Capitolo 2

Relazioni, funzioni e calcolo combinatorio

2.1 Prodotto cartesiano

In questo capitolo introdurremo altre due importanti nozioni della teoria degli insiemi, quella di *relazione su un insieme* e quella di *funzione tra due insiemi*.
Per dare delle definizioni rigorose, abbiamo bisogno di introdurre prima una nuova operazione sugli insiemi, quella di *prodotto cartesiano*.

Definizione 2.1. Dati due insiemi A e B, si dice il **prodotto cartesiano di A per B**, denotato con $A \times B$, l'insieme di tutte le *coppie ordinate* (a,b) dove a è un elemento di A e b è un elemento di B. In simboli

$$A \times B = \{(a,b) \mid a \in A \wedge b \in B\} .$$

Dare una coppia ordinata (a,b) significa non solo dare due elementi a e b ma anche specificare che a è il primo e b è il secondo. In questo (a,b) si differenzia da $\{a,b\}$ (l'insieme degli elementi a e b) nel quale non importa l'ordine di a e b ma solo la loro appartenenza all'insieme dato[1]. Quindi, in generale, $(a,b) \neq (b,a)$ (a meno che non sia $a = b$).

Esempio 2.2. Siano $A = \{0,1,2,3\}$ e $B = \{3,4\}$. Ad esempio, la coppia $(0,3)$ appartiene a $A \times B$ in quanto la sua prima componente, 0, appartiene ad A e la seconda, 3, appartiene a B. Si ha in effetti

[1] Se volessimo dare una definizione insiemistica e rigorosa della coppia (a,b), questa potrebbe essere $(a,b) = \{\{a,b\},\{a\}\}$: l'insieme con due elementi ci dice quali sono gli elementi che compongono la coppia, l'insieme con un solo elemento ci dice chi è il primo della coppia. Tuttavia, non useremo mai tale notazione.

$$A \times B = \{(0,3),(0,4),(1,3),(1,4),(2,3),(2,4),(3,3),(3,4)\}$$

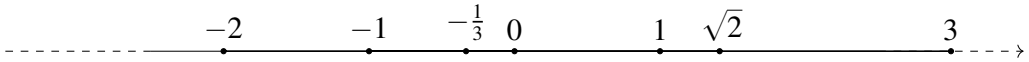
Si noti che il numero degli elementi di $A \times B$ è dato dal prodotto del numero degli elementi di A per il numero degli elementi di B. Infatti, per ogni elemento di A scelto come primo elemento della coppia, il secondo elemento può essere scelto tra uno qualunque degli elementi di B.

Come caso particolare possiamo considerare il prodotto di un insieme per se stesso. Per esempio, se $A = \{1,2,3\}$ si ha

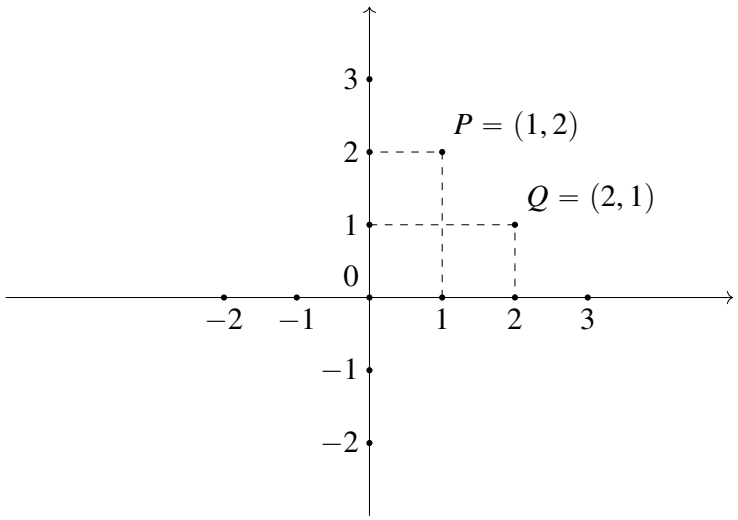
$$A \times A = \{(1,1),(1,2),(1,3),(2,1),(2,2),(2,3),(3,1),(3,2),(3,3)\}$$

Il prodotto cartesiano $A \times A$ di un insieme con se stesso si denota anche con A^2, con una notazione presa in prestito da quella usata per il prodotto usuale tra numeri dove effettivamente si scrive $a \cdot a = a^2$.

Osservazione 2.3. Il prodotto cartesiano $\mathbb{R}^2 = \mathbb{R} \times \mathbb{R}$ ha particolare importanza in geometria in quanto, fissato un sistema di coordinate cartesiane, rappresenta l'insieme dei punti del piano. Più precisamente, sappiamo che i numeri reali si possono rappresentare come i punti di una retta scegliendo prima un punto per rappresentare lo 0. Poi a distanza fissata (da noi scelta come unità di misura) gli interi, positivi da un lato e negativi dall'altro (quindi la retta risulta orientata). Inserendo successivamente le frazioni tra i numeri interi e infine i restanti numeri reali che non possono essere rappresentati come frazioni (gli *irrazionali*, ovvero i numeri decimali con infinite cifre dopo la virgola, non periodici):



A questo punto, se nel piano scegliamo due rette orientate (dette *assi cartesiani*), perpendicolari tra loro e che si incontrano nel punto scelto per lo 0, a ogni punto P si può assegnare una coppia di numeri reali (le sue *coordinate*) che sono determinati proiettando P sui due assi.



Si noti che è importante che le coordinate siano rappresentate mediante una coppia ordinata, in quanto, ad esempio, il punto di coordinate $(1,2)$ è diverso dal punto di coordinate $(2,1)$.

