima quantità in termini della funzione di ripartizione F di X . E poi ottenere la densità di $\phi(X)$ derivando $G(t)$.

ESEMPIO 9.3. Sia X una variabile continua di densità f . Qual è la densità di X^2 ? Sia G la funzione di ripartizione di X^2 . Allora

$$G(t) = P(X^2 \leq t) = P(-\sqrt{t} \leq X \leq \sqrt{t}) = F_X(\sqrt{t}) - F_X(-\sqrt{t}).$$

Se questa funzione è derivabile possiamo calcolare la densità g di X^2

$$g(t) = G'(t) = -\frac{1}{2\sqrt{t}}(f(\sqrt{t}) - f(-\sqrt{t})).$$

ESEMPIO 9.4. Se X ha densità f , qual è la densità di $aX + b$ (con $a \neq 0$)? Se $a > 0$ si ha $G(t) = P(aX + b \leq t) = P(X \leq \frac{t-b}{a}) = F(\frac{t-b}{a})$ e quindi

$$g(t) = G'(t) = \frac{1}{a}f\left(\frac{t-b}{a}\right).$$

In generale, si ottiene similmente $g(t) = \frac{1}{|a|}f\left(\frac{t-b}{a}\right)$.

ESEMPIO 9.5. Siano $X \sim \text{Exp}(a)$ ed $Y \sim \text{Exp}(b)$ variabili indipendenti. Determinare la densità della variabile $Z = \max(X, Y)$. Facendo un calcolo simile a quello fatto nel caso di variabili geometriche nell'esempio 7.5 si ha

$$F_Z(t) = P(X \leq t)P(Y \leq t) = F_X(t)F_Y(t) = (1 - e^{-at})(1 - e^{-bt}).$$

E per derivazione otteniamo

$$f_Z(t) = ae^{-at}(1 - e^{-bt}) + be^{-bt}(1 - e^{-at}).$$

Se poniamo $W = \min(X, Y)$ possiamo invece verificare che $W \sim \text{Exp}(a + b)$.

6. Densità normali

Passiamo ora a considerare quelle che sono probabilmente le variabili aleatorie continue più rilevanti.

Si può dimostrare (e Gauss lo ha fatto per primo) che

$$(6) \quad \int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} dx = \sqrt{2\pi}.$$

Ne segue che la funzione

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

è una densità ed è rappresentata nella Figura 4. Questa densità viene detta normale standard o

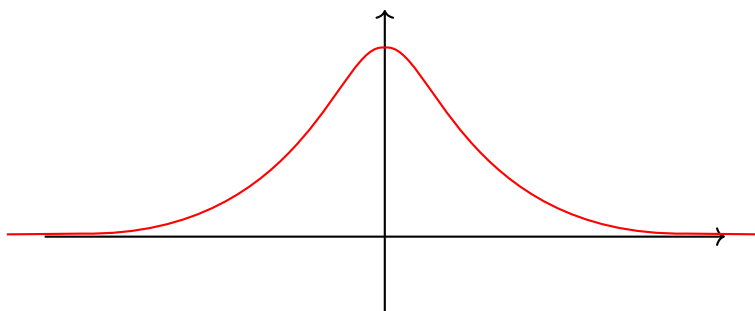


FIGURE 4. Densità normale standard.

gaussiana standard. Se una variabile X ammette questa densità scriveremo

$$X \sim N(0, 1).$$

Utilizzeremo il simbolo ζ_0 per indicare una variabile normale standard.

Siano ora $\sigma > 0$ e μ due parametri reali e $\zeta_0 \sim N(0, 1)$. Consideriamo la variabile $\zeta = \sigma\zeta_0 + \mu$. Questa variabile avrà densità

$$f_\zeta(s) = \frac{1}{\sigma} f_{\zeta_0}\left(\frac{s - \mu}{\sigma}\right) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(s - \mu)^2}{2\sigma^2}}.$$

Scriveremo in questo caso $\zeta \sim N(\mu, \sigma^2)$.

Sarà sovente il caso, ogni qual volta abbiamo a che fare con una variabile $\zeta \sim N(\mu, \sigma^2)$, di effettuare il cambio di variabile $\zeta_0 = \frac{\zeta - \mu}{\sigma}$. La variabile ζ_0 così definita è una normale standard. E in questo modo potremo fare i calcoli con la variabile ζ_0 per poi ritradurre i risultati ottenuti nella variabile ζ .

Il parametro μ trasla la densità standard verso destra o verso sinistra, mentre il parametro σ aumenta o diminuisce l'ampiezza della campana. In Figura 5 sono riportate diverse densità normali.

