

# Lezioni di matematica discreta e probabilità

Fabrizio Caselli



# Contents

Introduzione al corso	5
Chapter 1. Calcolo Combinatorio	7
1. Cardinalità finite, numerabile e del continuo	7
2. Costruzioni elementari di insiemi	8
3. Metodi per contare	9
4. Il principio di inclusione-esclusione	15
5. Partizioni di un insieme	18
Chapter 2. Statistica descrittiva	21
1. Popolazione, campione e caratteri	21
2. Classi e istogrammi	21
3. Indici di posizione o centralità e di dispersione	22
4. Correlazione	26
5. Altri indici di posizione	29
6. Esercizi riepilogativi	30
Chapter 3. Spazi di probabilità	33
1. Fenomeni deterministici e casuali	33
2. Spazi di probabilità	33
3. La probabilità condizionale.	36
Chapter 4. Variabili aleatorie discrete	41
1. Variabili discrete	41
2. Densità congiunte e marginali	47
3. Variabili aleatorie indipendenti	49
4. Cambi di variabile	51
5. Calcoli con densità	51
6. La media o speranza matematica	53
7. Varianza	57
8. La legge dei grandi numeri	59
Chapter 5. Variabili aleatorie continue	63
1. Funzione di ripartizione e densità	63
2. La media e la varianza	64
3. Densità uniforme	65
4. Densità esponenziale	65
5. Cambio di variabili	66
6. Densità normali	68
7. Il teorema limite centrale	70

Chapter 6. Statistica inferenziale	73
1. Stime e intervalli di confidenza	73
2. Test statistici	75

## 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.

Informazioni su:

- Contatti: durante il corso mi trovate in ufficio (stanza 4146) a Cesena il mercoledì dalle 08.15 alle 18 circa, esclusi ovviamente gli orari di lezione e una pausa per il pranzo. Potete venire liberamente in qualunque orario, 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)
- Esami: l'esame è costituito da uno scritto e da un orale.

Nello scritto viene chiesto di risolvere alcuni esercizi (tipicamente 3) ed eventualmente di rispondere a una domanda di carattere teorico. Gli esercizi vengono valutati in trentesimi e si è ammessi all'orale se il risultato ottenuto è di almeno 15/30. Sulla domanda di carattere teorico viene espresso un giudizio che verrà utilizzato nella valutazione complessiva.

Nell'orale, di circa 10/15 minuti, verrà chiesto di descrivere uno o due aspetti teorici trattati durante il corso. L'orale può essere effettuato nello stesso appello dello scritto oppure nell'appello successivo. 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.

- Orario lezioni: le lezioni si svolgono MER 09-12 e GIO 15-17 in Aula 3.7.
- Libri consigliati: “Barnabei e Bonetti, Matematica Discreta Elementare, Ed. Pitagora”, per la parte di matematica discreta e “Baldi, Introduzione alla probabilità con elementi di statistica, Ed McGraw-Hill”, per la parte di statistica e probabilità.

Prerequisiti del corso: un po' di analisi matematica e di algebra lineare e soprattutto tanto buon senso.

Struttura del corso. (1) Calcolo Combinatorio, 16 ore; (2) Statistica descrittiva, 10 ore; (3) Probabilità, 30 ore; (4) Statistica inferenziale, 4 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

# Calcolo Combinatorio

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.

### 1. Cardinalità finite, numerabile e del continuo

Come già accennato in precedenza, il calcolo combinatorio si occupa principalmente di “contare” quanti elementi contiene un insieme finito. 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 sono  $\mathbb{N}, \mathbb{Z}, \mathbb{Q}, \mathbb{R}, \mathbb{R}^2$ .

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.

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).

ESERCIZIO 1.1. Ogni sottoinsieme infinito di  $\mathbb{N}$  è numerabile.

Per stabilire se un insieme  $A$  è numerabile dobbiamo riuscire ad “elencare” gli elementi di  $A$  dicendo chi è il primo, chi il secondo e così via. Ad esempio possiamo scrivere  $\mathbb{Z} = \{0, +1, -1, +2, -2, \dots\}$  da cui discende che anche  $\mathbb{Z}$  è numerabile. L'insieme  $\mathbb{Q}$  è anche numerabile, infatti,

$$\mathbb{Q} = \{0, 1, -1, \frac{1}{2}, -\frac{1}{2}, 2, -2, \frac{1}{3}, -\frac{1}{3}, \frac{2}{3}, -\frac{2}{3}, \frac{3}{2}, -\frac{3}{2}, 3, -3, \dots\}.$$

Gli elementi di  $\mathbb{Q}$  li abbiamo elencati in questo modo: abbiamo prima scritto  $0, +1, -1$  poi abbiamo scritto tutti i numeri razionali che si possono ottenere utilizzando il numero 2 e eventualmente numeri minori di 2 (al numeratore e al denominatore), poi quelli che utilizzano il 3 ed eventualmente numeri minori di 3 e così via. È chiaro che in questo modo possiamo elencare tutti i numeri razionali. Chiaramente questo non è l'unico modo per farlo e anzi ce ne sono infiniti.

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

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 1.2. *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

$$B = \{b_1, b_2, b_3, \dots\},$$

dove quindi le  $b_i$  sono *tutte* le sequenze binarie.

Costruiamo a questo punto la sequenza binaria  $b$  nel seguente modo: la prima coordinata di  $b$  è diversa dalla prima coordinata di  $b_1$ . La seconda coordinata di  $b$  è diversa dalla seconda coordinata di  $b_2$ . La terza coordinata di  $b$  è diversa dalla terza coordinata di  $b_3$  e così via. Si ha che la sequenza  $b$  così costruita è diversa dalla sequenza  $b_i$  per ogni  $i$  avendone la coordinata di posto  $i$  diversa. Abbiamo quindi una contraddizione.  $\square$

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 ad esso ha la cardinalità del continuo. In questo caso si può scrivere  $|\mathbb{R}| = \aleph_1$ . 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 1.3 (per appassionati). Dimostrare che  $\mathbb{R}^2$  ha la stessa cardinalità di  $\mathbb{R}$ .

## 2. Costruzioni elementari di insiemi

Ricordiamo ora alcune costruzioni fondamentali che permettono di costruire nuovi insiemi partendo da insiemi dati.

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

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

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  $A^n$  anziché  $A \times \dots \times A$  ( $n$  volte).

Una *lista finita* di lunghezza  $n$  di elementi di  $A$  è un elemento  $(a_1, a_2, \dots, a_n)$  del prodotto cartesiano  $A^n$ . Talvolta una lista  $(a_1, \dots, a_n)$  di lunghezza  $n$  può essere vista come una funzione

$$a : \{1, 2, \dots, n\} \rightarrow A.$$

in cui  $a_1$  è l'immagine di 1,  $a_2$  è l'immagine di 2 e così via.

Una lista infinita, o *successione*, in  $A$  è semplicemente una funzione  $f : \mathbb{N} \rightarrow A$  e può quindi essere rappresentata tramite  $a_1, a_2, \dots$ , dove i puntini di sospensione indicano infiniti elementi.

Vediamo alcuni esempi di liste:

- una sequenza di 5 quadratini colorati con rosso, blu e nero rappresenta una lista di lunghezza 5 nell'insieme  $A = \{\text{rosso}, \text{blu}, \text{nero}\}$ ;
- le parole di 5 lettere possono essere interpretate come liste di lunghezza 5 nell'insieme di tutte le lettere.
- le sequenze binarie di una data lunghezza sono liste in  $\{0, 1\}$ . In particolare i byte sono liste di lunghezza 8 in  $\{0, 1\}$ .



**L'insieme delle parti.** 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.

**PROPOSIZIONE 2.1.** *Sia  $A$  un insieme finito,  $|A| = n$ . Esiste una corrispondenza biunivoca tra  $\mathcal{P}(A)$  e le liste in  $\{0, 1\}$  di lunghezza  $n$ .*

**DIMOSTRAZIONE.** Chiamiamo  $a_1, \dots, a_n$  gli elementi di  $A$ . Consideriamo la seguente funzione

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

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

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

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

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

Si ha chiaramente che  $F$  e  $G$  sono una l'inversa dell'altra e quindi sono delle corrispondenza biunivoche. Se per te questo non è chiaro prova a farti degli esempi. Se ancora non riesci chiedi aiuto a qualcuno. □

**ESEMPIO 2.2.** Se  $A = \{1, 2, 3\}$  l'applicazione  $F$  descritta nella precedente proposizione agisce nel modo seguente

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

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

### 3. Metodi per contare

**Induzione.** Uno strumento fondamentale nel calcolo combinatorio (e non solo) è il principio di induzione che avete senz'altro già visto e che ora ricordiamo:

*Principio di induzione.* Sia  $A \subseteq \mathbb{N}$  un sottoinsieme che soddisfa le seguenti proprietà:

- $0 \in A$  (passo base);
- se  $n \in A$  allora  $n + 1 \in A$  (passo induttivo).

Allora  $A = \mathbb{N}$ .

Il principio di induzione è utile (e a volte necessario) per dimostrare alcune formule del tipo

$$1 + 2 + \dots + n = \frac{n(n+1)}{2}$$

o

$$1 + 3 + 5 + \dots + (2n-1) = n^2,$$

ma come vedremo lo sarà anche per calcolare la cardinalità di alcuni insiemi finiti. Vediamo un esempio di applicazione: sia  $A \subseteq \mathbb{N}$  dato da

$$A = \{n \in \mathbb{N} : 1 + 2 + \cdots + n = \frac{n(n+1)}{2}\}.$$

Vogliamo mostrare che  $A = \mathbb{N}$  e per fare ciò utilizziamo il principio di induzione. Osserviamo che  $0 \in A$ : infatti la somma dei numeri da 1 a 0, essendo vuota, è 0 e infatti  $0(0+1)/2 = 0$ . Vediamo il passo induttivo e quindi supponiamo che  $n \in A$ , cioè che  $1 + 2 + \cdots + n = \frac{n(n+1)}{2}$ . Allora

$$1 + 2 + \cdots + (n+1) = 1 + 2 + \cdots + n + (n+1) = \frac{n(n+1)}{2} + (n+1) = \frac{(n+1)(n+2)}{2}$$

e quindi  $n+1 \in A$ . Il passo induttivo è così completato e quindi  $A = \mathbb{N}$ .

**Principi fondamentali.** 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 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 i risultati possibili nel lancio di due dadi sono in corrispondenza biunivoca con  $\{1, 2, 3, 4, 5, 6\}^2$  e quindi questi risultati sono  $6^2 = 36$ .

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

**Prodotti condizionati.** 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  $C \subseteq A_1 \times A_2 \times \cdots \times A_k$ . 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  $C$ .

Possiamo dimostrare, ad esempio per induzione, che un prodotto condizionato di tipo  $n_1, \dots, n_k$  ha esattamente  $n_1 \cdot n_2 \cdots n_k$  elementi. Un esempio è costituito dalle liste di 5 cifre decimali aventi tutte le cifre distinte. Un altro ancora dalle liste 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 3.1. 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.

**Contare le liste e i sottoinsiemi.** Possiamo a questo punto iniziare a contare quanti elementi contengono alcuni degli insiemi introdotti nella sezione precedente. Le liste di lunghezza  $m$  in un insieme finito  $A$  sono  $|A|^m$ . Questo segue direttamente dal principio di moltiplicazione.

L'insieme delle parti  $\mathcal{P}(A)$  di un insieme  $A$  contiene  $2^{|A|}$  elementi: abbiamo infatti visto che l'insieme delle parti di  $A$  è in biiezione con le liste binarie di lunghezza  $|A|$  (principio di uguaglianza).

ESEMPIO 3.2. 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.

**Disposizioni e permutazioni.** Una lista  $(a_1, \dots, a_k)$  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 3.3. *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 3.4. Le disposizioni di lunghezza 3 nell'insieme  $\{a, b, c, d, e\}$  sono  $5 \cdot 4 \cdot 3 = 60$ . Alcune di esse sono  $(a, b, d)$ ,  $(c, d, e)$  ... Osserviamo che nelle disposizioni, essendo liste, 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).