Il prodotto cartesiano può anche essere definito per più di due insiemi. Ad esempio, per 3 insiemi A, B, C basta usare invece della nozione di coppia ordinata quella di *terna ordinata* (a,b,c), nella quale, analogamente a quanto fatto per la coppia, diamo tre elementi in un ordine assegnato. In questo caso definiamo

$$A \times B \times C = \{(a,b,c) \mid a \in A \wedge b \in B \wedge c \in C\}.$$

Possiamo generalizzare tale definizione a un numero arbitrario di insiemi $A_1, A_2, \ldots, A_n$: in tal caso, useremo la nozione di *n-upla ordinata* $(a_1, a_2, \ldots, a_n)$ (che coincide con una coppia per $n = 2$ e con una terna per $n = 3$) e definiremo

$$A_1 \times A_2 \times \cdots \times A_n = \{(a_1, a_2, \ldots, a_n) \mid a_1 \in A \wedge a_2 \in A_2 \wedge \cdots \wedge a_n \in A_n\}.$$

Analogamente a quanto già detto per il prodotto tra due insiemi, denoteremo con A^n il prodotto cartesiano $A \times A \times \cdots \times A$ di A con se stesso n volte. Vedremo in un capitolo successivo che il prodotto $\mathbb{R}^n$ ha una particolare importanza in algebra vettoriale.

2.2 Relazioni

Introduciamo la definizione rigorosa di relazione su un insieme tramite un esempio concreto. Supponiamo che X sia un insieme i cui elementi sono persone, e supponiamo che alcuni elementi dell'insieme siano figli di altri sempre appartenenti all'insieme dato. In

altre parole, alcuni elementi dell'insieme X sono legati dalla relazione padre-figlio ad altri elementi di X.
Per raccogliere l'informazione su quali sono gli elementi di X in tale relazione tra loro, si può fare ricorso proprio alla nozione di coppia ordinata: ogniqualvolta si ha che x è figlio di x' per due elementi $x, x' \in X$, formiamo la coppia ordinata (x, x'). L'insieme delle coppie così costruite è un sottoinsieme del prodotto cartesiano $X \times X$ che ci dice esattamente chi è figlio di chi all'interno dell'insieme.
Si noti che per avere l'informazione corretta è necessario usare la coppia *ordinata*, perchè distinguendo tra la prima e la seconda componente specifica quale dei due elementi è figlio dell'altro. Quest'informazione sarebbe persa se usassimo solo l'insieme $\{x, x'\}$, che ci direbbe solo che tra x e x' uno è figlio dell'altro senza specificare chi. Questo ci porta alla seguente definizione generale.

Definizione 2.4. Una **relazione** su un insieme X è un sottoinsieme del prodotto cartesiano $X \times X$.

Esempio 2.5. Consideriamo sull'insieme dei numeri naturali la relazione "essere minore di". Questa relazione è rappresentata dal sottoinsieme S (infinito) di tutte le coppie (n, m) di numeri naturali in cui $n < m$:

$$S = \{(0,1), (0,2), \dots, (1,2), (1,3), \dots\}.$$

Se invece la relazione fosse quella di uguaglianza, il sottoinsieme sarebbe

$$S = \{(0,0), (1,1), (2,2), (3,3) \dots\}$$

in cui ogni elemento è in relazione solo con se stesso (in quanto uguale solo a se stesso).

Quando un elemento $x \in X$ è nella relazione data con un altro $x' \in X$, useremo la simbologia $x \sim x'$.

Come abbiamo osservato sopra, si usano coppie ordinate per formalizzare la definizione di relazione perchè in generale può essere $x \sim x'$ ma non $x' \sim x$. Ad esempio, per la relazione "essere figlio di", se x è figlio di x' sicuramente x' non è figlio di x.
Tuttavia, per alcune relazioni succede che

$$x \sim x' \Rightarrow x' \sim x. \tag{2.1}$$

In tal caso, si dice che la relazione è **simmetrica**.
Ad esempio, la relazione di uguaglianza è simmetrica in quanto se $x = x'$ sicuramente possiamo anche scrivere $x' = x$.
Un'altra proprietà soddisfatta dalla relazione di uguaglianza ma non dalle relazioni in generale, è che per ogni $x \in X$ si ha

$$x \sim x. \tag{2.2}$$

Ovvero ogni elemento è in relazione con se stesso (nel caso dell'uguaglianza, chiaramente questo vale in quanto $x = x$). Tra molte altre, la relazione "essere minore di" non soddisfa questa proprietà in quanto $x < x$ non è verificata. Quando vale questa proprietà, si dice che la relazione è **riflessiva**.
Infine, l'uguaglianza soddisfa una terza importante proprietà:

$$x \sim x' \text{ e } x' \sim x'' \Rightarrow x \sim x'' . \tag{2.3}$$

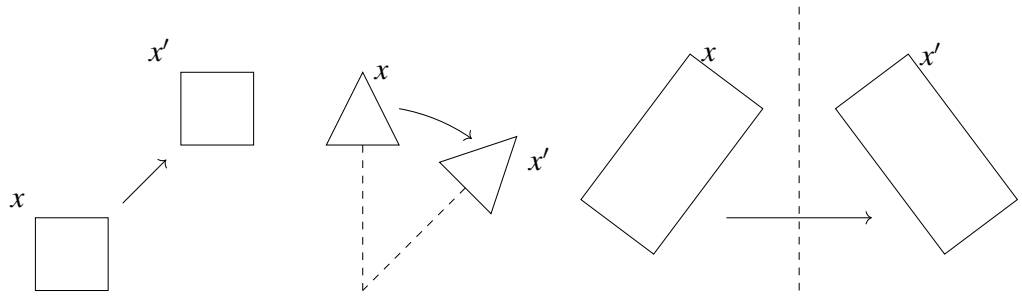
Ovvero se un elemento è in relazione con un secondo, e il secondo elemento è in relazione con un terzo, allora il primo è in relazione con il terzo. Quando vale ciò, la relazione si dice **transitiva**.
L'uguaglianza è chiaramente una relazione transitiva in quanto se $x = x'$ e $x' = x''$, allora $x = x''$. Anche la relazione "essere minore di" soddisfa tale proprietà in quanto $x < x'$ e $x' < x''$ implica $x < x''$. Diversamente la relazione "essere figlio di" invece non è transitiva, in quanto se x è figlio di x' e x' è figlio di x'', non è vero che x è figlio di x'' (sarà semmai il nipote di x'').