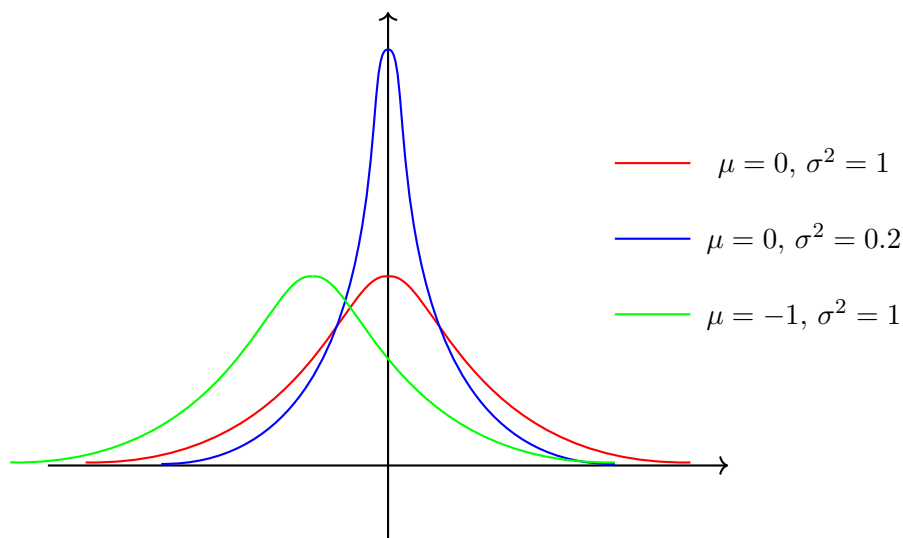


FIGURE 5. Densità normali al variare dei parametri.

Poniamo $\Phi(x)$ la funzione di ripartizione di una variabile normale standard. I suoi valori sono riportati su tabelle, osservando che $\Phi(x)$ è molto prossimo a zero se $x < -3$ e molto prossimo a 1 se $x > 3$. Inoltre grazie alla simmetria della densità normale standard si ha $\Phi(-x) = 1 - \Phi(x)$ e sarà sufficiente calcolare la $\Phi(x)$ solo per valori compresi tra 0 e 3.

ESEMPIO 9.6. Studi sperimentali mostrano che l'altezza di un uomo adulto ha densità normale $N(\mu, \sigma^2)$ con $\mu = 173\text{cm}$ e $\sigma = 9\text{cm}$. Una divisa per reclute è disponibile in 5 taglie

Taglia	Altezza
XS	$h < 162$
S	$162 < h < 170$
M	$170 < h < 180$
L	$180 < h < 190$
XL	$h > 190$

Dovendo vestire 500 nuove reclute quante divise di ogni taglia comprereste?

Sia ζ la variabile altezza di un uomo adulto scelto a caso abbiamo

$$P(\zeta < 162) = P\left(\frac{\zeta - 173}{9} < \frac{162 - 173}{9}\right) = P(\zeta_0 < -1.22) = \Phi(-1.22) = 1 - \Phi(1.22) = 1 - 0.88877 = 0,11123.$$

Similmente calcoliamo

$$\begin{aligned} P(162 < \zeta < 170) &= P(-1.22 < \zeta_0 < -0,33) = \Phi(-0,33) - \Phi(-1.22) = \Phi(1.22) - \Phi(0.33) \\ &= 0.88877 - 0.62930 = 0,25947. \end{aligned}$$

$$\begin{aligned} P(170 < \zeta < 180) &= P(-0,33 < \zeta_0 < 0.78) = \Phi(0,78) - \Phi(-0,33) = \Phi(0,78) - 1 + \Phi(0,33) \\ &= 0,78230 - 1 + 0.62930 = 0,4116 \end{aligned}$$

$$P(180 < \zeta < 190) = P(0,78 < \zeta_0 < 1.89) = \Phi(1,89) - \Phi(0,78) = 0,97062 - 0,7823 = 0,18832$$

$$P(\zeta > 190) = P(\zeta_0 > 1,89) = 1 - \Phi(1,89) = 1 - 0,97062 = 0,02938$$

Il numero di divise XS è stimato da $0,11123 \cdot 500 = 55,615$, le S da $0,25947 \cdot 500 = 129,735$, le M da $0,4116 \cdot 500 = 205,8$, le L da $0,18832 \cdot 500 = 94,16$, le XL da $0,02938 \cdot 500 = 14,69$. Io ne comprerei quindi 55 XS, 130 S, 206 M, 94 L, e 15 XL.

PROPOSIZIONE 9.7. Sia $\zeta_0 \sim N(0,1)$. Allora $E[\zeta_0] = 0$ e $\text{Var}(\zeta_0) = 1$.