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 3.5. 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 più avanti il problema di contare il numero di anagrammi di una parola qualunque.

**Combinazioni e sottoinsiemi di data cardinalità.** Abbiamo già visto che un insieme di  $n$  elementi contiene  $2^n$  sottoinsiemi. Come facciamo invece a contare quanti sono i sottoinsiemi che hanno un certo numero prefissato, diciamo  $k$ , di elementi? Possiamo utilizzare i risultati appena ottenuti per rispondere a questa domanda.

PROPOSIZIONE 3.6. *Un insieme di  $n$  elementi ha esattamente*

$$\frac{(n)_k}{k!} = \frac{n!}{(n-k)!k!}$$

*sottoinsiemi di cardinalità  $k$ .*

DIMOSTRAZIONE. Osserviamo che ogni disposizione di lunghezza  $k$  individua un sottoinsieme con  $k$  elementi. Una qualunque permutazione degli elementi di questa disposizione individua lo stesso sottoinsieme e abbiamo quindi che ogni sottoinsieme è individuato da esattamente  $k!$  disposizioni. Concludiamo quindi che il numero di disposizioni è pari a  $k!$  volte il numero di sottoinsiemi e il risultato segue.  $\square$

I sottoinsiemi di  $A$  con  $k$  elementi si chiamano anche *combinazioni* di lunghezza  $k$  in  $A$ . Ad 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.

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:

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

La scelta di un sottoinsieme con  $k$  elementi comporta automaticamente la scelta di un sottoinsieme con  $n-k$  elementi (il suo complementare). Possiamo quindi vedere una combinazione di lunghezza  $k$  in un insieme di  $n$  elementi come una suddivisione dell'insieme in due sottoinsiemi, il primo di cardinalità  $k$  e il secondo di cardinalità  $n-k$ . Chiameremo quindi una combinazione di lunghezza  $k$  anche una combinazione di tipo  $(k, n-k)$ . Possiamo generalizzare questo concetto suddividendo il nostro insieme in tre o più sottoinsiemi di cardinalità  $a, b, c$  rispettivamente, con  $a + b + c = n$ .

DEFINIZIONE. Sia  $S$  un insieme con  $|S| = n$  e  $a, b, c \in \mathbb{N}$  tali che  $a + b + c = n$ . Una combinazione di tipo  $(a, b, c)$  in  $S$  è una terna (ordinata) di sottoinsiemi a due a due disgiunti  $A, B, C$  tali che  $|A| = a$ ,  $|B| = b$ ,  $|C| = c$ .

Chiaramente ci poniamo il problema di contare quante sono le combinazioni di tipo  $(a, b, c)$ . I prodotti condizionati vengono ancora in nostro aiuto.

LEMMA 3.7. *Sia  $S$  un insieme con  $|S| = n$ . L'insieme delle combinazioni di tipo  $a, b, c$  in  $S$  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. 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.  $\square$

Come ci si può aspettare, vale una formula analoga se vogliamo suddividere l'insieme in più di 3 sottoinsiemi. Le combinazioni di tipo  $(a, b, c)$  prendono anche il nome di coefficienti trinomiali in quanto 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.

**PROPOSIZIONE 3.8.** *Consideriamo una parola in cui compare  $a$  volte la lettera  $x$ ,  $b$  volte la lettera  $y$  e  $c$  volte la lettera  $z$ . Allora gli anagrammi di questa parola sono*

$$\binom{a+b+c}{a, b, c}.$$

**DIMOSTRAZIONE.** Ogni anagramma corrisponde ad una combinazione di  $\{1, 2, \dots, n\}$ , dove  $n = a+b+c$ , di tipo  $(a, b, c)$ . La combinazione associata  $(A, B, C)$  si ottiene prendendo come  $A$  i posti dove compare la  $x$ ,  $B$  i posti dove compare la  $y$ ,  $C$  i posti dove compare la  $z$ . Ad esempio se l'anagramma è  $xyxz$  abbiamo  $A = \{1, 3\}$ ,  $B = \{2\}$ , e  $C = \{4\}$ . Questa corrispondenza è chiaramente biunivoca e il risultato segue.  $\square$

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

**PROPOSIZIONE 3.9 (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.  $\square$

Abbiamo anche una “formula di Stifel” per le combinazioni di tipo  $a, b, c$ . Si ha infatti:

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

per ogni  $a, b, c > 0$  tali che  $a + b + c = n$ .

Osserviamo anche la seguente identità

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

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.

**ESERCIZIO 3.10.** Calcolare il numero di permutazioni di 8 che hanno numeri pari nelle posizioni pari.

**ESERCIZIO 3.11.** Per ogni  $k = 0, 1, 2, 3, 4, 5$  stabilire quante sono le matrici  $5 \times 5$  a coefficienti in  $0, 1$  che hanno esattamente  $k$  righe nulle.

**ESERCIZIO 3.12.** Quante sono le liste decrescenti di lunghezza  $k$  nell'insieme  $\{1, 2, \dots, n\}$ ?

**ESEMPIO 3.13.** 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 bisogna effettuare  $m$  passi verso

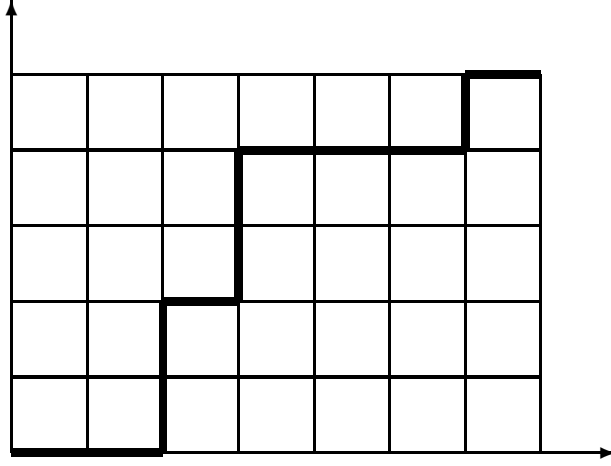


FIGURE 1. Esempio di cammino reticolare

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. Ogni cammino resta quindi individuato dalla scelta delle  $m$  posizioni (tra le  $m + n$  disponibili) in cui inserire gli 0 e quindi tali cammini sono esattamente

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

**I numeri di Fibonacci.** Ci chiediamo quanti sono i sottoinsiemi di  $\{1, 2, \dots, n\}$  che non contengono 2 interi consecutivi. Per rispondere ricordiamo intanto la corrispondenza biunivoca tra l'insieme delle parti di  $\{1, 2, \dots, n\}$  e le liste 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 domanda può essere riformulata chiedendoci quante sono le liste binarie di lunghezza  $n$  in cui non compaiono due 1 consecutivi. Chiamiamo  $F_n$  tale numero. 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). Possiamo dare una formula esplicita per calcolarli.

PROPOSIZIONE 3.14. *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 3.15. *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  $H_n = F_{n-2}$ .  $\square$

#### 4. Il principio di inclusione-esclusione

Vogliamo ora determinare la cardinalità dell'unione di due o più insiemi nel caso siano note le cardinalità delle intersezioni tra alcuni 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 4.1. 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

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

da cui 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

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

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. 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

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

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

**ESERCIZIO 4.2.** Quanti sono i numeri compresi tra 1 e 100 che sono multipli di 2, 3 o 5?

**ESEMPIO 4.3.** 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

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

**ESERCIZIO 4.4.** Determinare quanti sono i numeri compresi tra 1 e 50 che non sono multipli né di 2, né di 3, né di 5.

**Funzioni iniettive e suriettive.** Siano  $A$  e  $B$  due insiemi finiti di cardinalità  $m$  e  $n$  rispettivamente. Ci chiediamo quante sono le funzioni iniettive e quante quelle suriettive da  $A$  a  $B$ .

Le funzioni iniettive esistono se e solo se  $m \leq n$ . Osserviamo che se chiamiamo  $a_1, \dots, a_m$  gli elementi di  $A$

$$A = \{a_1, \dots, a_m\}$$

abbiamo che ogni funzione iniettiva  $f$  definisce una disposizione in  $B$  di lunghezza  $m$  data da

$$(f(a_1), \dots, f(a_m)).$$

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, e in tal caso abbiamo bisogno di utilizzare il principio di inclusione esclusione in modo opportuno. Osserviamo intanto che abbiamo necessariamente in questo caso  $m \geq n$ .



Poniamo  $S$  l'insieme delle funzioni suriettive e, denotando gli elementi di  $B$  tramite  $B = \{b_1, \dots, b_n\}$ , per ogni  $i = 1, \dots, n$  poniamo

$$A_i = \{f : A \rightarrow B : b_i \notin f(A)\}.$$

Abbiamo chiaramente che  $A_1 \cup \dots \cup A_n$  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

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

in quanto  $\bigcap_{i \in I} A_i$  sono esattamente le funzioni  $f : A \rightarrow B$  le cui immagini non contengono gli elementi  $b_i$ ,  $i \in I$ . 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  $\binom{n}{i}$ .

**Scombussolamenti.** Uno scombussolamento è una permutazione  $f$  dell'insieme  $\{1, 2, \dots, n\}$  in cui nessun elemento rimane al proprio posto. Ad esempio 4213 non è uno scombussolamento perché il numero 2 è al posto 2, mentre 4123 è uno scombussolamento. Per  $n = 3$  abbiamo solo due scombussolamenti, cioè 312 e 231. Ma quanti sono in generale? Il principio di inclusione-esclusione ci permette ancora una volta di rispondere a questa domanda.

Per semplificare la trattazione prendiamo  $A = \{1, 2, \dots, n\}$ . Chiamiamo  $S$  l'insieme degli scombussolamenti 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 due termini in questa sommatoria ottenendo

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

ESEMPIO 4.5. In quanti modi 6 persone si possono scambiare il proprio cappello in modo che nessuno riceva il proprio cappello? La risposta è data dagli scombussolamenti su un insieme di 6 elementi e sono quindi

$$\sum_{i=0}^6 (-1)^i (6)_{6-i} = (6)_6 - (6)_5 + (6)_4 - (6)_3 + (6)_2 - (6)_1 + (6)_0 = 720 - 720 + 360 - 120 + 30 - 6 + 1 = 265.$$

ESERCIZIO 4.6. Quanti sono i numeri compresi tra 1 e  $10^6$  che non sono divisibili né per 2, né per 3, né per 5, né per 7?

ESERCIZIO 4.7. Tra le liste di lunghezza 10 nell'insieme  $\{1, 2, 3, 4, 5\}$ , quante sono quelle in cui ciascun numero tra 1 e 5 compare almeno una volta?

ESERCIZIO 4.8. In quanti modi si possono distribuire  $n$  persone in  $k$  stanze? E in quanti modi se si richiede che nessuna stanza rimanga vuota?

ESERCIZIO 4.9 (Difficile). Sia  $D_n$  il numero di scombussolamenti su un insieme con  $n$  elementi. Mostrare che per ogni  $n > 1$  si ha  $D_n = (n-1)(D_{n-1} + D_{n-2})$ .

## 5. 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. 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* della partizione.

ESEMPIO 5.1. 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 5.2. Per ogni numero primo  $p$  poniamo  $A_p = \{n \geq 2 : p \text{ è il più grande numero primo che divide } n\}$ . 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 parti, ognuna delle quali contiene un numero infinito di elementi.

ESERCIZIO 5.3. Sia  $A = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$ . E consideriamo i seguenti sottoinsiemi di  $A$ :  $A_1 = \{1, 2, 3, 4\}$ ,  $A_2 = \{5, 6, 7\}$ ,  $A_3 = \{4, 5, 7, 9\}$ ,  $A_4 = \{4, 8, 10\}$ ,  $A_5 = \{8, 9, 10\}$ ,  $A_6 = \{1, 2, 3, 6, 8, 10\}$ . Quante e quali sono le partizioni di  $A$  che si possono ottenere utilizzando questi sottoinsiemi?

**I numeri di Bell.** È evidente che se  $A$  è infinito ammetterà infinite partizioni. Ci vogliamo quindi concentrare sul caso finito.