Definizione 2.6. Una relazione su un insieme X che sia riflessiva, simmetrica e transitiva si dice **relazione di equivalenza**.

La definizione precedente è di grande importanza per la matematica e in questo corso in particolare. Per questo motivo sarà l'oggetto della prossima sezione.

2.3 Relazioni di equivalenza

Come visto nella sezione precedente, la relazione di uguaglianza è una relazione di equivalenza. In generale, una relazione di equivalenza è una sorta di "uguaglianza in senso largo", utile per identificare, ovvero considerare in un qualche senso uguali, oggetti che uguali non sono.
A sostegno di questa affermazione, consideriamo l'esempio seguente. Sia X l'insieme delle *figure* del piano, ovvero quadrati, triangoli, circonferenze e in generale qualunque insieme di punti.
Definiamo su tale insieme la seguente relazione: date due figure x e x', diciamo che $x \sim x'$ se x' può essere ottenuta da x attraverso una sequenza di traslazioni, rotazioni o riflessioni di x rispetto a qualche retta (ovvero tramite cosiddetti *movimenti rigidi*), come nel disegno seguente.



Mostriamo che si tratta di una relazione di equivalenza (che prende il nome di *congruenza*). Se $x \sim x'$, ovvero x' si ottiene da x mediante traslazioni, rotazioni o riflessioni, allora anche $x' \sim x$, in quanto x si può riottenere da x' applicando le trasformazioni inverse nell'ordine inverso. Questa relazione è quindi simmetrica.
Inoltre, se $x \sim x'$ e $x' \sim x''$, questo significa che x' si ottiene da x mediante certi movimenti rigidi, e x'' si ottiene da x' mediante altri movimenti di tale tipo. Ma allora anche x'' si ottiene da x mediante movimenti rigidi (e quindi $x \sim x''$): basta applicare in sequenza quelli che servono prima per ottenere x' e poi quelli che servono per ottenere x'' da x'. Quindi abbiamo anche la transitività.
Quanto alla riflessività, ovvero $x \sim x$, possiamo dire che vale in quanto x si ottiene da se stesso applicando per esempio una rotazione di angolo 0.
Ora, questo esempio chiarisce perché, come abbiamo detto sopra, una relazione di equivalenza può essere considerata come un'uguaglianza in senso largo. I due quadrati della figura precedente, così come i due triangoli o i due rettangoli, pur essendo figure diverse (non sono formate dagli stessi punti, quindi non possiamo dire che siano uguali), rappresentano tutte in un certo senso "la stessa figura".

Esempio 2.7. Tra le relazioni di parentela, un esempio di relazione di equivalenza è "essere fratello di", nel senso di avere gli stessi genitori. La riflessività è chiara: ogni x ha gli stessi genitori di se stesso. La simmetria è altrettanto palese, dal momento che se x è fratello di x', anche x' è fratello di x. Infine si ha anche la transitività, poiché se x è fratello di x' e x' è fratello di x'', allora è chiaro che x è fratello di x''.

Data una relazione di equivalenza, tutti gli elementi equivalenti tra loro possono essere raggruppati in in una cosiddetta classe di equivalenza.

Definizione 2.8. Sia X un insieme sul quale è definita una relazione d'equivalenza e sia $x \in X$. La **classe di equivalenza** di x è l'insieme

$$[x] = \{y \in X \mid y \sim x\} \tag{2.4}$$

cioè l'insieme di tutti gli elementi y che sono in relazione con (ovvero equivalenti a) x.

Elementi distinti x e x' possono dar luogo alla stessa classe di equivalenza, ovvero $[x] = [x']$. Questo accade esattamente quando $x \sim x'$.

Proposizione 2.9. *Si consideri una relazione di equivalenza $\sim$ su un insieme X. Per ogni $x, x' \in X$, si ha $[x] = [x'] \Leftrightarrow x \sim x'$.*