PROOF. Se $\zeta_0 \sim N(0,1)$ si ha $E[\zeta_0] = 0$ in quanto la densità di ζ_0 è pari. Per calcolarne la varianza dobbiamo determinare la media di ζ_0^2 facendo un'integrazione per parti:

$$\begin{aligned} E[\zeta_0^2] &= \int_{-\infty}^{+\infty} x^2 \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx \\ &= \int_{-\infty}^{+\infty} x \cdot \frac{1}{\sqrt{2\pi}} x e^{-\frac{x^2}{2}} dx \\ &= \left[-x \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \right]_{-\infty}^{+\infty} - \int_{-\infty}^{+\infty} -\frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx \\ &= 0 + 1, \end{aligned}$$

dove si è sfruttato (6). Concludiamo quindi che $\text{Var}(\zeta_0) = 1$. □

COROLLARIO 9.8. Sia $\zeta \sim N(\mu, \sigma^2)$ una variabile normale. Allora

$$E[\zeta] = \mu,$$

e

$$\text{Var}[\zeta] = \sigma^2.$$

DIMOSTRAZIONE. Il risultato segue immediatamente dal fatto che $\zeta = \sigma\zeta_0 + \mu$ e utilizzando le proprietà della media e della varianza:

$$E[\zeta] = \sigma E[\zeta_0] + \mu = \mu$$

e

$$\text{Var}[\zeta] = \text{Var}(\sigma\zeta_0 + \mu) = \sigma^2 \text{Var}(\zeta_0) + \text{Var}(\mu) = \sigma^2.$$

□

Abbiamo quindi che i parametri μ e σ^2 sono proprio la media e la varianza di una variabile normale $\zeta \sim N(\mu, \sigma^2)$.

ESEMPIO 9.9. La lunghezza delle foglie di lauro ha densità normale di media $\mu = 151\text{mm}$ e scarto quadratico medio $\sigma = 15\text{mm}$. Determinare la probabilità che una foglia scelta a caso abbia

- (1) lunghezza compresa tra i 120 e i 155 millimetri;
- (2) lunghezza maggiore di 185mm

Se ζ è la variabile lunghezza delle foglie di lauro si ha $\zeta = N(151, (15)^2)$ e quindi Si ha

$$P(120 < \zeta < 155) = P(-31 < \zeta - 151 < 4) = P(-2,07 < \frac{\zeta - 151}{15} < 0,27).$$

Ricordando a questo punto che $\zeta_0 = \frac{\zeta - 151}{15}$ è una variabile normale standard abbiamo

$$\begin{aligned} P(120 < \zeta < 155) &= P(-2,07 < \zeta_0 < 0,27) = \Phi(0,27) - \Phi(-2,07) \\ &= \Phi(0,27) - 1 + P(2,07) = 0,60642 - 1 + 0,98077 \\ &= 58.7\% \end{aligned}$$

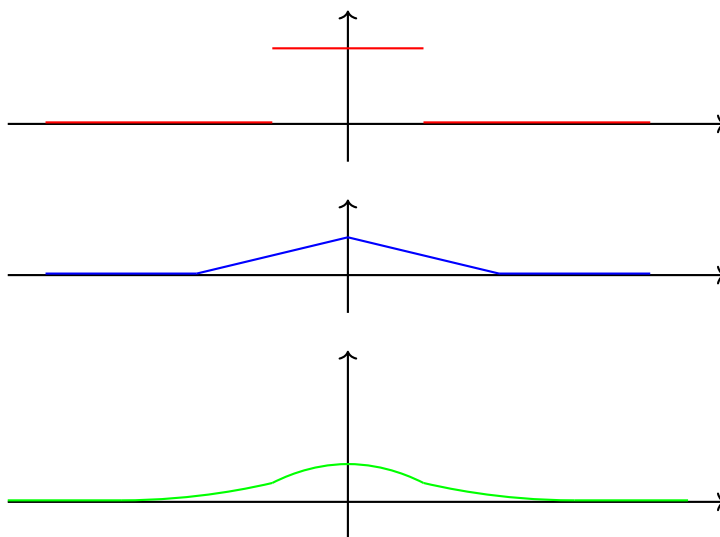
Similmente

$$P(\zeta > 185) = P(\zeta_0 > \frac{185 - 151}{15}) = P(\zeta_0 > 2,27) = 1 - \Phi(2,27) = 1 - 0,98840 = 1,16\%.$$

7. Il teorema del limite centrale

Consideriamo una sequenza di variabili X_1, X_2, \dots uniformi nell'intervallo $[-1, 1]$ e tutte indipendenti tra loro. Non abbiamo sviluppato gli strumenti per determinare la densità della variabile $S_2 = X_1 + X_2$. Possiamo però renderci conto che è ragionevole che questa densità sia $\frac{1}{4}(2 - t)$ per $0 < t < 2$ e che sia simmetrica rispetto all'asse delle ordinate. Se provassimo a calcolare la densità di $S_3 = X_1 + X_2 + X_3$ otterremmo 3 rami di parabola negli intervalli $[-3, -1]$, $[-1, 1]$ e $[1, 3]$ che messi insieme ricordano un po' la forma a campana della densità gaussiana. Questa non è una caratteristica delle variabili uniformi, vedremo tra breve che si tratta di una proprietà molto più generale.