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

Il numero di partizioni di  $\{1, 2, \dots, n\}$  è detto numero di Bell e si indica con  $B_n$ . Abbiamo già calcolato  $B_1 = 1$ ,  $B_2 = 2$  e  $B_3 = 5$ . Vogliamo ora determinare una formula ricorsiva per calcolare  $B_n$ .

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.

**I numeri di Stirling.** Possiamo raffinare il calcolo del numero di partizioni di un insieme finito se vogliamo tener conto anche del numero di sottoinsiemi coinvolti. Chiamiamo numero di Stirling di seconda specie il numero  $S_{n,k}$  di partizioni di  $\{1, 2, \dots, n\}$  costituite da esattamente  $k$  sottoinsiemi. Per convenzione si pone  $S_{0,0} = 1$  e  $S_{0,k} = 0$  per ogni  $k > 0$ .

Vogliamo innanzitutto stabilire un legame tra i numeri di Stirling e il numero di funzioni suriettive tra due insiemi. Siano quindi  $A, B$  insiemi finiti, con  $|A| = n$  e  $|B| = k$ . È chiaro che 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 5.4.** *Per ogni  $n, k > 1$  si ha*

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

**DIMOSTRAZIONE.** Ancora una volta facciamo ricorso al principio di somma. Consideriamo le partizioni in  $k$  parti di cui una è  $\{n\}$  e le partizioni in  $k$  parti 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  $k - 1$  sottoinsiemi). Quelli del secondo tipo si ottengono considerando una qualunque partizione di  $\{1, 2, \dots, n - 1\}$  in  $k$  parti per poi aggiungere  $n$  ad una di queste  $k$  parti. Otteniamo quindi  $k S_{n-1,k}$  partizioni.  $\square$



## CHAPTER 2

# Statistica descrittiva

### 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 è una funzione che ha come dominio la popolazione.

ESEMPIO 1.1. 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 1.2. 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 1.3. 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.

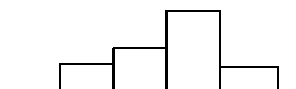
ESEMPIO 2.1. Voto conseguito alla maturità dagli iscritti a ISI. Non sarà conveniente considerare tutti i possibili voti da 60 a 100, ma sarà più opportuno raggrupparli in classi, ad esempio 60-69, 70-79, ..., 90-100

ESEMPIO 2.2. 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.

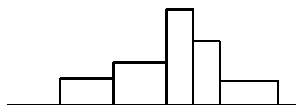
L'*istogramma* è un grafico che permette di visualizzare alcune informazioni relative ad un carattere. Ad ogni modalità viene associato un rettangolo avente *area* proporzionale alla frequenza della modalità stessa. Se le modalità sono state raggruppate in classi si richiede che la base del rettangolo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

ESEMPIO 2.3. Studenti e voto di maturità. 60-69 5, 70-79 8, 80-89 15, 90-100 5



Notare che l'ultimo rettangolo è un po' più largo e un po' più basso del primo, anche se rappresentano classi aventi la medesima frequenza.

Gli stessi dati potrebbero essere raggruppati in classi in modo diverso. Ad esempio potremmo decidere di essere più precisi dove si presentano più elementi del campione, cioè nella classe 80-89 e considerare la seguente ripartizione in classi: Studenti e voto di maturità. 60-69 5, 70-79 8, 80-84 9 85-89 6, 90-100 5 In questo caso l'istogramma diventa



### 3. Indici di posizione o centralità e di dispersione

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.

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

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

**La media campionaria.** Consideriamo un carattere quantitativo  $x$  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

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

Calcolo della media campionaria nell'Esempio 1.3.

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 3.1.** *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 3.2. 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  $x_1 = 50$ ,  $x_2 = 55$ ,...da cui

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

e quindi  $\bar{y} = (49 - 32)\frac{100}{180} = 9.44^\circ\text{C}$ . 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 3.3. 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

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

Caso dell'unione di 2 campioni. Consideriamo due campioni costituiti da  $n$  e  $m$  elementi ciascuno. Consideriamo 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 3.4. 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  $\frac{14+7}{21} = 1$ .

**La media ponderata.** La media ponderata si utilizza se si vuole assegnare un 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:

$$\bar{x}_w = \frac{w_1 x_1 + \dots + w_n x_n}{w_1 + \dots + w_n}.$$

ESEMPIO 3.5. 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

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

**La varianza campionaria e lo scarto quadratico medio campionario.** 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

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

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  $\sigma_x$ . Per il calcolo della varianza campionaria può essere conveniente utilizzare la seguente formula

PROPOSIZIONE 3.6. *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 nell'intervallo  $[\bar{x} - 3\sigma_x, \bar{x} + 3\sigma_x]$  centrato nella media campionaria  $\bar{x}$  e di semiampiezza  $3\sigma_x$  si trovi la maggior parte dei valori registrati.

ESERCIZIO 3.7. 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 lo scarto quadratico medio è circa 19 e che in effetti i dati si trovano all'interno dell'intervallo  $[\bar{x} - 3\sigma_x, \bar{x} + 3\sigma_x] = [35, 149]$ .

Cosa succede alla varianza campionaria per i cambi di variabile lineari  $y = ax + b$ ?

PROPOSIZIONE 3.8. *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 3.6 per calcolare la varianza di  $y$  e le proprietà della media della Proposizione 3.1. Abbiamo:

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

□

#### 4. Correlazione

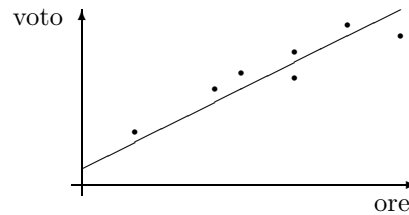
In questa sezione studiamo il legame che ci può essere tra due caratteri (quantitativi) definiti sullo stesso campione. Studieremo in particolare quando esiste una sorta di proporzionalità diretta (correlazione positiva) o inversa (correlazione negativa). Studiamo questo problema tenendo a mente il seguente 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
10	10
60	28
50	30
40	25
40	20
30	21
25	18

ore <sup>2</sup>	voto <sup>2</sup>
100	100
3600	784
2500	900
1600	625
1600	400
900	441
625	324

ore · voto
100
1680
1500
1000
800
630
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.

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  $ax_i + b$ .

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

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

Questa 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 4.1. *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  $\pm 1$  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.

ESERCIZIO 4.2.  $x$ =concentrazione di ozono e  $y$ =irraggiamento solare

$x$	$y$	$x^2$	$y^2$	$xy$
41	190	1681	36100	7790
16	256	256	65536	4096
30	322	900	103688	9660
45	252	2025	63504	11340
21	191	441	36481	4011
32	236	1024	55696	7552
10	264	100	69696	2640

Determinare la retta ai minimi quadrati. Che valore dell'irraggiamento dobbiamo aspettarci in un giorno in cui si registra una concentrazione di ozono pari a 20?

## 5. Altri indici di posizione

**La mediana e gli altri quantili.** 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. Un 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.

**Altri tipi di medie.** E' 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? La variazione complessiva è del -6,16% che corrisponde ad una variazione media del -1,58%. La media da utilizzare in questo caso è la geometrica  $\bar{x}_{\log} = (x_1 \cdots x_n)^{1/n}$  tra le percentuali espresse come fattore moltiplicativo, 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. Bisogna fare la media quadratica  $\sqrt{\frac{1}{n}(x_1^2 + \cdots + x_n^2)} = \sqrt{\frac{1+4+9}{3}} = 2,16$ .
- Passeggiata in montagna. All'andata si va a 3 km/h. Al ritorno a 7 km/h. Qual è la velocità media? Si fa la media armonica  $\left(\frac{x_1^{-1} + \cdots + x_n^{-1}}{n}\right)^{-1}$  che in questo caso fa 4,2 km/h.

## 6. Esercizi riepilogativi

ESERCIZIO 6.1. Si consideri l'esercizio 3.7. Determinare la mediana e lo scarto interquartile.

ESERCIZIO 6.2. In 25 scatole di 100 viti si sono contati il numero di pezzi difettosi ottenendo i seguenti dati

1, 4, 3, 1, 3, 2, 2, 1, 2, 5, 3, 0, 1, 4, 3, 7, 1, 3, 1, 7, 2, 1, 2, 4, 8.

Determinare le modalità, le frequenze, la media (2.84), i quartili (1,2,4), la moda, la varianza campionaria (4.21), la differenza interquartile e il range.

ESERCIZIO 6.3. Sono state recensite 512 famiglie di 6 figli. Per ognuna è stato osservato il numero di figlie femmine. I dati sono riportati nella tabella seguente.

Numero di figlie	frequenza
0	23
1	64
2	131
3	123
4	107
5	48
6	16

Qual è il numero medio di figlie? Calcolare la varianza con entrambe le formule.

Si ha  $\bar{x} = \frac{1459}{512} = 2,84906$ . Per la varianza abbiamo

$$\begin{aligned}\sigma_x^2 &= \frac{23(2.85)^2 + 64(1.85)^2 + 131(0.85)^2 + 123(0,15)^2 + 107(1,15)^2 + 48(2,15)^2 + 16(3,15)^2}{512} \\ &= \frac{1025,42}{512} = 2,0028\end{aligned}$$

Calcoliamo la media dei quadrati  $\overline{x^2} = \frac{5183}{512} = 10,123$ . E quindi

$$\sigma_x^2 = \frac{5183}{512} - \left(\frac{1459}{512}\right)^2 = \frac{5183 \cdot 512 - 1459^2}{512^2} = \frac{525015}{262144} = 2,0028.$$

ESERCIZIO 6.4. Il peso di 50 giocatori di football americano è rappresentato in questa tabella.

Peso	Frequenza
98-100	2
100-102	4
102-104	8
104-106	8
106-108	18
108-110	8
110-112	0
112-114	2

Determinare la media (105.8), la varianza (9.28) e la mediana (106.3) dei dati.

ESERCIZIO 6.5. Abbiamo 3 statuette in rame della Tour Eiffel di altezza rispettivamente 5, 10 e 20 centimetri. Decidiamo di fonderle e di ricavarne 3 statuette uguali. Quanto bisogna farle alte se si vuole utilizzare tutto il rame a disposizione?

Bisogna fare in questo caso la media cubica  $\bar{x}_f$  dove  $f(x) = x^3$ . Si ha quindi

$$\bar{x}_f = \left(\frac{5^3 + 10^3 + 20^3}{3}\right)^{\frac{1}{3}} = 14,49.$$





## CHAPTER 3

# Spazi di probabilità

### 1. Fenomeni deterministici e casuali

Un fenomeno è un qualunque accadimento che porta ad un certo *risultato*, 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  $\Omega$ . Nei casi precedenti abbiamo  $\Omega = \{1, 2, \dots, 6\}$ ,  $\{(0, 0), (0, 1), (1, 0), \dots\}$ ,  $\{2017, 2018, \dots\}$ ,  $[50, 200]$ .

Un *evento* è un sottoinsieme di  $\Omega$ . In altre parole, un evento è un insieme di possibili risultati di un fenomeno aleatorio. Negli esempi precedenti possiamo avere  $E = \{2, 4, 6\}$ ,  $\{(x, y) : x > y\}$ ,  $\{x : x > 2020\}$ ,  $\{x : x < 60\}$ . In realtà vedremo più avanti che solo *alcuni* sottoinsiemi di  $\Omega$  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 tutte 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.

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. Se lo spazio dei risultati è continuo vogliamo anche un criterio sensato per stabilire quali sono gli eventi da considerare. Può essere senz'altro complicato (e per certi versi inutile e per altri impossibile) considerare come eventi tutti i possibili sottoinsiemi dell'insieme dei risultati possibili.

### 2. Spazi di probabilità

Sia dato un insieme  $\Omega$  (di possibili risultati di un fenomeno aleatorio). Gli eventi saranno per noi i sottoinsiemi di  $\Omega$  di cui vorremo e potremo calcolare le probabilità. Pertanto, NON HA SENSO calcolare la probabilità di un sottoinsieme di  $\Omega$  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 terminologia. 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  $\Omega$ . Diciamo che  $\mathcal{A}$  è una *famiglia coerente di eventi* su  $\Omega$  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  $\mathcal{A}$ .