Dimostrazione. Trattandosi di una doppia implicazione, dobbiamo dimostrare sia $[x] = [x'] \Rightarrow x \sim x'$ che $x \sim x' \Rightarrow [x] = [x']$.
Per dimostrare la prima implicazione, ricordiamoci che, per definizione di uguaglianza tra insiemi, $[x] = [x']$ significa che ogni elemento della classe $[x]$ sta anche nella classe $[x']$ e viceversa ogni elemento di $[x']$ appartiene anche a $[x]$. Ma nella classe $[x]$ c'è sicuramente x stesso, perchè essendo la relazione di equivalenza riflessiva vale $x \sim x$, e quindi $x \in \{y \in X \mid y \sim x\} = [x]$. Quindi, essendo $x \in [x]$ e $[x] = [x']$, si ha anche $x \in [x']$, ovvero $x \sim x'$, come volevamo.
Per dimostrare l'implicazione inversa $x \sim x' \Rightarrow [x] = [x']$, dobbiamo far vedere che (sotto l'ipotesi $x \sim x'$), ogni elemento di $[x]$ appartiene anche a $[x']$ e viceversa.
Sia $y \in [x]$: questo significa che $y \sim x$; ma poiché abbiamo $x \sim x'$, possiamo applicare la transitività delle relazioni di equivalenza e concludere che $y \sim x'$, ovvero $y \in [x']$. Abbiamo allora dimostrato che ogni elemento di $[x]$ sta anche in $[x']$, ovvero che $[x] \subseteq [x']$. Per dimostrare che vale proprio l'uguaglianza, prendiamo viceversa un $y \in [x']$: questo significa che $y \sim x'$. Combinando con $x' \sim x$ (l'ipotesi sarebbe $x \sim x'$, ma è la stessa cosa visto che una relazione di equivalenza è anche simmetrica) e applicando la transitività si ottiene $y \sim x$, ovvero $y \in [x]$. Quindi è vero anche che ogni elemento di $[x']$ sta in $[x]$, cioè abbiamo finalmente $[x] = [x']$. □

Le classi di equivalenza hanno l'importante proprietà di ripartire l'insieme X su cui è definita la relazione in sottoinsiemi non vuoti e disgiunti, come affermato dal seguente risultato.

Proposizione 2.10. *Data una relazione di equivalenza $\sim$ su un insieme X, le classi di equivalenza hanno le seguenti tre proprietà:*

(i) *per ogni $x \in X$, si ha $[x] \neq \varnothing$,*

(ii) *l'unione $\bigcup_{x \in X} [x]$ di tutte le classi di equivalenza è uguale ad X e*

(iii) *due classi di equivalenza distinte sono disgiunte, ovvero $[x] \neq [x'] \Rightarrow [x] \cap [x'] = \varnothing$.*

Dimostrazione. La (i) si dimostra semplicemente ricordando (lo abbiamo visto nella dimostrazione della Proposizione 2.9) che x appartiene sempre alla sua classe di equivalenza $[x]$, quindi questa non è sicuramente vuota.

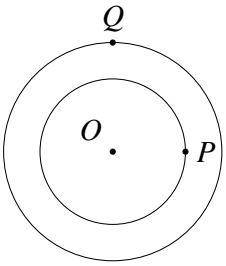
Per dimostrare la (ii), in base alla definizione di unione basta mostrare che ogni elemento di X sta in almeno una classe di equivalenza. Questo è vero per quanto appena detto, in quanto ogni elemento x sta almeno nella sua classe $[x]$.
Per dimostrare la (iii), dimostriamo la sua contronominale

$$[x] \cap [x'] \neq \varnothing \Rightarrow [x] = [x'] \tag{2.5}$$

che come sappiamo dal capitolo precedente è equivalente all'implicazione in (iii).
L'ipotesi $[x] \cap [x'] \neq \varnothing$ significa che esiste un z che appartiene sia a $[x]$ che a $[x']$. Allora, per definizione di classe , valgono sia $z \sim x$ (o, equivalentemente visto che la relazione è simmetrica, $x \sim z$) che $z \sim x'$. Mettendo insieme $x \sim z$ e $z \sim x'$, per la transitività abbiamo $x \sim x'$. Ma la Proposizione 2.9 ci dice che allora $[x] = [x']$, e la (2.5) è dimostrata. □

In generale, dato un insieme X, una famiglia di sottoinsiemi non vuoti, disgiunti e la cui unione sia tutto X si chiama una **partizione** di X. Quindi, la Proposizione 2.10 afferma che una relazione di equivalenza su X determina una partizione di X (in classi di equivalenza).

Esempio 2.11. Si consideri l'insieme X dei punti del piano, si fissi un punto O e per ogni $P, P' \in X$, si ponga $P \sim P'$ quando $|OP| = |OP'|$ (cioè quando i due punti sono alla stessa distanza da O).
È facile vedere che si tratta di una relazione di equivalenza. Infatti si ha $P \sim P$ in quanto ovviamente $|OP| = |OP|$. Inoltre $P \sim P'$ implica $P' \sim P$ in quanto $|OP| = |OP'|$ implica (per simmetria della relazione di uguaglianza) $|OP'| = |OP|$. Infine se $P \sim P'$ e $P' \sim P''$, ovvero $|OP| = |OP'|$ e $|OP'| = |OP''|$, visto che l'uguaglianza gode della proprietà transitiva si ha $|OP| = |OP''|$, ovvero $P \sim P''$.
Chi sono le classi di equivalenza di questa relazione? Dato un punto P, che dista $|OP| = R$ da O, la sua classe di equivalenza è data da tutti i punti che distano anch'essi R da O, ovvero è la circonferenza centrata in O su cui si trova P



Questo esempio illustra come le classi di equivalenza diano una partizione dell'insieme dato. In questo caso le circonferenze costituiscono una partizione del piano. Due cir-

conferenze di raggio diverso sono disgiunte e la loro unione, al crescere del raggio, ci dà tutto il piano[2].

Esempio 2.12. Se la relazione di equivalenza è l'uguaglianza, ogni classe di equivalenza contiene un solo elemento, cioè $[x] = \{x\}$: in questo caso la partizione è quella in singoli elementi.