Consideriamo la variabile $S_n = X_1 + \dots + X_n$ e ricordiamo la variabile media campionaria $\bar{X}_n = \frac{1}{n}S_n$. Se le variabili X_i hanno tutte la stessa densità (discreta o continua) di media μ e

FIGURE 6. Densità della somma di 1,2,3 variabili $U([-1, 1])$ indipendenti.

varianza σ^2 e sono indipendenti sappiamo calcolare media e varianza di S_n e \bar{X}_n . In particolare

$$\begin{aligned} E[S_n] &= n\mu \\ \text{Var}(S_n) &= n\sigma^2 \\ E[\bar{X}_n] &= \mu \\ \text{Var}(\bar{X}_n) &= \frac{\sigma^2}{n} \end{aligned}$$

Il seguente fondamentale teorema ci dice che se n è abbastanza grande questo è tutto quello che ci serve per risolvere problemi di probabilità riguardanti le variabili aleatorie S_n e \bar{X}_n .

TEOREMA 9.10 (del limite centrale). *Sia X_1, X_2, \dots una successione di variabili aleatorie indipendenti aventi tutte la stessa densità (discreta o continua) di media μ e varianza σ^2 . Allora per n abbastanza grande si ha con buona approssimazione*

$$S_n \sim N(n\mu, n\sigma^2)$$

e

$$\bar{X}_n \sim N(\mu, \sigma^2/n).$$

ESEMPIO 9.11. Una certa lampadina ha tempo di vita di densità esponenziale di media 10 giorni. Avendo comprato 40 di queste lampadine, qual è la probabilità che mi bastino per un anno? Quante lampadine dovrei comprare se voglio che questa probabilità sia almeno del 90%?

Il tempo di vita di una lampadina è una variabile esponenziale di parametro $a = 0.1$ e quindi la sua varianza è 100. Ne segue che $S_{40} \sim N(400, 4000)$ e quindi

$$P(S_{40} > 365) = P(\zeta > 365) = P(\zeta_0 > \frac{365 - 400}{\sqrt{4000}}) = P(\zeta_0 > -0.55) = \Phi(0.55) = 70.9\%.$$

Consideriamo la variabile S_n questa ha densità approssimativamente normale $N(10n, 100n)$ e quindi

$$P(\zeta > 365) = P(\zeta_0 > \frac{365 - 10n}{10\sqrt{n}}) > 0.90.$$

Guardando la tabella questo capita se $\frac{365-10n}{10\sqrt{n}} < -1.28$ che ha come soluzione $n \geq 46$.

Se X è una variabile discreta (che assuma solo valori interi) approssimata da una variabile normale ζ sarà sempre preferibile, per ottenere dei risultati più accurati, effettuare quella che si chiama “correzione di continuità”. Se ad esempio X assume solo valori interi l’evento $X = 5$ viene sostituito da $4.5 < \zeta < 5.5$, cioè ogni numero intero viene considerato come l’intervallo di ampiezza 1 centrato in se stesso. Per fare altri esempi abbiamo $2 \leq X \leq 5$ viene sostituito da $1.5 < \zeta < 5.5$ oppure $X > 4$ viene rimpiazzato da $\zeta > 4.5$.

ESEMPIO 9.12. In un volo di 24 posti, sappiamo che ogni passeggero ha probabilità di non presentarsi del 30%. Quanti biglietti posso vendere se ammetto una probabilità del 10 % che qualcuno non trovi posto?

Supponiamo di vendere n biglietti. Sia X_i la variabile di Bernoulli che vale 1 se il passeggero i si presenta. La variabile $S_n = X_1 + \dots + X_n$ è una variabile binomiale che per il teorema centrale del limite può essere approssimata da una variabile normale $\zeta \sim N(\frac{7}{10}n, \frac{21}{100}n)$, dove ricordiamo che se $X \sim B(1, p)$ allora $E[X] = p$ e $\text{Var}(X) = p(1 - p)$. L’evento “qualcuno non trova posto” si traduce in $S_n \geq 25$. Utilizzando l’approssimazione di continuità abbiamo che $S_n \geq 25$ corrisponde a $\zeta > 24.5$. Si ha quindi

$$P(\zeta > 24.5) = P\left(\zeta_0 > \frac{24.5 - \frac{7}{10}n}{\sqrt{\frac{21}{100}n}}\right) < 0.10.$$

Guardando alla tabella si ha che questo accade se $\frac{24.5 - \frac{7}{10}n}{\sqrt{\frac{21}{100}n}} > 1.28$. Risolvendo questa disequazione si ottiene $n \leq 30$. Si possono quindi vendere 30 biglietti correndo un rischio del 10% che qualcuno non trovi posto in aereo.