**ESEMPIO 2.1.** 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  $\sigma$ -algebra. Avremmo potuto scegliere anche solo unione e intersezioni finite anziché numerabili, ma si preferisce utilizzare questo tipo di definizione.

Data una  $\sigma$ -algebra  $\mathcal{A}$  una 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) non viene richiesta nel caso di unioni 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  $(\Omega, \mathcal{A}, P)$ , dove  $\Omega$  è un insieme,  $\mathcal{A}$  è una famiglia coerente di eventi su  $\Omega$  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.

**La probabilità uniforme.** Prendiamo il caso in cui  $\Omega$  è un insieme finito ed i risultati sono tutti equiprobabili. In questo caso si sceglie  $\mathcal{A}$  come l'insieme di tutti i possibili sottoinsiemi di  $\Omega$ . E si deduce facilmente che  $P(A) = \frac{|A|}{|\Omega|}$ . 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 2.2.** Qual è la probabilità di lanciare un dado ed ottenere un numero pari? E qual è la probabilità di indovinare un terno al lotto? Qual è la probabilità che esca un numero fissato al lotto?

Ci chiediamo ora qual è la probabilità che un certo numero esca come secondo estratto al lotto. I risultati possibili come secondo estratto sono i numeri da 1 a 90 e 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  $1/90$ .

ESERCIZIO 2.3. Costruire uno spazio di probabilità (non uniforme) sull'insieme  $\Omega = \{1, 2, a, b\}$ .

ESERCIZIO 2.4. Costruire uno spazio di probabilità sull'insieme  $\mathbb{N}$  dei numeri naturali.

ESEMPIO 2.5. 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”. 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 è 0 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, perché nella modellizzazione uno fa come vuole. Ma se i numeri interi li scrivessimo in questo ordine  $0, 1, 3, 2, 6, 4, 9, 5, 12, 7, 15, 8, 18, 10, 21, 11, 24, 13, 27, 14, \dots$ ? 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  $\sigma$ -algebra si registrano se  $\Omega$  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  $\sigma$ -algebra troppo complessa da poter essere esaminata. Ma ci torneremo più in là.

**Prime proprietà.** 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

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

ESEMPIO 2.6. Abbiamo una scatola con due palline rosse e una nera e una scatola con due palline nere e due rosse. Estraendo 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  $\Omega$  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

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

Questo è un esempio della seguente costruzione che vale molto più in generale. Sia  $\{A_1, A_2, \dots\}$  una partizione (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  $\Omega$ .

Sia quindi  $A_1, A_2, \dots$  una partizione in eventi di  $\Omega$ ; 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

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

Vediamone un esempio. 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.

### 3. La probabilità condizionale.

Dato uno spazio  $(\Omega, \mathcal{A}, P)$  e fissato un evento  $A$  tale che  $P(A) > 0$  formiamo un nuovo spazio di probabilità  $(\Omega, \mathcal{A}, P')$  ponendo

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

Verifichiamo che  $P'$  è effettivamente una probabilità.

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

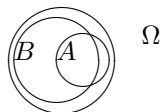
$$P'(B_1 \cup B_2 \cup \dots) = \frac{P(A \cap (B_1 \cup B_2 \cup \dots))}{P(A)} = \frac{P((A \cap B_1) \cup (A \cap B_2) \cup \dots)}{P(A)} = P'(B_1) + P'(B_2) + \dots.$$

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

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 2.6. 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.

Vediamo l'Esempio 2.9 del libro. Si ha  $P(M) = P(M \cap F) + P(M \cap N)$ .

**La formula delle probabilità totali.** Se  $A_1, \dots, A_n$  è una partizione in eventi di  $\Omega$  si ha per ogni evento  $B$

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

Questa formula, che abbiamo già implicitamente utilizzato per risolvere il problema dell'esempio 2.6 e nell'esempio 2.9 del libro, si ottiene facilmente dalla eq. (3) insieme alla definizione di probabilità condizionale.

**ESERCIZIO 3.1.** 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  $\Omega$  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  $\Omega$  per cui, usando la formula delle probabilità totali otteniamo

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

**La formula di Bayes.** 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 3.2.** Qual è la probabilità che un malato scelto a caso nella regione dell'esercizio 3.1 sia un fumatore? Dobbiamo calcolare  $P(F|M)$  e per farlo utilizziamo la formula di Bayes

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

**ESEMPIO 3.3.** 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  $\Omega$  l'insieme di tutte le  $n$ -ple ordinate di palline distinte scelte tra le  $b + r$  nell'urna (tutte le possibili estrazioni). Su  $\Omega$  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  $P(B_1) = \frac{b}{b+r}$ . 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 3.4.** 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?

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 3.5.** 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?

**ESEMPIO 3.6.** 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 3.7.** 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

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

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

**Eventi indipendenti.** Due eventi  $A$  e  $B$  si dicono indipendenti se  $P(A \cap B) = P(A) \cdot P(B)$ . 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  $P(B) \neq 0$ ). Detto in altri termini 2 eventi si dicono indipendenti se il verificarsi di uno dei due non modifica in alcun modo la probabilità che accada l'altro. Questo è fondamentale nella fase di modellizzazione quando, per motivi evidenti 2 eventi risultano essere indipendenti.

**ESEMPIO 3.8.** Lanciando un dado diciamo che otteniamo un successo se esce un numero  $\geq 5$ . 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

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

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.

**Esercizi.**

ESERCIZIO 3.9. Componenti prodotti da una ditta presentano indipendentemente 2 tipi di difetti con probabilità del 3% e del 7%.

- Qual è la probabilità che un componente presenti entrambi i difetti? Detti  $A$  e  $B$  gli eventi relativi ai 2 difetti si ha

$$P(A \cap B) = P(A)P(B) = \frac{3}{100} \frac{7}{100} = \frac{21}{10000}.$$

- Qual è la probabilità che un componente sia difettoso?

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = \frac{3}{100} + \frac{7}{100} - \frac{21}{10000} = \frac{300 + 700 - 21}{10000} = \frac{979}{10000}.$$

- Qual è la probabilità che presenti il difetto 1 sapendo che è difettoso?

$$P(A|A \cup B) = \frac{P(A)}{P(A \cup B)} = \frac{3}{100} \frac{10000}{979} = \frac{300}{979} = 0,3064.$$

- Qual è la probabilità che presenti solo uno dei 2 difetti sapendo che è difettoso?

$$P((A \setminus B) \cup (B \setminus A)|A \cup B) = 1 - P(A \cap B|A \cup B) = 1 - \frac{P(A \cap B)}{P(A \cup B)} = 1 - \frac{21}{10000} \frac{10000}{979} = \frac{958}{979} = 0.9785.$$

ESERCIZIO 3.10. Dieci urne: la  $i$ -ma ha 4 palline rosse e  $i$  bianche. Vengono estratte 2 palline da un'urna scelta a caso.

- Qual è la probabilità che le palline siano diverse? In questo caso  $\Omega$  è l'insieme delle coppie (non ordinate) di palline appartenenti alla stessa urna. Poniamo  $A_i$  l'evento "coppie di palline dell'urna  $i$ " e poi consideriamo gli eventi  $BB$ ,  $RR$ ,  $BR$  (2 bianche, 2 rosse, 1 rossa e 1 bianca). Si ha

$$\begin{aligned} P(BR) &= \sum_{i=1}^{10} P(A_i)P(BR|A_i) = \sum_{i=1}^{10} \frac{1}{10} \frac{4i}{\binom{4+i}{2}} \\ &= \frac{1}{10} \left( \frac{4}{10} + \frac{8}{15} + \frac{12}{21} + \frac{16}{28} + \frac{20}{36} + \frac{24}{45} + \frac{28}{55} + \frac{32}{66} + \frac{36}{78} + \frac{40}{91} \right) \\ &= 0.5060 \end{aligned}$$

- Sapendo di aver estratto una bianca e una rossa, qual è la probabilità di aver estratto dalla urna  $i$ ? Qual è l'urna più probabile?

$$P(A_i|BR) = P(BR|A_i) \frac{P(A_i)}{P(BR)} = \frac{4i}{\binom{4+i}{2}} \frac{1/10}{0.5060}.$$

Le probabilità di aver estratto dall'urna  $i$  sono quindi date rispettivamente da

$$7.9\% \quad 10.5\% \quad 11.3\% \quad 11.3\% \quad 11.0\% \quad 10.5\% \quad 10.1\% \quad 9.6\% \quad 9.1\% \quad 8.7\%$$

Le 2 urne che hanno la massima probabilità sono la 3 e la 4.

In alternativa, per stabilire quale sia l'urna più probabile, si può valutare la quantità

$$\frac{P(A_i|BR)}{P(A_{i+1}|BR)} = \frac{P(BR|A_i)}{P(BR|A_{i+1})} = \frac{\frac{4i}{\binom{4+i}{2}}}{\frac{4(i+1)}{\binom{5+i}{2}}} = \frac{(5+i)i}{(3+i)(i+1)} = \frac{i^2 + 4i + i}{i^2 + 4i + 3},$$

da cui risulta evidente che il massimo si ottiene per le urne 3 e 4.

ESERCIZIO 3.11. Qual è la probabilità che due studenti scelti a caso in aula festeggino il compleanno lo stesso giorno? Diciamo che in aula ci sono 40 studenti. Allora  $\Omega = S^{40}$ , dove  $S = \{1, 2, \dots, 365\}$ . Calcoliamo la probabilità dell'evento complementare, cioè che siano nati tutti in giorni diversi. Si ha

$$P(E^C) = \frac{D_{40}^{365}}{365^{40}} \cong 0,11.$$

La probabilità che 2 studenti in aula festeggino il compleanno lo stesso giorno è quindi di circa l'89%.

ESERCIZIO 3.12. Nel poker classico si gioca con  $16 + 4n$  carte dove  $n$  è il numero di giocatori. Stabilire per quali valori di  $n$  è più probabile avere un poker che 5 carte dello stesso seme.

Il poker si può fare in  $4 + n$  modi, tutti chiaramente equiprobabili. Si ha quindi

$$\begin{aligned} P(\text{Poker}) &= (4 + n)P(\text{Poker d'assi}) = (4 + n) \frac{12 + 4n}{\binom{16+4n}{5}} \\ &= 5! \frac{(4 + n)(12 + 4n)}{(16 + 4n)(15 + 4n)(14 + 4n)(13 + 4n)(12 + 4n)} = \frac{30}{(4n + 15)(4n + 14)(4n + 13)}. \end{aligned}$$

Vediamo ora la probabilità di avere 5 carte dello stesso seme (colore):

$$P(\text{colore}) = 4P(\text{colore di cuori}) = 4 \frac{\binom{4+n}{5}}{\binom{16+4n}{5}} = \frac{(2 + n)(1 + n)n}{(15 + 4n)(14 + 4n)(13 + 4n)}.$$

Bisogna quindi stabilire per quali valori di  $n$  si ha  $30 \leq \frac{(2+n)(1+n)n}{4}$ . Il membro di destra è monotono crescente in  $n$ . Si ha uguaglianza per  $n = 4$ . Per  $n > 4$  è più probabile il colore, per  $n < 4$  è più probabile il poker.



## CHAPTER 4

### Variabili aleatorie discrete

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 0.1. 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. O mediamente quanto mi devo aspettare di vincere.

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

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

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

ESEMPIO 0.2. Si consideri  $\Omega = \mathbb{N}$  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 (perché?). Mentre la funzione  $X(n) = (-1)^n$  lo è.

Osserviamo che  $\{\omega \in \Omega : X(\omega) > a\}$  è un evento perché complementare del precedente. Abbiamo anche che  $\{\omega \in \Omega : X(\omega) < a\}$  è un evento in quanto

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

e quindi è una unione di una successione di eventi. Ne deduciamo anche che

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

è un evento per ogni  $a \in \mathbb{R}$  (perché?).

Siamo interessati a considerare la funzione

$$A \mapsto P(\{\omega : X(\omega) \in A\}).$$

Questa funzione si dice *legge o distribuzione* della variabile  $X$ . Osserviamo che non è definita per ogni sottoinsieme di  $\mathbb{R}$ . D'ora in avanti considereremo quasi sempre variabili aleatorie in termini della loro legge, 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  $P(\{\omega \in \Omega : X(\omega) \leq t\})$ .