Osservazione 2.13. Notiamo che se definissimo anche per una relazione $\sim$ che non sia di equivalenza le classi mediante la (2.4), ovvero dicendo che la classe $[x]$ di x è l'insieme di tutti gli elementi in relazione con x, non sarebbe più vero che tali classi determinano una partizione.
Ad esempio, consideriamo un insieme di quattro individui a_1, a_2, a_3, a_4 nel quale a_2 e a_3 sono figli di a_1 e a_4 è figlio di a_2. Rispetto alla relazione "essere figlio di", allora, la classe di a_1 (cioè l'insieme degli x tali che x è figlio di a_1) è $\{a_2, a_3\}$, la classe di a_2 è $\{a_4\}$, mentre le classi di a_3 e a_4 sono vuote, in quanto non esistono figli di a_3 e a_4 nell'insieme dato. Quindi non vale la (i) della Proposizione 2.10 (ci sono classi vuote) e neanche la (ii), in quanto unendo le classi $\{a_2, a_3\}$, $\{a_4\}$ e le due classi vuote si ottiene $\{a_2, a_3, a_4\}$, che non è tutto l'insieme.

Per vedere invece un esempio di relazione non di equivalenza in cui non vale la (iii) della Proposizione 2.10, si prenda ad esempio l'insieme $\mathbb{N}$ dei numeri naturali e la relazione $<$ (minore). Allora, la classe di 0 è vuota (non esiste nessun $x < 0$), la classe di 1 è $\{0\}$ (zero è l'unico numero naturale per cui $x < 1$ è verificata), la classe di 2 è $\{0, 1\}$, di 3 è $\{0, 1, 2\}$ e così via. Come si vede, ci sono classi diverse che non sono disgiunte, diversamente da quanto affermato dalla (iii) della Proposizione 2.10 nel caso di relazioni di equivalenza.

Nel prossimo capitolo studieremo un'altra importante relazione di equivalenza sull'insieme $\mathbb{Z}$ degli interi che ci porterà a comprendere alcuni moderni metodi crittografici.

Esempio 2.14. Per dare un'anticipazione, illustrata da un esempio concreto, prendiamo l'insieme degli interi $\mathbb{Z}$ e supponiamo che lo 0 rappresenti la mezzanotte di un dato giorno fissato, mentre ogni intero positivo (risp. negativo) rappresenti il numero di ore successive (risp. precedenti) alla mezzanotte.
Decidiamo di considerare equivalenti due interi x e y se essi corrispondono alla stessa ora del giorno (eventualmente di giorni diversi): definiamo cioè la relazione (nel capitolo successivo, dimostreremo che si tratta di una relazione di equivalenza[3]) $x \sim y$ se e solo se la differenza tra x e y è un multiplo di 24.
Ad esempio, 0, 24, 48, -24 sono equivalenti in quanto rappresentano la mezzanotte (rispettivamente del giorno in questione, del giorno successivo, di due giorni dopo e del

[2] Si noti che la classe di equivalenza di O contiene solo O, ovvero è una circonferenza di raggio 0.

[3] Anzi dimostreremo più in generale che, fissato un qualunque k, se per due interi n, m si definisce $n \sim m$ se e solo se la differenza $n - m$ è un multiplo di k questo dà una relazione di equivalenza.

giorno precedente). Gli interi $+1$, $+25$, $+49$, -23 sono equivalenti in quanto rappresentano l'una del mattino (rispettivamente del giorno in questione, del giorno successivo, di due giorni dopo e del giorno precedente).

Esempio 2.15. Si noti che se al posto di 24 avessimo scelto 2, ovvero avessimo definito due numeri sono equivalenti se e solo se la loro differenza è un multiplo di 2, avremmo avuto due sole classi di equivalenza: la classe di 0, data da $\{0, \pm 2, \pm 4, \pm 6 \dots\}$, e la classe di 1, data da $\{\pm 1, \pm 3, \pm 5, \pm 7 \dots\}$. In altre parole, la partizione di $\mathbb{Z}$ determinata da questa relazione è quella in numeri pari e numeri dispari.

2.4 Relazioni d'ordine e grafi

Vediamo ora un'altra importantissima classe di relazioni, le *relazioni d'ordine*.

Definizione 2.16. Una relazione $\sim$ su un insieme X si dice **relazione d'ordine** se soddisfa le seguenti tre proprietà:

(1) $x \sim x$ per ogni $x \in X$ (ovvero $\sim$ è riflessiva),

(2) se $x \sim y$ e $y \sim x$, allora $x = y$ (ovvero $\sim$ è antisimmetrica) e

(3) se $x \sim y$ e $y \sim z$, allora $x \sim z$ (ovvero $\sim$ è transitiva).

Esempio 2.17. La relazione di "minore o uguale" (denotata con $\leqslant$) sull'insieme dei numeri naturali $\mathbb{N} = \{0, 1, 2, 3, \dots\}$ è un esempio di relazione d'ordine, come è facile verificare.

Esempio 2.18. Un altro importante esempio di relazione d'ordine è dato dall'inclusione di sottoinsiemi. Più precisamente, si consideri un insieme X e l'insieme $P(X)$ dei suoi sottoinsiemi (cioè il cosiddetto **insieme potenza** di X). La relazione d'inclusione $A \subseteq B$, dove $A, B \in P(X)$ (cioè A e B sono sottoinsiemi di X) è una relazione d'ordine. Infatti,

(1) Si ha sempre $A \subseteq A$, in quanto tale inclusione, per definizione stessa di inclusione tra insiemi, significa che per ogni x vale $x \in A \Rightarrow x \in A$, che è un'implicazione sempre vera. Quindi vale la proprietà riflessiva.

(2) se $A \subseteq B$ e $B \subseteq A$, allora la prima inclusione ci dice che ogni elemento di A sta anche in B, e la seconda che viceversa ogni elemento di B sta anche in A, quindi per definizione di uguaglianza tra insiemi si ha $A = B$. Vale quindi anche la proprietà antisimmetrica.