ESEMPIO 9.13. Alle ultime elezioni un partito ha ottenuto il 55% dei consensi. Scelti 100 votanti a caso, qual è la probabilità che almeno la metà di essi abbia votato per quel partito?

In questo caso abbiamo 100 variabili X_1, \dots, X_{100} , tutte indipendenti di densità $B(1, 0.55)$. Ci stiamo chiedendo qual è la probabilità che

$$P(\bar{X}_{100} \geq 0.50).$$

Per il teorema centrale del limite la variabile \bar{X}_{100} è approssimata da una normale $\zeta \sim N(0.55, \frac{0.55 \cdot 0.45}{100})$. Siccome la variabile \bar{X}_{100} è una variabile discreta possiamo anche effettuare una correzione di continuità e calcolare

$$\begin{aligned} P(\bar{X}_{100} \geq 0.50) &= P(\zeta > 0.495) = P\left(\zeta_0 > \frac{0.495 - 0.55}{\sqrt{\frac{0.55 \cdot 0.45}{100}}}\right) \\ &= P(\zeta_0 > -0.055/0,049749) = P(\zeta_0 > -1,10) = \Phi(1.10) = 86.4\% \end{aligned}$$

ESEMPIO 9.14. Sia X la variabile “diametro di un'albicocca di un certo campo”. Si sa che $E[X] = 5\text{cm}$ con $\sigma_X = 1$. Scegliendo 100 albicocche, qual è la probabilità che la media dei diametri delle albicocche sia superiore a 5.1cm? Dette X_1, \dots, X_{100} . Dobbiamo valutare

$$P(\bar{X}_n > 5.1).$$

La variabile \bar{X}_n è approssimativamente una $\zeta \sim N(5, \frac{\sigma^2}{n}) = N(5, 0.01)$. Ne segue che

$$P(\zeta > 5.1) = P(\zeta_0 > \frac{5.1 - 5}{0.01}) = P(\zeta_0 > 10) = 0.$$

CHAPTER 10

Statistica inferenziale

1. Stime e intervalli di confidenza

Il teorema del limite centrale ci permette di calcolare facilmente delle probabilità riguardanti un campione numeroso. Cioè se X è una variabile aleatoria definita su una certa popolazione di cui conosciamo media e varianza possiamo calcolare probabilità che riguardano la media di X su un certo campione, cioè probabilità del tipo

$$P(a < \bar{X}_n < b),$$

semplicemente riformulando il problema in termini di una variabile normale standard ζ_0 .

Il problema principale in questo tipo di approccio è che in generale non si hanno a disposizione media e varianza di una variabile. Ma anzi vorremmo dei metodi che ci permettano di stimarle. E vorremmo anche un metodo per stabilire quanto precise siano queste stime.

Ci sarà utile la seguente notazione. Dato $0 < c < 1$ sia z_c tale che $P(-z_c < \zeta_0 < z_c) = c$. Si ha quindi $\Phi(z_c) - \Phi(-z_c) = c$ e quindi $2\Phi(z_c) - 1 = c$ da cui $\Phi(z_c) = \frac{c+1}{2}$. Ad esempio, se $c = 0.95$ abbiamo $\Phi(z_c) = 0.975$ da cui $z_{0.95} = 1.96$.

Sia quindi X una variabile aleatoria definita su una certa popolazione. Siano $\mu_X = E[X]$ e $\sigma_X^2 = \text{Var}(X)$ la media e la varianza di X . Consideriamo la variabile media campionaria \bar{X}_n definita sullo spazio costituito dai campioni di n elementi scelti dalla nostra popolazione. Per il teorema centrale del limite sappiamo

$$\bar{X}_n \sim N(E[\bar{X}_n], \text{Var}(\bar{X}_n)) = N(\mu_X, \frac{\sigma_X^2}{n}).$$