#### 1. Variabili discrete

Una variabile aleatoria  $X$  si dice *discreta* se i valori che assume sono finiti oppure numerabili. Verifichiamo che una funzione  $X : \Omega \rightarrow \mathbb{R}$  che assume una quantità al più numerabile di valori è una variabile aleatoria se e solo se  $\{X = t\}$  è un evento per ogni  $t \in \mathbb{R}$ . Infatti, in questo caso, per ogni  $t \in \mathbb{R}$  siano  $t_1, t_2, \dots$  i valori  $\leq t$  che possono essere assunti da  $X$ : abbiamo

$$\{X \leq t\} = \bigcup_i \{X = t_i\}$$

e quindi, come unione di eventi, è automaticamente un evento.

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

Osserviamo che la funzione  $p_X$  soddisfa le seguenti proprietà.

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

Osserviamo che la somma che appare in (3) è una serie numerica in quanto, per (1) i termini diversi da zero sono al più numerabili. E' inoltre ben definita in quanto i termini sono tutti non negativi grazie alla proprietà (2). Una qualunque funzione  $p$  che soddisfa queste proprietà si dice una *densità discreta*.

Osserviamo anche che per ogni  $A \subset \mathbb{R}$  si ha  $\{X \in A\}$  è un evento. Infatti siano  $a_1, a_2, \dots$  i valori assunti da  $X$  che stanno anche in  $A$ : avremo che  $\{X \in A\} = \cup_i \{X = a_i\}$  è unione numerabile di eventi.

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

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

ESEMPIO 1.1. Nel lancio di un dado sia  $\Omega = \{1, 2, 3, 4, 5, 6\}$  l'insieme dei possibili risultati. Allora la funzione  $X(\omega) = \omega$  è una variabile aleatoria uniforme  $X \sim U(1, 2, 3, 4, 5, 6)$ . Cosa possiamo dire della funzione  $Y(\omega) = \cos(\omega\pi)$ ? Si ha che anche  $Y$  è uniforme, ed in particolare  $Y \sim U(-1, 1)$ .

**1.2. La funzione caratteristica e la densità di Bernoulli.** Se  $A$  è un evento la sua funzione caratteristica  $\chi_A$  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  $\{\chi_A \leq t\}$  è un evento per ogni  $t \in \mathbb{R}$ :

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

La sua densità è data da

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

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  $\chi_A$  è una variabile di Bernoulli

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

### 1.3. La densità binomiale.

ESERCIZIO 1.2. 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  $X$ .

**ESERCIZIO 1.3.** 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  $X$ .

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". Uno schema successo-insuccesso consiste nel ripetere questo fenomeno un certo 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 3.8 del Cap. 2 e nella risoluzione degli Esercizi 1.2 e 1.3 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 dipenda 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

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

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

**1.4. 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  $b$  bianche e  $r$  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.

Lo spazio  $\Omega$  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

$$p(t) = \begin{cases} \frac{\binom{b}{t} \binom{r}{n-t}}{\binom{b+r}{n}} & \text{se } t \in \{0, 1, \dots, \min(n, b)\} \\ 0 & \text{altrimenti.} \end{cases}$$

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

I prossimi due esempi vanno visti in parallelo.

**ESEMPIO 1.4.** 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

**ESEMPIO 1.5.** Un'urna contiene 8 palline rosse e 2 bianche. Ne estraiamo 3 senza rimpiazzo. Qual è la probabilità di estrarne al più una bianca? 0.933

**ESEMPIO 1.6.** 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 è data da

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

ESEMPIO 1.7. 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

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

**1.5. La densità geometrica.** 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  $T \sim \tilde{G}(p)$ . La densità geometrica standard si può vedere come una densità geometrica modificata in cui è stato effettuato uno shift di 1. O più semplicemente 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 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 1.8. 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. Determinare la densità delle tre variabili  $X$ ,  $Y$  e  $Z$ . Determinare inoltre la probabilità di ottenere 6 con il dado rosso prima di ottenere 6 con il dado blu.

ESERCIZIO 1.9. Consideriamo 4 scatole contenenti 2 palline ciascuna indistinguibili al tatto. La scatola  $i$  ha una pallina numerata con 1 e l'altra numerata con  $i$  (quindi nella prima entrambe le palline sono numerate 1). Si estrae una pallina per scatola. Si denoti con  $X$  la variabile data dalla somma delle palline estratte e con  $Y$  la variabile data dal numero di palline estratte con il numero 1. Si determinino le densità di  $X$  e di  $Y$ . Determinare inoltre la probabilità di aver estratto la pallina con il 4 sapendo che  $X = 7$ .

Sia  $X$  una variabile aleatoria che assume solo valori naturali. 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 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 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 che se  $X$  è 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 è una variabile geometrica. 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

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

**1.6. La densità di Poisson.** La densità 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  $\lambda$  è un parametro reale strettamente positivo. Se  $X$  è una variabile aleatoria di Poisson di parametro  $\lambda$  scriviamo  $X \sim P(\lambda)$ . Il fatto che questa sia una densità viene dallo sviluppo di Taylor  $e^x = \sum_{k \geq 0} \frac{x^k}{k!}$ , per  $x = \lambda$ . Ricordando anche l'altro limite notevole

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

per  $n \leftarrow +\infty$ , vogliamo verificare che una variabile di Poisson approssima sotto certe ipotesi una variabile binomiale  $X \sim B(n, \lambda/n)$ . Infatti in tal caso la 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  $np < 10$ .

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

**ESEMPIO 1.10.** 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. È 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.

### Esercizi.

**ESERCIZIO 1.11.** Una certa azienda ha 1000 clienti. La probabilità che un singolo cliente chiami al call-center di questa azienda in un certo giorno è di 0,001. Stabilire la probabilità che in un giorno il call-center riceva non più di due telefonate. Approssimiamo la variabile  $X =$  numero di chiamate al giorno con una variabile di Poisson di parametro  $\lambda = n \cdot p = 1$ . Abbiamo

$$P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2) = e^{-1} \left( \frac{1^0}{0!} + \frac{1^1}{1!} + \frac{1^2}{2!} \right) = \frac{5}{2} e^{-1} = 0,91.$$

Il calcolo è quindi molto semplice che potrebbe anche essere fatto a mano. Che tipo di calcoli avremmo dovuto fare utilizzando la densità binomiale?

ESERCIZIO 1.12. In un'azienda il numero di guasti al mese che si verificano ha una densità di Poisson. La probabilità che in un mese non vi siano guasti è data da  $e^{-1}$ .

(a) Qual è la probabilità che nel prossimo mese si verifichi più di un guasto?

$$P(X > 1) = 1 - P(X \leq 1) = 1 - e^{-1} - e^{-1} = 0,264$$

E esattamente 2?  $P(X = 2) = e^{-1} 1^2 / 2! = 0,184$

(b) Qual è la probabilità che nei prossimi 6 mesi si verifichino almeno 3 guasti? Bisogna adesso considerare una variabile di Poisson di parametro  $\lambda = 6$ . Questo perché possiamo pensare a questo problema come il precedente con il numero di macchinari sestuplicato e quindi anche  $\lambda = np$  sarà sestuplicato. Abbiamo

$$P(Y \geq 3) = 1 - P(Y \leq 2) = 1 - \frac{e^{-6} 6^0}{0!} - \frac{e^{-6} 6^1}{1!} - \frac{e^{-6} 6^2}{2!} = 1 - 25 \cdot e^{-6} = 0.938.$$

(c) Qual è la probabilità che esattamente in 2 dei prossimi 6 mesi non ci siano guasti? Dobbiamo qui considerare una variabile  $Z \sim B(6, e^{-1})$  e quindi

$$P(Z = 2) = \binom{6}{2} (e^{-1})^2 (1 - e^{-1})^4 = 0.324.$$

ESERCIZIO 1.13. In un volo di 24 posti si vendono 28 biglietti. Sapendo che ogni passeggero ha probabilità di non presentarsi del 10%, qual è la probabilità che qualcuno non trovi posto? (0.6944). E quanti biglietti posso vendere se ammetto una probabilità del 10 % che qualcuno non trovi posto? Per rispondere a domande di questo tipo ci sarà utile la teoria delle variabili aleatorie continue.

ESERCIZIO 1.14. Un programma ha bisogno di 3000 file suddivisi in 30 cartelle con 100 file ciascuno per funzionare correttamente. Ogni volta che viene lanciato utilizza 28 file. Se una cartella non è stata installata qual è la probabilità che il software funzioni correttamente? Da risolvere sia con che senza ripetizioni. Con ripetizioni siamo in presenza di una probabilità binomiale e la probabilità è  $(29/30)^{28} = 0,387$ . Senza ripetizioni siamo in presenza di una ipergeometrica e abbiamo

$$\frac{\binom{2900}{28}}{\binom{3000}{28}} = \frac{2900 \cdot 2899 \cdots 2873}{3000 \cdot 2999 \cdots 2973} = 0.385.$$

Osserviamo che le due probabilità sono molto simili. Perché?

ESERCIZIO 1.15. Esercizio 1 della prova scritta del 10/09/2009

ESERCIZIO 1.16. Un album contiene 100 figurine e se ne sono già attaccate 60. Comprando un pacchetto contenente 6 figurine, tutte distinte tra loro, determinare la probabilità di trovare almeno 4 figurine che non si hanno già. Possiamo modellizzare il nostro problema con uno schema successo-insuccesso *senza* rimpiazzo. Rappresentiamo cioè il nostro problema come se le figurine che abbiamo già siano palline rosse e quelle che ci mancano come palline bianche. Le figurine di un pacchetto sono quindi rappresentate dall'estrazione di 6 palline. Abbiamo quindi che la variabile  $X$  = numero di figurine che non ho già trovate è una variabile ipergeometrica  $X \sim H(6, 40, 60)$  da cui

$$P(X \geq 4) = P(X = 4) + P(X = 5) + P(X = 6) = \frac{\binom{40}{4} \binom{60}{2}}{\binom{100}{6}} + \frac{\binom{40}{5} \binom{60}{1}}{\binom{100}{6}} + \frac{\binom{40}{6} \binom{60}{0}}{\binom{100}{6}} = 0.172.$$

## 2. Densità congiunte e marginali

Siamo spesso interessati a considerare più di una variabile aleatoria 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  $\underline{x} \in \mathbb{R}^m$ . La densità di una variabile  $m$ -dimensionale è la funzione  $p : \mathbb{R}^m \rightarrow \mathbb{R}_{\geq 0}$  data da  $p(\underline{x}) = P(\underline{X} = \underline{x})$ . Questa densità si dice densità congiunta delle variabili  $X_1, \dots, X_m$ .

ESEMPIO 2.1. 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 2.2. 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

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

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 3.3 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



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

Questo esempio mostra come le densità marginali non determinino univocamente la densità congiunta. È invece vero il contrario: vediamo per semplicità il caso  $m = 2$ . Sia  $z \in \mathbb{R}$ . Allora

$$\begin{aligned}
p_1(z) &= P(X_1 = z) \\
&= P(\cup_{x_2 \in \mathbb{R}} (X_1 = z, X_2 = x_2)) \\
&= \sum_{x_2 \in \mathbb{R}} P(X_1 = z, X_2 = x_2) \\
&= \sum_{x_2 \in \mathbb{R}} p(z, x_2).
\end{aligned}$$

In questo calcolo abbiamo usato la simbologia  $\cup_{x \in \mathbb{R}}$  e  $\sum_{x \in \mathbb{R}}$  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.

### 3. Variabili aleatorie indipendenti

Date  $n$  variabili aleatorie  $X_1, \dots, X_n$ , diciamo che esse sono indipendenti se gli eventi  $\{X_1 \in A_1\}, \dots, \{X_n \in A_n\}$  sono indipendenti per ogni scelta degli insiemi  $A_i \subseteq \mathbb{R}$ . In altre parole le variabili  $X_1, \dots, X_n$  sono indipendenti se vale la relazione

$$P(X_1 \in A_1, \dots, X_n \in A_n) = P(X_1 \in A_1) \cdots P(X_n \in A_n).$$

Consideriamo ora  $n$  punti di  $\mathbb{R}$ ,  $x_1, \dots, x_n$ . Se le variabili  $X_1, \dots, X_n$  sono indipendenti sussiste la relazione

$$p(\underline{x}) = p_1(x_1) \cdots p_n(x_n).$$

Ma è vero anche il contrario se le variabili sono discrete!