(3) Infine, se $A \subseteq B$ e $B \subseteq C$, discende che $A \subseteq C$ (e quindi vale la proprietà transitiva). Infatti se $x \in A$, allora dalla prima inclusione $A \subseteq B$ deduciamo che $x \in B$ e, usando la seconda inclusione, deduciamo che $x \in C$. Avendo quindi dimostrato che $x \in A \Rightarrow x \in C$, concludiamo che $A \subseteq C$, come richiesto.

Questi due semplici esempi di relazioni d'ordine presentano alcune interessanti caratteristiche e significative differenze tra loro.

Ad esempio, osserviamo che in $\mathbb{N}$ con la relazione $\leqslant$ di minore o uguale, per due numeri $x, y \in \mathbb{N}$ vale sempre $x \leqslant y$ o $y \leqslant x$. Due elementi che soddisfano questa proprietà rispetto ad una relazione d'ordine si dicono *confrontabili*.
D'altro canto questo non si verifica sempre in $P(X)$ dotato della relazione $\subseteq$ di inclusione. Ad esempio se $X = \{a, b, c\}$, l'insieme dei sottoinsiemi di X è dato da

$$P(X) = \{\varnothing, \{a\}, \{b\}, \{c\}, \{a, b\}, \{a, c\}, \{b, c\}, \{a, b, c\}\}\,.$$

In questo caso esistono elementi *non confrontabili*. Ad esempio, per $\{a, b\}$ e $\{a, c\}$ non si ha né $\{a, b\} \subseteq \{a, c\}$ né $\{a, c\} \subseteq \{a, b\}$.
Nelle definizioni che daremo in seguito, scegliamo per ragioni di convenienza didattica di indicare la generica relazione d'ordine sempre con il simbolo "$\leqslant$", anche se non si tratterà necessariamente della relazione usuale di minore o uguale tra numeri (quando si tratterà invece di quest'ultima lo specificheremo esplicitamente).

Definizione 2.19. Una relazione d'ordine $\leqslant$ su insieme X si dice **ordine totale** se due elementi di X sono sempre confrontabili rispetto a $\leqslant$ (ovvero se, dati comunque $x_1, x_2 \in X$, si ha sempre o $x_1 \leqslant x_2$ o $x_2 \leqslant x_1$). In tal caso, l'insieme X dotato della relazione $\leqslant$ si dice **totalmente ordinato**.

Quindi possiamo dire che $\mathbb{N}$, dotato della relazione d'ordine usuale $\leqslant$ di minore o uguale tra numeri, è un insieme totalmente ordinato, mentre l'insieme $P(X)$ dei sottoinsiemi di un insieme dato, dotato della relazione d'ordine $\subseteq$, non è totalmente ordinato.

Un'altra importante differenza tra le relazioni $\leqslant$ su $\mathbb{N}$ e $\subseteq$ su $P(X)$ consiste nell'esistenza di *massimi e minimi*.

Definizione 2.20. Data una relazione d'ordine $\leqslant$ su insieme X, un elemento $x_0 \in X$ si dice **massimo** se per ogni $x \in X$ si ha $x \leqslant x_0$. Analogamente, x_0 si dice **minimo** se per ogni $x \in X$ si ha $x_0 \leqslant x$

Ad esempio, $\mathbb{N}$ dotato della relazione usuale $\leqslant$ di minore o uguale tra numeri ammette minimo, cioè lo zero 0, in quanto per ogni $x \in \mathbb{N}$ si ha $0 \leqslant x$. D'altra parte $\mathbb{N}$ non ammette massimo perché non esiste nessun numero naturale x_0 tale che $x \leqslant x_0$ per ogni $x \in \mathbb{N}$ (non esiste il numero naturale più grande).
Invece, qualunque sia X, l'insieme dei suoi sottoinsiemi $P(X)$ dotato della relazione di inclusione ha sempre sia minimo (dato dall'insieme vuoto $\varnothing$, in quanto per ogni altro sottoinsieme A si X si ha $\varnothing \subseteq A$) che massimo (dato da X stesso, in quanto per ogni altro sottoinsieme A si X si ha $A \subseteq X$)

Esempio 2.21. Consideriamo l'insieme

$$P' = \{\ \{a\}, \{b\}, \{c\}, \{a, b\}, \{a, c\}, \{b, c\}\ \}$$

ottenuto dall'insieme $P(X)$ dei sottoinsiemi di $X = \{a, b, c\}$ eliminando l'insieme vuoto e X stesso, e consideriamo su P' sempre la relazione $\subseteq$ d'inclusione. Allora, si vede che non c'è nessun S in P' tale che $Z \subseteq S$ per ogni Z di P', ovvero non c'è nessun massimo. Analogamente, si vede che non c'è nessun S in P' tale che $S \subseteq Z$ per ogni Z di P', ovvero non c'è nessun minimo.
Tuttavia, l'elemento $S = \{b, c\}$ ha la notevole proprietà che *non esiste nessun Z in P' diverso da S per cui $S \subseteq Z$*. Quindi, benché questo elemento S non possa essere considerato un massimo in base alla Definizione 2.20 data sopra, non esiste nessun elemento che sia "maggiore di lui" (rispetto alla relazione d'inclusione). Potremmo quindi considerare S una sorta di massimo in un senso più largo.