Sappiamo quindi che

$$\frac{\bar{X}_n - \mu_X}{\sigma_X/\sqrt{n}} \sim N(0, 1).$$

Di conseguenza abbiamo

$$P(-z_c < \frac{\bar{X}_n - \mu_X}{\sigma_X/\sqrt{n}} < z_c) = c$$

e moltiplicando tutti i membri per σ_X/\sqrt{n}

$$P(-z_c \frac{\sigma_X}{\sqrt{n}} < \bar{X}_n - \mu_X < z_c \frac{\sigma_X}{\sqrt{n}}) = c$$

e sottraendo da tutti i membri \bar{X}_n

$$P(-z_c \frac{\sigma_X}{\sqrt{n}} - \bar{X}_n < -\mu_X < z_c \frac{\sigma_X}{\sqrt{n}} - \bar{X}_n) = c$$

e infine cambiando i segni e rigirando le disuguaglianze otteniamo

$$P(\bar{X}_n - z_c \frac{\sigma_X}{\sqrt{n}} < \mu_X < \bar{X}_n + z_c \frac{\sigma_X}{\sqrt{n}}) = c.$$

Osserviamo che in quest'ultima equazione abbiamo rigirato il punto di vista: fa un'affermazione riguardo la media della variabile X su tutta la popolazione (μ_X) rispetto al valore medio registrato su un singolo campione \bar{X}_n . Il problema è che tale probabilità viene espressa in termini della varianza σ_X^2 che comunque non conosciamo. In questo caso allora sfruttiamo la legge dei grandi numeri per la varianza per ricordare che la varianza σ_X^2 è ben approssimata dalla varianza campionaria $\bar{\sigma}_n^2$. Utilizziamo questo valore al posto di σ_X^2 che possiamo calcolare nel campione e parliamo di confidenza anziché di probabilità. Abbiamo quindi

$$(7) \quad \text{Conf}(\bar{X}_n - z_c \frac{\bar{\sigma}_n}{\sqrt{n}} < \mu_X < \bar{X}_n + z_c \frac{\bar{\sigma}_n}{\sqrt{n}}) = c$$

e diciamo che la media μ_X è contenuta nell'intervallo $[\bar{X}_n - z_c \frac{\bar{\sigma}_n}{\sqrt{n}}, \bar{X}_n + z_c \frac{\bar{\sigma}_n}{\sqrt{n}}]$ con confidenza c .

ESEMPIO 10.1. Si consideri un campione di 100 albicocche di un certo frutteto. Si registra un diametro medio $m = 5.1\text{cm}$ ed uno scarto quadratico medio $s = 1\text{cm}$.

- Determinare l'intervallo di confidenza al 95% per la media di tutte le albicocche prodotte in questo frutteto;
- Determinare con che confidenza tale media appartiene all'intervallo $[5, 5.2]$.

Basterà sostituire nella formula (7) $\bar{\sigma}_n = 1$, $\bar{X}_n = 5.1$, $n = 100$ e $z_c = 1.96$. Otteniamo la semiampiezza $z_c \bar{\sigma}_n / \sqrt{n} = 1.96/10 = 0.196$ e quindi l'intervallo di confidenza al 95% per μ_X è

$$[5.1 - 0.196, 5.1 + 0.196] = [4.904, 5.296].$$

La confidenza con cui μ_X appartiene all'intervallo $[5, 5.2]$ sarà senz'altro inferiore al 95%. Vogliamo che la semiampiezza dell'intervallo sia 0.1, cioè

$$z_c \bar{\sigma}_n / \sqrt{n} = 0.1,$$

da cui deduciamo che $z_c = 0.1 \cdot 10/1 = 1$. Di conseguenza $c = 2\phi(z_c) - 1 = 2 \cdot 0.8413 - 1 = 68.3\%$.

Vediamo ora il caso particolare in cui la variabile che si vuole stimare è una variabile di Bernoulli. Questo capita ogni volta che si vuole stimare con che probabilità un dato fenomeno darà successo oppure no. Ad esempio: realizzare un rigore, passare un esame, ottenere testa dal lancio di una moneta truccata, vincere le elezioni...

In tutti questi casi abbiamo a che fare con una variabile $X \sim B(1, p)$ di cui non conosciamo il parametro p . E vogliamo determinare un intervallo di confidenza per questo parametro. Questo è un caso particolare dell'esempio trattato precedentemente, ma lo esaminiamo anche singolarmente data la sua importanza. Preso un campione di n elementi di una popolazione la variabile \bar{X}_n ci dà in questo caso la proporzione di successi ottenuti nel campione. Andiamo quindi a considerare ancora una volta la formula (7)

$$\text{Conf}(\bar{X}_n - z_c \frac{\bar{\sigma}_n}{\sqrt{n}} < \mu_X < \bar{X}_n + z_c \frac{\bar{\sigma}_n}{\sqrt{n}}) = c.$$