**LEMMA 3.1.** *Siano  $X, Y$  due variabili aleatorie discrete. Sia  $p(x, y)$  la densità congiunta e siano  $p_X(x)$  e  $p_Y(y)$  le marginali. Allora  $X$  e  $Y$  sono indipendenti se e solo se*

$$p(x, y) = p_X(x)p_Y(y),$$

per ogni  $x, y \in \mathbb{R}$ .

Dimostrazione (da non fare in aula): poniamo  $C = A \times B$ . Allora

$$P(X \in A, Y \in B) = P((X, Y) \in C) = \sum_{(x, y) \in C} p(x, y) = \sum_{x \in A, y \in B} p_X(x)p_Y(y).$$

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 2.2 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  $p(x, y) = p_X(x)p_Y(y)$  per ogni  $x, y \in \mathbb{R}$ , mentre per mostrare che sono dipendenti è sufficiente verificare che *esistono* due valori  $x, y \in \mathbb{R}$  tali che  $p(x, y) \neq p_X(x)p_Y(y)$ .

### Massimo e minimo di variabili aleatorie.

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 3.2. 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 3.3. 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 si è sfruttato che  $S$  e  $T$  sono indipendenti. □

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  $W = \min(S, T)$  e si calcola  $1 - F_W(k) = P(W > k)$ . Abbiamo in questo caso

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

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  $W \sim \tilde{G}(7/12)$ .

#### 4. Cambi di variabile

Consideriamo una variabile  $X$  di densità

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

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 (se  $\underline{X}$  è variabile aleatoria  $m$ -dimensionale discreta). Se  $\underline{X}$  assume i valori  $\underline{x}_1, \underline{x}_2, \dots$  allora  $\Phi(\underline{X})$  assume i valori  $\Phi(\underline{x}_1), \Phi(\underline{x}_2), \dots$  e quindi  $\Phi(\underline{X})$  assume al più una quantità numerabile di valori. Inoltre, dato  $y \in \mathbb{R}$  si ha

$$\{\Phi(\underline{X}) = y\} = \{\underline{X} \in \Phi^{-1}(y)\} = \cup_{\underline{x} \in \Phi^{-1}(y)} \{\underline{X} = \underline{x}\},$$

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

$$p_{\Phi(\underline{X})}(y) = \sum_{\underline{x} \in \Phi^{-1}(y)} p_{\underline{X}}(\underline{x}).$$

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  $\Phi, \Psi : \mathbb{R} \rightarrow \mathbb{R}$ .

$$\begin{aligned} P(\Phi(X) = z, \Psi(Y) = w) &= \sum_{\substack{x \in \Phi^{-1}(z) \\ y \in \Psi^{-1}(w)}} p_X(x) p_Y(y) \\ &= \sum_{x \in \Phi^{-1}(z)} p_X(x) \sum_{y \in \Psi^{-1}(w)} p_Y(y) \\ &= \sum_{\Phi(x)=z} p_X(x) \sum_{\Psi(y)=w} p_Y(y). \end{aligned}$$

#### 5. Calcoli con densità

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

**ESEMPIO 5.1.** 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

$$P(S < T) = \sum_{(x,y) \in A} p(x,y),$$

dove  $p$  è la densità congiunta delle variabili  $S$  e  $T$ , e  $A = \{(x, y) \in \mathbb{R}^2 : x < y\}$ . Siccome le variabili  $S \sim \tilde{G}(\frac{1}{2})$  e  $T \sim \tilde{G}(\frac{1}{6})$  sono indipendenti abbiamo

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

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

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

dove abbiamo sfruttato 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 il cambio di variabile  $\Phi(S, T) = S - T$  e studiato l'evento  $\{\Phi(S, T) > 0\}$ .

Abbiamo visto che

$$P(\Phi(X) = y) = \sum_{\underline{x} \in \Phi^{-1}(y)} p(\underline{x}).$$

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 5.2.** 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?

**Somma di variabili di Poisson.** Consideriamo due variabili  $X \sim P(\lambda)$  e  $Y \sim P(\mu)$  di Poisson di parametri rispettivamente  $\lambda$  e  $\mu$ , 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 5.3. 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

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

ESERCIZIO 5.4. Esercizio n.1 dell'esame del 15/06/2010 e esercizio n.1 dell'esame del 02/02/2011

## 6. La media o speranza matematica

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. In generale poniamo

DEFINIZIONE. Sia  $X$  una variabile aleatoria discreta di densità  $p(x)$ . Le media di  $X$  è data da

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

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

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

ESEMPIO 6.2. 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  $B(n, p)$ .

**Media nei cambi di variabili: la media della somma.** È importante saper calcolare la media nel caso in cui una variabile sia costruita a partire da altre.

ESEMPIO 6.3. Si consideri una variabile  $X$  di densità

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

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

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

dove abbiamo utilizzato i risultati della §4, da cui possiamo ottenere

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

Potremmo essere tentati di calcolare  $E[X^2]$  anche nel seguente modo

$$\sum_{x \in \mathbb{R}} x^2 p_X(x) = 1^2 \frac{1}{4} + (-1)^2 \frac{1}{4} + 2^2 \frac{1}{4} + (-2)^2 \frac{1}{8} + 3^2 \frac{1}{8} = \frac{25}{8} :$$

avremmo ottenuto lo stesso risultato. Il prossimo teorema ci assicura che possiamo utilizzare questa ultima formula anche in un contesto molto più generale.

**TEOREMA 6.4.** *Sia  $X = (X_1, \dots, X_m)$  una variabile aleatoria  $m$ -dimensionale discreta e  $\Phi := \mathbb{R}^m \rightarrow \mathbb{R}$ . Poniamo  $Z = \Phi(X)$ . Allora*

$$E[Z] = \sum_{\underline{x} \in \mathbb{R}^m} \Phi(\underline{x}) p(\underline{x}),$$

dove  $p(\underline{x})$  indica la densità di  $X$ .

**DIMOSTRAZIONE.** Sfruttando i risultati della sezione §4 si ha

$$\begin{aligned} E[Z] &= \sum_{z \in \mathbb{R}} z P(Z = z) \\ &= \sum_{z \in \mathbb{R}} z \sum_{\underline{x} \in \Phi^{-1}(z)} p(\underline{x}) \\ &= \sum_{\underline{x} \in \mathbb{R}^m} \Phi(\underline{x}) p(\underline{x}). \end{aligned}$$

□

**PROPOSIZIONE 6.5.** *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

$$\begin{aligned} E[cX] &= \sum_{x \in \mathbb{R}} (cx) p(x) = cE[X]; \\ E[X + Y] &= \sum_{(x,y) \in \mathbb{R}^2} (x + y) p(x, y) \\ &= \sum_{(x,y) \in \mathbb{R}^2} xp(x, y) + \sum_{(x,y) \in \mathbb{R}^2} yp(x, y) \\ &= \sum_{x \in \mathbb{R}} x \sum_{y \in \mathbb{R}} p(x, y) + \sum_{y \in \mathbb{R}} y \sum_{x \in \mathbb{R}} p(x, y) \\ &= \sum_{x \in \mathbb{R}} xp_X(x) + \sum_{y \in \mathbb{R}} yp_Y(y), \end{aligned}$$

dove abbiamo utilizzato la formula che esprime le densità marginali in funzione della densità congiunta. □

**La media di variabili binomiali, ipergeometriche e di Poisson.** Possiamo a questo punto calcolare facilmente la media di una variabile binomiale  $B(n, p)$ . 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  $\lambda$ . Ne segue (in modo non rigoroso) che se  $X \sim P(\lambda)$  allora  $E[X] = \lambda$ . In alternativa si può effettuare il calcolo esplicito

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

ESERCIZIO 6.6. Esercizio sul minimo di variabili geometriche modificate indipendenti.

ESERCIZIO 6.7. Esercizio sulla somma di variabili di Poisson

ESERCIZIO 6.8. Esercizio sulla media di variabili uniformi e geometriche modificate

**La media di una variabile geometrica modificata.** Lanciando un dado ci potremmo aspettare che mediamente dobbiamo effettuare sei lanci per ottenere un 6. Questo è confermato nel seguente risultato.

PROPOSIZIONE 6.9. 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  $f(x) = \frac{x}{1-x}$ . 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

$$\begin{aligned} E[X] &= \sum_{k=1}^{\infty} k p (1-p)^{k-1} \\ &= p f'(1-p) \\ &= p \frac{1}{(1-(1-p))^2} \\ &= \frac{1}{p}. \end{aligned}$$

□

ESEMPIO 6.10. 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 media di un prodotto di variabili.** Abbiamo visto che la media di una somma è la somma delle medie. Cosa si può dire del prodotto di due variabili? Vediamo 2 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 6.11.** *Se  $X$  ed  $Y$  sono variabili aleatorie indipendenti, allora*

$$E[XY] = E[X]E[Y].$$



DIMOSTRAZIONE. Sfruttiamo il Teorema 6.4 per calcolare  $E[XY]$ :

$$\begin{aligned} E[XY] &= \sum_{(x,y) \in \mathbb{R}^2} xyp(x,y) \\ &= \sum_{(x,y) \in \mathbb{R}^2} xyp_X(x)p_Y(y) \\ &= \sum_{x \in \mathbb{R}} xp_X(x) \sum_{y \in \mathbb{R}} yp_Y(y) \\ &= E[X]E[Y]. \end{aligned}$$

□

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 6.12. Si consideri la variabile bidimensionale  $(X, Y)$  di densità

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

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 6.13. 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

$$E[X] = E[X_1] + E[X_2] = 6 + 6 = 12.$$

## 7. Varianza

Se  $X$  è una variabile aleatoria discreta definiamo la varianza di  $X$  ricordando la definizione di varianza di un carattere in statistica descrittiva

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

Scriveremo talvolta  $\sigma_X^2$  al posto di  $\text{Var}(X)$ . Valgono le stesse osservazioni fatte nel capitolo 1 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 la stessa e quindi non la ripetiamo. 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.

LEMMA 7.1. *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  $\lambda$  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  $\lambda$  in una variabile di Poisson rappresenta sia la media che la varianza della variabile stessa.

PROPOSIZIONE 7.2 (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. □

ESEMPIO 7.3. Stimare, utilizzando la disuguaglianza di Chebyshev, la probabilità che una variabile  $X \sim B(10, 1/2)$  soddisfi  $|X - 5| \geq 3$ . Abbiamo

$$P(|X - 5| \geq 3) = P(|X - 5| > 2.99) \leq \frac{10/4}{2.99^2} = 0.28.$$

### 8. La legge dei grandi numeri

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 legge dei grandi numeri permette di dare un senso preciso a questo tipo di intuizione. Abbiamo 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 variabile aleatoria  $X$  se per ogni  $\varepsilon$  fissato si ha

$$\lim_{n \rightarrow \infty} P(|X_n - X| > \varepsilon) = 0.$$

**TEOREMA 8.1** (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  $\mu$  e varianza  $\sigma^2$ ). 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  $\mu$ .

**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 8.2** (Legge dei grandi numeri per la varianza). Stesse ipotesi del Teorema 8.1. 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  $\sigma^2$ .

**DIMOSTRAZIONE.** 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 8.1 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 convergenza 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*.

**ESEMPIO 8.3.** Si lancia un dado per 4 volte. Sia  $X$  la variabile che conta quante volte è uscito 6 e  $Y$  la variabile che conta il numero di volte in cui è uscito un numero dispari seguito da un numero pari.

- (a) Determinare le densità marginali delle variabili  $X$  ed  $Y$ .