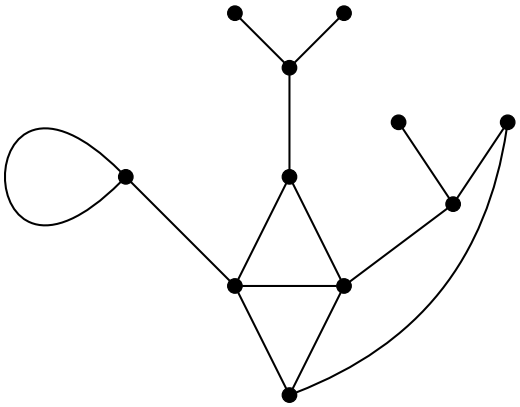
Definizione 2.22. Data una relazione d'ordine $\leqslant$ su insieme X, un elemento $x_0 \in X$ si dice **massimale** se non esiste nessun $y \in X$ diverso da x e tale che $x_0 \leqslant y$. Analogamente, un elemento $x_0 \in X$ si dice **minimale** se non esiste nessun $y \in X$ diverso da x e tale che $y \leqslant x_0$.

Equivalentemente, possiamo dire un elemento x_0 massimale se $x_0 \leqslant y$ implica $y = x_0$, o minimale se $y \leqslant x_0$ implica $y = x_0$.
Ad esempio, sempre nell'insieme $P' = \{\ \{a\}, \{b\}, \{c\}, \{a, b\}, \{a, c\}, \{b, c\}\ \}$ si ha che $\{a, b\}, \{a, c\}, \{b, c\}$ sono massimali, mentre $\{a\}, \{b\}, \{c\}$ sono minimali. Si noti quindi che in un insieme ordinato possiamo avere diversi elementi massimali e diversi elementi minimali (mentre, come si vede facilmente, se esiste un massimo o un minimo questo è unico).

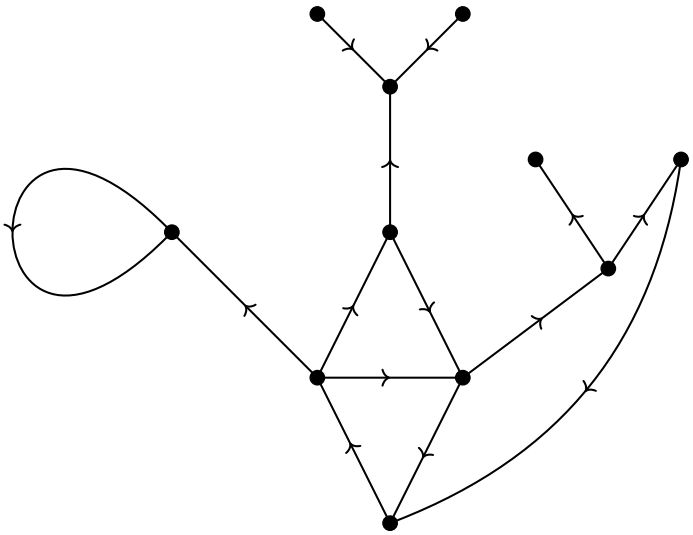
Osservazione 2.23. Si noti che la distinzione tra massimale e massimo (o tra minimale e minimo) ha senso se stiamo lavorando con un ordine non totale, cioè nel quale esistono elementi non confrontabili tra loro. Il motivo per cui $\{b, c\}$ è massimale ma non massimo in $P' = \{\ \{a\}, \{b\}, \{c\}, \{a, b\}, \{a, c\}, \{b, c\}\ \}$ rispetto alla relazione $\subseteq$ d'inclusione è che in P' ci sono elementi con i quali $\{b, c\}$ non è confrontabile, come ad esempio $\{a, c\}$. Quindi, pur non esistendo nessun elemento "maggiore" di $\{b, c\}$ non possiamo dire che $\{b, c\}$ sia maggiore di tutti gli altri, sicuramente non di quelli con i quali non è confrontabile.
Al contrario, in un ordine totale in cui dati due elementi x e y si ha sempre o $x \leqslant y$ o $y \leqslant x$, un massimale x_0 è automaticamente anche il massimo. Questo perché se per nessun y si ha $x_0 \leqslant y$ (come prevede la definizione di massimale), dovendo valere comunque una delle due disuguaglianze allora si ha necessariamente $y \leqslant x_0$ per ogni y, e quindi x_0 è in effetti un massimo.

Un modo efficace per visualizzare un insieme ordinato e l'eventuale presenza di massimali, minimali, massimi o minimi è quello di rappresentarlo mediante un *grafo*.

Definizione 2.24. Un **grafo** è un insieme di punti, detti *vertici*, e di segmenti di curva, detti *lati*, che collegano coppie di vertici.[4]



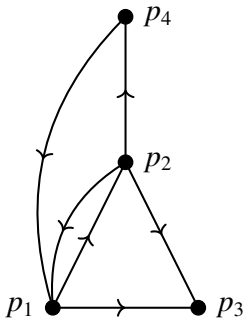
Si noti che un grafo può presentare **cappi**, ovvero segmenti che collegano un vertice con se stesso.
Se i lati del grafo sono dotati di un verso di percorrenza (simboleggiato da una freccia) si parla di **grafo orientato.**



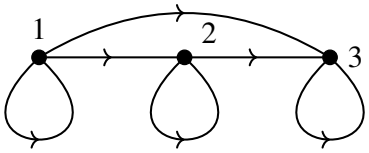
Una qualunque relazione $\sim$ su un insieme X può essere rappresentata mediante un grafo orientato nel modo seguente. Si rappresentano gli elementi di X mediante punti (i vertici del grafo) e si traccia un segmento orientato dal punto che rappresenta x al punto che rappresenta x' se e solamente se $x \sim x'$, ovvero x è in relazione con x'. Si osservi che

[4] Questa non è una definizione rigorosa di grafo ma è sufficiente ai nostri scopi. In generale un grafo è definito come una coppia di insiemi: quello dei vertici e quello dei lati.