In questo caso abbiamo $\mu_X = p$. Quanto vale $\bar{\sigma}_n$? Basterà ricordare come si calcola la varianza di un campione in statistica descrittiva. Sugli n elementi del campione abbiamo ottenuto $n\bar{X}_n$ volte 1 e $n(1 - \bar{X}_n)$ volte 0. Ricordando che la varianza è la media dei quadrati meno il quadrato della media abbiamo

$$\bar{\sigma}_n^2 = \bar{X}_n - \bar{X}_n^2$$

e abbiamo quindi in questo caso

$$\text{Conf}(\bar{X}_n - z_c \sqrt{\frac{\bar{X}_n - \bar{X}_n^2}{n}} < p < \bar{X}_n + z_c \sqrt{\frac{\bar{X}_n - \bar{X}_n^2}{n}}) = c.$$

ESEMPIO 10.2. Abbiamo una moneta che non sappiamo se è truccata o no. La lanciamo 80 volte ed otteniamo 43 volte testa. Determinare l'intervallo di confidenza al 90% per la probabilità che la moneta dia testa.

La variabile aleatoria da considerare è una variabile di Bernoulli $X \sim B(1, p)$ che vale 1 se la moneta dà testa. Si vuole quindi stimare la probabilità p che la moneta dia testa. Calcoliamo intanto il coefficiente

$$z_{0.90} = 1.64.$$

In questo caso abbiamo $n = 80$ e $\bar{X}_n = \frac{43}{80}$. L'intervallo di confidenza al 90% ha quindi semiampiezza

$$z_c \sqrt{\frac{\bar{X}_n - \bar{X}_n^2}{n}} = 1.64 \cdot \sqrt{\frac{\frac{43}{80} - \frac{43^2}{80^2}}{80}} = 0,0557$$

e quindi l'intervallo di confidenza è dato da

$$\left[\frac{43}{80} - 0,0557, \frac{43}{80} + 0,0557\right] = [0.482, 0.593].$$

2. Test statistici

Nell'ultimo esempio ci siano trovati di fronte ad una moneta non sapendo se fosse buona o truccata e abbiamo dato una stima per la probabilità che dia testa. Ora però potrebbe essere naturale porsi la domanda “la moneta è buona o è truccata?” e dover dare una risposta che può essere solo sì o no. Questo tipo di domanda capita spesso in uno studio sulla qualità di un prodotto: un prodotto viene sottoposto ad un test un certo numero di volte e si vuole decidere se il prodotto passa il test oppure no.

In un test statistico si ha a che fare con una variabile aleatoria X e si formulano due tipi di ipotesi che denotiamo con H_0 e H_1 , che si chiamano solitamente ipotesi nulla e ipotesi alternativa. Nei nostri primi esempi l'ipotesi H_0 sarà del tipo

$$H_0 : E[X] = \mu,$$

mentre l'ipotesi H_1 è del tipo

$$H_1 : E[X] \neq \mu.$$

Alla fine del test dobbiamo prendere una decisione: accettare l'ipotesi H_0 , oppure rifiutarla e quindi accettare l'ipotesi alternativa H_1 . E' chiaro che in tal modo di procedere si può incorrere in due tipi di errori:

- Errore del primo tipo: H_0 è vera, ma viene rifiutata;
- Errore del secondo tipo: H_0 è falsa ma viene accettata.

In questa breve trattazione dei test statistici ci occupiamo solo di errori del primo tipo: la *significatività* di un test è la probabilità di commettere un errore del primo tipo. E' importante, quando si effettua un test, stabilire a priori, cioè prima di effettuare il test stesso, i criteri di accettazione o di rifiuto dell'ipotesi H_0 .

ESEMPIO 10.3. Si vuole stabilire se una moneta è buona o truccata con un test con significatività al 5%. Si supponga di effettuare 80 lanci. In questo caso formuliamo l'ipotesi H_0 : la moneta è buona, cioè $p = 0.5$ dove $X \sim B(1, p)$ è la variabile che vale 1 se e solo se la moneta dà testa. Se l'ipotesi H_0 è vera vogliamo determinare l'intervallo in cui ricade la variabile $\bar{X}_{80} =$ “proporzione di teste ottenute” con probabilità 95%. Sotto l'ipotesi H_0 sappiamo che \bar{X}_{80} è una variabile approssimativamente normale di densità $N(0.5, 0.003125)$ e quindi

$$\frac{\bar{X}_{80} - 0.5}{\sqrt{0.003125}} = \frac{\bar{X}_{80} - 0.5}{0.0559} \sim N(0, 1)$$

Ricordando che $z_{0.95} = 1.96$ si ha quindi che $\zeta_0 \in [-1.96, 1.96]$ con probabilità 95%. Si ha quindi

$$0.401 = -1.96 \cdot 0.0559 + 0.5 < \bar{X}_{80} < 1.96 \cdot 0.0559 + 0.5 = 0.609,$$

con probabilità al 95%. Queste proporzioni corrispondono ad un numero di teste pari a $0.401 \cdot 80 = 32.08$ e $0.609 \cdot 80 = 48.72$. Possiamo quindi affermare che rifiutare l'ipotesi “la moneta è buona” avendo ottenuto un numero di teste inferiore a 33 o superiore a 48, la probabilità di sbagliare è inferiore al 5%. Questo test ha quindi una significatività al 5%.