$X$  è una variabile binomiale  $B(4, 1/6)$ . Facciamo qualche osservazione sulla variabile  $Y$ . Questa è somma di 3 variabili di Bernoulli aventi tutte la stessa legge, ma non è né binomiale né ipergeometrica. Vediamola più nel dettaglio. Sia

$$Y_i = \begin{cases} 1 & \text{se il lancio } i \text{ dà dispari e il lancio } i+1 \text{ dà pari} \\ 0 & \text{altrimenti} \end{cases}$$

Si ha chiaramente  $Y = Y_1 + Y_2 + Y_3$ . Le variabili  $Y_1$  e  $Y_3$  sono indipendenti (e quindi  $Y_1 + Y_3$  è binomiale), mentre  $Y_2$  è (molto) dipendente dalle altre 2 (ad esempio  $Y_2 = 1$  implica  $Y_1 = Y_3 = 0$ ). Comunque sia vediamo nel dettaglio come determinare la densità di  $Y$ .  $Y$  può assumere solo valori compresi tra 0 e 2. Abbiamo

$$P(Y = 2) = \frac{1}{16}.$$

in quanto  $Y = 2$  solo se abbiamo una sequenza DPDP.

$$P(Y = 1) = P(Y_2 = 1) + P(Y_1 = 1)P(Y_3 = 0) + P(Y_1 = 0)P(Y_3 = 1) = \frac{1}{4} + \frac{1}{4} \frac{3}{4} + \frac{3}{4} \frac{1}{4} = \frac{5}{8},$$

dove abbiamo sfruttato che  $Y_2 = 1$  esclude sia  $Y_1 = 1$  che  $Y_3 = 1$ , e che le variabili  $Y_1$  e  $Y_3$  sono indipendenti. Per differenza abbiamo anche

$$P(Y = 0) = 1 - \frac{1}{16} - \frac{5}{8} = \frac{5}{16},$$

ma puoi provare a calcolarlo anche direttamente.

- (b) Detta  $p_{X,Y}$  la densità congiunta di  $X$  ed  $Y$  determinare  $p_{X,Y}(3, 1)$ .

$$p_{X,Y}(3, 1) = P(D666) + P(6D66) + P(66D6) = 3 \frac{1}{2} \frac{1}{6^3} = \frac{1}{144}.$$

- (d) Le variabili  $X$  ed  $Y$  sono indipendenti? No, basta verificare che  $p_{X,Y}(3, 1) \neq p_X(3)p_Y(1) = \frac{25}{2592}$ .  
 (e) Determinare la varianza di  $X$  ed  $Y$ . La varianza di  $X$  è  $4 \frac{1}{6} \frac{5}{6} = \frac{5}{9}$ . Per calcolare la varianza di  $Y$  determiniamo la sua media

$$E[Y] = E[Y_1] + E[Y_2] + E[Y_3] = 3 \frac{1}{4} = \frac{3}{4}.$$

mentre

$$E[Y^2] = 0 \cdot \frac{5}{16} + 1 \cdot \frac{5}{8} + 4 \cdot \frac{1}{16} = \frac{7}{8}.$$

$$\text{E quindi } \text{Var}(Y) = E[Y^2] - E[Y]^2 = \frac{7}{8} - \frac{9}{16} = \frac{5}{16}.$$

**ESERCIZIO 8.4.** Si hanno 112 dadi di cui 56 truccati. Quelli truccati danno 1 con probabilità  $1/2$ , mentre gli altri 5 risultati hanno tutti probabilità  $1/10$ . Si prende un dado a caso e si lancia.

- (a) Qual è la probabilità che esca 3?

Sia  $X$  = risultato del lancio,  $A$  l'evento " il dado scelto è truccato" e  $B = A^C$ . Allora

$$P(X = 3) = P(A)P(X = 3|A) + P(B)P(X = 3|B) = \frac{2}{15}.$$

- (b) Qual è il risultato medio? Determiniamo la densità  $p(x)$  di  $X$ . Chiaramente per simmetria si ha  $p(2) = p(3) = p(4) = p(5) = p(6) = \frac{2}{15}$ . Di conseguenza  $p(1) = 1 - 5 \frac{2}{15} = \frac{1}{3}$  e quindi

$$E[X] = \sum_{k=1}^6 kp(k) = 1 \cdot \frac{1}{3} + (2 + 3 + 4 + 5 + 6) \frac{2}{15} = 3.$$

(c) Si consideri ora questo nuovo fenomeno aleatorio. Si sceglie un dado a caso e si lancia 2 volte. Si consideri la variabile bidimensionale  $(X, Y)$  data dai risultati dei 2 lanci e sia  $p(x, y)$  la corrispondente densità congiunta. Determinare  $p(2, 3)$ .

$$\begin{aligned} p(2, 3) &= P(X = 2 \cap Y = 3) \\ &= P(A)P(X = 2 \cap Y = 3|A) + P(B)P(X = 2 \cap Y = 3|B) \\ &= \frac{17}{900} \\ &= 1.89\% \end{aligned}$$

(d) Le variabili  $X$  ed  $Y$  sono indipendenti? No, ad esempio,

$$1.89\% = \frac{17}{900} = P(X = 2 \cap Y = 3) \neq P(X = 2)P(Y = 3) = \frac{2}{15} \frac{2}{15} = \frac{4}{225} = 1.78\%$$

(e) Determinare la media della variabile  $X + Y$  somma dei 2 lanci. Siccome per il punto (b) sia che  $X$  che  $Y$  hanno media 3 concludiamo che  $E[X + Y] = 3 + 3 = 6$  (anche se le variabili  $X$  ed  $Y$  sono dipendenti).

ESEMPIO 8.5. Siano  $X$  ed  $Y$  due variabili aleatorie definite sullo stesso spazio di probabilità la cui densità congiunta è espressa dalla seguente tabella

$X \backslash Y$	-1	0	2
3	$\frac{1}{24}$	$\frac{1}{12}$	$\frac{1}{8}$
4	$\frac{1}{24}$	$\frac{1}{12}$	$\frac{1}{8}$
6	$\frac{1}{12}$	$\frac{1}{6}$	$\frac{1}{4}$

(1) Determinare  $P(X + Y \geq 6)$ .

*Si vede facilmente che  $X + Y \geq 6$  nei 3 casi  $(6, 0)$ ,  $(4, 2)$  e  $(6, 2)$ . Si ha quindi*

$$P(X + Y \geq 6) = \frac{1}{8} + \frac{1}{6} + \frac{1}{4} = \frac{13}{24}.$$

(2) Determinare le densità marginali di  $X$  ed  $Y$ ;

*Basta fare la somma delle righe e delle colonne della tabella. Si ha*

$$p_X(k) = \begin{cases} \frac{1}{4} & \text{se } k = 3, 4; \\ \frac{1}{2} & \text{se } k = 6; \\ 0 & \text{altrimenti} \end{cases}$$

e

$$p_Y(k) = \begin{cases} \frac{1}{6} & \text{se } k = -1; \\ \frac{1}{3} & \text{se } k = 0; \\ \frac{1}{2} & \text{se } k = 2; \\ 0 & \text{altrimenti} \end{cases}$$

- (3) Stabilire se  $X$  ed  $Y$  sono indipendenti;

*Si vede facilmente che la densità congiunta  $p(x, y)$  è uguale al prodotto delle marginali  $p_X(x) \cdot p_Y(y)$  PER OGNI coppia di valori  $x$  ed  $y$  e quindi le variabili  $X$  ed  $Y$  sono indipendenti.*

- (4) Determinare la media di  $X + Y$ .

*Calcoliamo la media di  $X$  e di  $Y$  utilizzando le densità marginali. Abbiamo*

$$E[X] = 3 \cdot \frac{1}{4} + 4 \cdot \frac{1}{4} + 6 \cdot \frac{1}{2} = \frac{19}{4},$$

$$E[Y] = -1 \cdot \frac{1}{6} + 0 \cdot \frac{1}{3} + 2 \cdot \frac{1}{2} = \frac{5}{6}$$

*e quindi*

$$E[X + Y] = E[X] + E[Y] = \frac{67}{12} = 5.58.$$

- (5) Determinare la varianza di  $X - Y$ .

*Siccome  $X$  e  $-Y$  sono indipendenti abbiamo*

$$\text{Var}(X - Y) = \text{Var}(X) + \text{Var}(-Y) = \text{Var}(X) + (-1)^2 \text{Var}(Y) = \text{Var}(X) + \text{Var}(Y).$$

*Calcoliamo le medie dei quadrati per poter determinare la varianza di  $X$  ed  $Y$ . Abbiamo*

$$E[X^2] = (16 + 9) \frac{1}{4} + 36 \frac{1}{2} = \frac{97}{4}$$

*e*

$$E[Y^2] = \frac{1}{6} + 4 \frac{1}{2} = \frac{13}{6}.$$

*Di conseguenza  $\text{Var}(X) = E[X^2] - E[X]^2 = \frac{97}{4} - \frac{361}{16} = \frac{27}{16}$  e  $\text{Var}(Y) = E[Y^2] - E[Y]^2 = \frac{13}{6} - \frac{25}{36} = \frac{53}{36}$ .*

*Concludiamo che*

$$\text{Var}(X - Y) = \frac{27}{16} + \frac{53}{36} = 3.16.$$

## Variabili aleatorie continue

### 1. Funzione di ripartizione e densità

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  $\mathbb{R}$ .

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

se  $a \leq b$ .

LEMMA 1.1. *Se una variabile aleatoria  $X$  è continua in un intervallo aperto  $I$ , allora  $P(X = t) = 0$  per ogni  $x \in I$ .*

DIMOSTRAZIONE. Si ha

$$P(X = t) \leq P(t - \frac{1}{n} < X \leq t) = F(t) - F(t - \frac{1}{n}),$$

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 1.1.

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: cambiando 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

$$(2) \quad \int_{-\infty}^{+\infty} f(s) ds = 1.$$

Una funzione che soddisfa queste 2 proprietà verrà detta una *densità continua*.

OSSERVAZIONE. Se una variabile  $X$  ammette densità continua allora la sua funzione di ripartizione è continua e quindi  $X$  stessa è continua.

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  $X$ .

La nozione di indipendenza tra variabili rimane la stessa, cioè diremo che 2 variabili  $X$  ed  $Y$  sono indipendenti se presi comunque due intervalli reali  $I$  e  $J$  si ha

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

In particolare, scegliendo intervalli illimitati a sinistra otteniamo

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

## 2. La media e la varianza

In generale la media di una variabile continua di densità  $f$  si definisce tramite

$$E[X] = \int_{-\infty}^{+\infty} sf(s) ds.$$

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

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  $F_X(t)$  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  $[0, 1]$ .

La variabile uniforme nell'intervallo  $[a, b]$ . È definita in modo che la sua densità sia costante nell'intervallo  $[a, b]$  e sia nulla altrove. Si ha quindi che la densità uniforme data da

$$f(s) = \begin{cases} \frac{1}{b-a} & s \in [a, b] \\ 0 & \text{altrimenti.} \end{cases}$$

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

Ricordiamo che una primitiva della funzione  $e^{-\lambda t}$  è data da  $-\frac{1}{\lambda}e^{-\lambda t}$  da cui abbiamo che:

$$(4) \quad \int_0^{+\infty} e^{-\lambda t} dt = \frac{1}{\lambda}.$$

ESEMPIO 4.1. Si consideri la funzione

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

dove  $\lambda$  è un parametro reale positivo. Verifichiamo che si tratta di una densità continua. Si ha chiaramente  $f(s) \geq 0$  per ogni  $s \in \mathbb{R}$ . Quanto all'integrale

$$\int_{-\infty}^{+\infty} f(s) ds = \int_0^{\infty} \lambda e^{-\lambda s} ds = 1$$

grazie a (4). Una variabile  $X$  che ammette tale densità si dice *esponenziale* di parametro  $\lambda$  e scriveremo  $X \sim \text{Exp}(\lambda)$ .