ESEMPIO 10.4. Il tempo di vita medio di un campione di 100 lampadine prodotte da una fabbrica è 1570h, con uno scarto quadratico medio di 120 h. Se μ è il tempo di vita medio di tutte le lampadine prodotte da questa fabbrica, provare l'ipotesi $\mu = 1600h$ con significatività al 5% e all'1%.

Dobbiamo decidere tra le 2 ipotesi $H_0 : E[X] = 1600$ e $H_1 : E[X] \neq 1600$. Nell'ipotesi H_0 la variabile \bar{X}_{100} ha media 1600 e varianza $\text{Var}(X)/100$. Non conoscendo la varianza di X utilizziamo in questo caso la varianza campionaria 120^2 per stimarla. Abbiamo quindi che

$$(\bar{X}_{100} - 1600)/12$$

è una normale standard. Possiamo quindi concludere che accettiamo l'ipotesi se

$$-z_c < (\bar{X}_{100} - 1600)/12 < z_c.$$

Si ha $(\bar{X}_{100} - 1600)/12 = -2.5$. Essendo $z_{0.95} = 1.96$ rifiutiamo l'ipotesi con significatività al 5%. Siccome invece $z_{0.99} = 2.58$ accettiamo l'ipotesi con significatività all'1%.

Talvolta può essere naturale considerare come ipotesi nulla $H_0 : E[X] \geq \mu$ e come ipotesi alternativa $H_1 : E[X] < \mu$ (o $H_0 : E[X] \leq \mu$ e $H_1 : E[X] > \mu$). Ad esempio se si vuole testare se una certa medicina è efficace in almeno una certa percentuale di casi, o se si vuole testare se le mele i un certo frutteto pesano mediamente almeno un certo peso prefissato. In questi casi si effettua un cosiddetto test ad una coda in quanto si va a considerare solo una delle due code della densità normale per rifiutare l'ipotesi. In altre parole poniamo w_c quel valore tale che

$$\phi(w_c) = c$$

e in un test ad una coda con significatività $1-c$ in cui $H_0 : E[X] \geq \mu$ e $H_1 : E[X] < \mu$ rifiutiamo l'ipotesi se il valore registrato standardizzato è inferiore a $-w_c$. Vediamone un esempio.

ESEMPIO 10.5. Un tizio dice di avere poteri extrasensoriali. Gli viene chiesto di indovinare il colore (rosso o blu) di una carta scelta da un mazzo di 50 carte e ne indovina 32. Stabilire se il tizio ha davvero poteri extrasensoriali o no con una significatività del 5% e dell'1%.

In questo caso abbiamo l'ipotesi nulla $H_0 : p \leq 0.5$ e $H_1 : p > 0.5$, dove p è la probabilità che il tizio indovini una carta scelta a caso. Effettuiamo quindi un test ad una coda e rifiutiamo l'ipotesi H_0 se il risultato cade nella coda di destra relativa all'ultimo 5% o all'ultimo 1% rispettivamente. Accettando l'ipotesi H_0 nel caso limite $p = 0.5$ abbiamo che \bar{X}_{50} è approssimativamente normale di media 0.5 e scarto quadratico medio $\sqrt{0.5 \cdot 0.5/50} = 0,070$. Rifiutiamo l'ipotesi quindi se

$$(\bar{X}_{50} - 0.5)/0.070 > w_c.$$

Si ha $(\bar{X}_{50} - 0.5)/0.070 = 2$. Siccome abbiamo $w_{0.95} = 1.64$ e $w_{0.99} = 2.33$ nel primo test rifiutiamo l'ipotesi H_0 mentre nel secondo la accettiamo.

ESEMPIO 10.6. Il fabbricante di una certa medicina assicura che è efficace nel 90% dei casi nell'alleviare i sintomi di una certa allergia. In un campione di 200 individui che soffrono per questa allergia, la medicina si è dimostrata utile per 160. Sottoporre a test l'affermazione del fabbricante.

In questo caso poniamo $H_0 : E[X] \geq 0.90$ e l'affermazione è legittima, mentre $H_1 < 0.90$ e l'affermazione è falsa.