Lo stesso calcolo mostra che la funzione di ripartizione di una variabile  $X \sim \text{Exp}(\lambda)$  è data da

$$F_X(t) = 1 - e^{-\lambda t}$$

Una variabile esponenziale soddisfa la proprietà di mancanza di memoria. Infatti

$$P(X > t + u | X > t) = \frac{P(X > t + u)}{P(X > t)} = \frac{e^{-\lambda(t+u)}}{e^{-\lambda t}} = e^{-\lambda u} = P(X > u).$$

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  $X \sim \text{Exp}(\lambda)$ . Si ha

$$\begin{aligned} E[X] &= \int_0^\infty x \lambda e^{-\lambda x} dx \\ &= \left[ -x e^{-\lambda x} \right]_0^\infty + \int_0^\infty e^{-\lambda x} dx \\ &= \frac{1}{\lambda} \end{aligned}$$

dove abbiamo utilizzato (4). Mentre

$$\begin{aligned} E[X^2] &= \int_0^\infty x^2 \lambda e^{-\lambda x} dx \\ &= \left[ -x^2 e^{-\lambda x} \right]_0^\infty + \int_0^\infty 2x e^{-\lambda x} dx \\ &= 0 + \frac{2}{\lambda} \int_0^\infty x \lambda e^{-\lambda x} dx \\ &= \frac{2}{\lambda} \frac{1}{\lambda} \\ &= \frac{2}{\lambda^2} \end{aligned}$$

dove abbiamo sfruttato anche il calcolo precedente per determinare  $E[X]$ . Concludiamo che

$$\text{Var}(X) = E[X^2] - E[X]^2 = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}.$$

## 5. Cambio di 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  $\phi$  è continua tranne al più in un numero finito di punti. In tal caso come si può fare per calcolare la densità di una variabile composta? Sia  $f$  la densità di  $X$  e  $G$  la funzione di ripartizione di  $\phi(X)$ . Allora

$$G(t) = P(\phi(X) \leq t) = P(X \in \phi^{-1}((-\infty, t])) = \int_{\phi^{-1}((-\infty, t])} f(x) dx$$

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 5.1. 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 5.2. 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 5.3. Siano  $X \sim \text{Exp}(\lambda)$  ed  $Y \sim \text{Exp}(\mu)$  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 3.2 si ha

$$F_Z(t) = P(X \leq t)P(Y \leq t) = F_X(t)F_Y(t) = (1 - e^{-\lambda t})(1 - e^{-\mu t}).$$

E per derivazione otteniamo

$$f_Z(t) = \lambda e^{-\lambda t}(1 - e^{-\mu t}) + \mu e^{-\mu t}(1 - e^{-\lambda t}).$$

Se poniamo  $W = \min(X, Y)$  possiamo invece verificare che  $W \sim \text{Exp}(\lambda + \mu)$ .

ESERCIZIO 5.4. Sia  $X$  una variabile continua uniforme nell'intervallo  $[0, 1]$  ed  $Y$  la variabile (discreta) data da  $Y = \lfloor 5X \rfloor$ , dove  $\lfloor x \rfloor$  indica la parte intera di  $x$ , cioè il più grande numero intero minore o uguale a  $x$ . Determinare la densità di  $Y$ .

ESERCIZIO 5.5. Una certa pianta ha tempo di vita esponenziale di media 8 anni. Compro due di queste piante. Per quanto tempo, mediamente, vivrà almeno una delle due piante?

ESERCIZIO 5.6. Si consideri la seguente funzione  $f: \mathbb{R} \rightarrow \mathbb{R}$

$$f(t) = \begin{cases} e^{-2t}, & \text{se } t \geq 0; \\ e^{2t}, & \text{se } t \leq 0. \end{cases}$$

- (1) Verificare che  $f$  è la densità di una variabile aleatoria continua  $X$ ;

Si ha  $f(t) \geq 0$  e

$$\int_{-\infty}^{+\infty} f(t) dt = 2 \int_0^{+\infty} e^{-2t} dt = 1,$$

dove si è sfruttato che la funzione  $f$  è pari, e quindi  $f$  è una densità continua.

- (2) Determinare la funzione di ripartizione di  $X$ ;

Se  $t < 0$  si ha

$$F(t) = \int_{-\infty}^t e^{2s} ds = \left[ \frac{1}{2} e^{2s} \right]_{-\infty}^t = \frac{1}{2} e^{2t}.$$

Se  $t > 0$  si ha

$$F(t) = \int_{-\infty}^t f(s) ds = \int_{-\infty}^0 e^{2s} ds + \int_0^t e^{-2s} ds = \frac{1}{2} + \left[ -\frac{1}{2} e^{-2s} \right]_0^t = 1 - \frac{1}{2} e^{-2t}.$$

- (3) Determinare la media di  $X$ ; Si ha  $E[X] = 0$  perché la densità è pari.

- (4) Determinare la probabilità che  $X > 1$ ;

$$P(X > 1) = 1 - P(X \leq 1) = 1 - F_X(1) = 1 - \left(1 - \frac{1}{2} e^{-2}\right) = \frac{1}{2} e^{-2} = 6.7\%$$

## 6. Densità normali

Si può dimostrare che

$$(5) \quad \int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} dx = \sqrt{2\pi}.$$

Ne segue che la funzione

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

è una densità. Questa densità viene detta normale standard o gaussiana standard. Se una variabile  $X$  ammette questa densità scriveremo

$$X \sim N(0, 1).$$

Utilizzeremo il simbolo  $\zeta_0$  per indicare una variabile normale standard. Siano ora  $\sigma > 0$  e  $\mu$  due parametri reali e  $\zeta_0 \sim N(0, 1)$ . Consideriamo la variabile  $\zeta = \sigma\zeta_0 + \mu$ . Questa variabile avrà densità

$$f_{\zeta}(y) = \frac{1}{\sigma} f_{\zeta_0}\left(\frac{y - \mu}{\sigma}\right) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(y - \mu)^2}{2\sigma^2}}.$$

Scriveremo in questo caso  $\zeta \sim N(\mu, \sigma^2)$ .

Sarà sovente il caso, ogni qual volta abbiamo a che fare con una variabile  $\zeta \sim N(\mu, \sigma^2)$  effettuare il cambio di variabile  $\zeta_0 = \frac{\zeta - \mu}{\sigma}$ . La variabile  $\zeta_0$  così definita è una normale standard. E in questo modo potremo fare i calcoli con la variabile  $\zeta_0$  per poi ritradurre i risultati ottenuti nella variabile  $\zeta$ .

Analisi del grafico di una densità normale 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.

Esercizio. 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  $\zeta$  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.

ESEMPIO 6.1. Se  $\zeta_0 \sim N(0, 1)$  si ha  $E[\zeta_0] = 0$  in quanto la densità di  $\zeta_0$  è pari. Per calcolarne la varianza dobbiamo determinare la media di  $\zeta_0^2$  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 (5). Concludiamo quindi che  $\text{Var}(\zeta_0) = 1$ .

ESEMPIO 6.2. Se  $\zeta = \sigma\zeta_0 + \mu \sim N(\mu, \sigma^2)$  è una variabile normale si ha

$$E[\zeta] = \sigma E[\zeta_0] + \mu = \mu,$$

mentre

$$\text{Var}[\zeta] = \text{Var}(\sigma\zeta_0 + \mu) = \sigma^2 \text{Var}(\zeta_0) + \text{Var}(\mu) = \sigma^2.$$

Abbiamo quindi che i parametri  $\mu$  e  $\sigma^2$  sono proprio la media e la varianza di una variabile normale  $\zeta \sim N(\mu, \sigma^2)$ .

ESEMPIO 6.3. 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  $\zeta$  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

$$P(\zeta > 185) = P(\zeta_0 > \frac{185 - 151}{15}) = P(\zeta_0 > 2,27) = 1 - \Phi(2,27) = 1 - 0,98840 = 1,16\%.$$

ESEMPIO 6.4. Esercizio 4.9. Un componente viene prodotto da 2 linee di produzione, di tempo di vita esponenziale  $\lambda$  e  $\mu$  (con  $\mu > \lambda$ ). Vengono prodotti in proporzioni  $p$  e  $q$ , con  $p + q = 1$ . Un pezzo viene scelto a caso e indichiamo con  $T$  il suo tempo di vita. Qual è la densità di  $T$ ?  $E(T)$ ? Sapendo che il componente è ancora funzionante al tempo  $s$  qual è la probabilità che venga dalla prima linea?

Denotiamo con  $L_i$  l'evento "il pezzo viene dalla linea di produzione  $i$  e  $T_i$  i corrispondenti tempi di vita. Determiniamo la funzione di ripartizione di  $T$ . Si ha

$$\begin{aligned} F_T(t) &= P(T \leq t) = P(T \leq t, L_1) + P(T \leq t, L_2) \\ &= P(L_1)P(T \leq t|L_1) + P(L_2)P(T \leq t|L_2) \\ &= pF_{T_1}(t) + qF_{T_2}(t) \end{aligned}$$

La densità di  $T$  è quindi data da

$$f_T(t) = F'_T(t) = pf_{T_1}(t) + qf_{T_2}(t),$$

e quindi la sua media è data da  $E[T] = pE[T_1] + qE[T_2] = \frac{p}{\lambda} + \frac{q}{\mu}$ . Infine, si ha

$$P(L_1|T > s) = \frac{P(T > s|L_1)P(L_1)}{P(T > s)} = \frac{e^{-\lambda s}p}{1 - (p(1 - e^{-\lambda s}) + q(1 - e^{-\mu s}))} = \frac{e^{-\lambda s}p}{pe^{-\lambda s} + qe^{-\mu s}} = \frac{p}{p + qe^{(\lambda - \mu)s}},$$

ed osserviamo, come è ragionevole che sia, che questa quantità tende a 1 per  $s \rightarrow +\infty$ .

## 7. Il teorema 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  $\mu$  e varianza  $\sigma^2$  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 7.1 (Limite centrale).** *Sia  $X_1, X_2, \dots$  una successione di variabili aleatorie indipendenti aventi tutte la stessa densità (discreta o continua) di media  $\mu$  e varianza  $\sigma^2$ . 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 7.2.** 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  $\lambda = 0.1$  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 approssimata da una variabile normale  $\zeta$  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 7.3. Torniamo all’Esempio 1.13. 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 limite centrale 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 7.4. 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 limite centrale 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 7.5. 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

$$P(\zeta > 5.1) = P(\zeta_0 > \frac{5.1 - 5}{0.01}) = P(\zeta_0 > 10) = 0.$$

Esercizi di riepilogo sulle variabili continue



## Statistica inferenziale

### 1. Stime e intervalli di confidenza

Il teorema 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  $\zeta_0$ .

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 limite centrale 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  $\sigma_X/\sqrt{n}$

$$P(-z_c \frac{\sigma_X}{\sqrt{n}} < \bar{X}_n - \mu_X < z_c \frac{\sigma_X}{\sqrt{n}}) = c$$

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 ( $\mu_X$ ) rispetto al valore medio registrato su un singolo campione  $\bar{X}_n$ . Il problema è che tale probabilità viene espressa in termini della varianza  $\sigma_X^2$  che comunque non conosciamo. In questo caso allora sfruttiamo la legge dei grandi numeri per la varianza pe ricordare che

la varianza  $\sigma_X^2$  è ben approssimata dalla varianza campionaria  $\bar{\sigma}_n^2$ . Utilizziamo questo valore al posto di  $\sigma_X^2$  che possiamo calcolare nel campione e parliamo di confidenza anziché di probabilità. Abbiamo quindi

$$(6) \quad \text{Conf}(\bar{X}_n - z_c \frac{\bar{\sigma}_n}{\sqrt{n}} < \mu_X < \bar{X}_n + z_c \frac{\bar{\sigma}_n}{\sqrt{n}}) = c$$

e diciamo che la media  $\mu_X$  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 1.1. 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 (6)  $\bar{\sigma}_n = 1$ ,  $\bar{X}_n = 5.1$ ,  $n = 100$  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  $\mu_X$  è

$$[5.1 - 0.196, 5.1 + 0.196] = [4.904, 5.296].$$

La confidenza con cui  $\mu_X$  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 (6)

$$\text{Conf}(\bar{X}_n - z_c \frac{\bar{\sigma}_n}{\sqrt{n}} < \mu_X < \bar{X}_n + z_c \frac{\bar{\sigma}_n}{\sqrt{n}}) = c.$$

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 1.2. 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 2.1.** 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 2.2.** 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  $\mu$  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 in 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 2.3. 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 2.4. 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  $H_1 : E[X] < 0.90$  e l'affermazione è falsa.