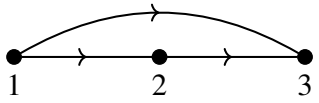
orientare il segmento è necessario per capire se $x \sim x'$ o $x' \sim x$ per cui una relazione d'ordine definisce sempre un grafo orientato. Ad esempio, supponiamo di avere un insieme $X = \{p_1, p_2, p_3, p_4\}$ di pagine internet. Vogliamo definire la relazione: $p \sim p'$ se e solo se esiste un link dalla pagina p alla pagina p'. Se ad esempio da p_1 abbiamo un link verso p_2 e uno verso p_3, da p_2 un link verso p_1, uno verso p_3 e uno verso p_4, da p_3 nessun link e da p_4 solo un link verso p_1, allora il grafo orientato che rappresenta la relazione corrispondente è



Parlando in particolare di relazioni d'ordine, se consideriamo ad esempio su $X = \{1, 2, 3\}$ l'usuale relazione d'ordine $\leqslant$ di minore o uguale, allora il grafo corrispondente è



Si noti che da ogni vertice parte un cappio, in quanto ogni elemento x è in relazione $x \leqslant x$ con se stesso (le relazione è riflessiva). Se, sapendo che la relazione è d'ordine e quindi riflessiva, sottointendiamo la presenza di questi cappi, potremmo semplificare il grafo nel modo seguente



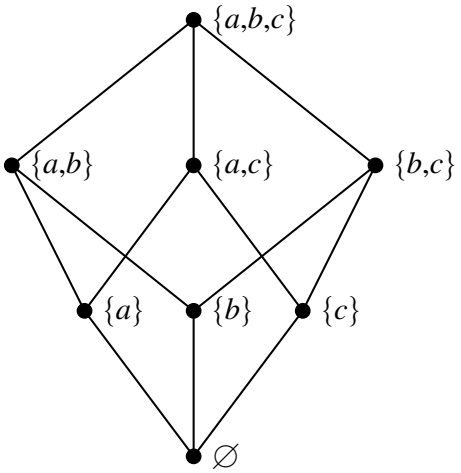
Un'ulteriore semplificazione può essere fatta osservando che, essendo la relazione d'ordine transitiva, la presenza del segmento orientato da 1 a 3 può essere dedotta dalla presenza del segmenti orientati da 1 a 2 e da 2 a 3 (infatti, per la transitività $1 \leqslant 2$ e $2 \leqslant 3$ implica $1 \leqslant 3$). Sottointendendo quindi anche i segmenti che sono conseguenza della transitività, il grafico diventa

Infine, possiamo pensare di evitare di orientare i segmenti disponendo il grafico in modo che i versi degli stessi vadano sempre pensati dal basso verso l'alto:

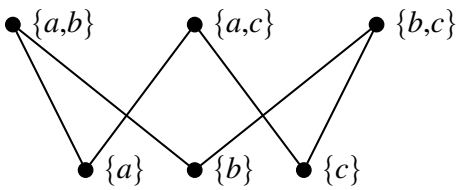
Il grafo di una relazione d'ordine così semplificato prende anche il nome di *diagramma di Hasse*.
Ad esempio, consideriamo di nuovo l'insieme

$$P(X) = \{\varnothing, \{a\}, \{b\}, \{c\}, \{a,b\}, \{a,c\}, \{b,c\}, \{a,b,c\}\}$$

dei sottoinsiemi di $X = \{a,b,c\}$ con la relazione d'ordine data dall'inclusione $\subseteq$. Allora, il suo diagramma di Hasse è



Come si vede, l'insieme ordinato possiede un massimo (il vertice più alto del diagramma) e un minimo (il vertice più basso del diagramma). Se ci limitassimo alla relazione d'inclusione sui sottoinsiemi $\{a\}, \{b\}, \{c\}, \{a,b\}, \{a,c\}, \{b,c\}$, il diagramma sarebbe invece



nel quale non esiste un solo vertice più alto o più basso, ma più vertici nel livello superiore e in quello inferiore (che rappresentano rispettivamente gli elementi massimali o minimali).

Osservazione 2.25. L'estrema importanza dei grafi in matematica e nelle sue applicazioni è giustificata dall fatto che un grafo può modellizzare numerose situazioni e permettere di visualizzare geometricamente una grande varietà di problemi. Una rete di trasporti (ad esempio la cartina delle strade di una città con i suoi incroci) può essere rappresentata da un grafo ordinato, un albero genealogico è un grafo, un circuito stampato può essere pensato come un grafo e molti altri.
Queste situazioni concrete hanno ispirato una serie di problemi e domande relativi ai grafi, quali: dato un grafo, è possibile percorrerlo tramite un cammino continuo che tocchi tutti i suoi vertici una sola volta ? (si parla di *cammino hamiltoniano*); o ancora, è possibile percorrerlo tramite un cammino continuo che tocchi tutti i suoi lati una sola volta ? (si parla di *cammino euleriano*). Questi e tanti altri problemi (alcuni risolti, altri ancora aperti) sono affrontati da una branca della matematica detta appunto teoria dei grafi.

2.5 Funzioni

Vedremo ora un'altra nozione fondamentale per tutta la matematica, quella di *funzione*.

Definizione 2.26. Dati due insiemi X e Y, una **funzione** f da X a Y (si scrive $f\colon X \longrightarrow Y$) è una legge che assegna ad ogni elemento x di X un unico elemento y di Y. L'elemento y è detto **immagine** di x e si denota con $f(x)$. Per indicare che f associa $f(x)$ ad x si scrive anche $x \mapsto f(x)$. Il tutto è espresso dalla scrittura compatta

$$\begin{aligned} f\colon X &\longrightarrow Y \\ x &\mapsto f(x) \end{aligned}$$

L'insieme X di partenza si dice **dominio** della funzione, quello Y di arrivo **codominio**.

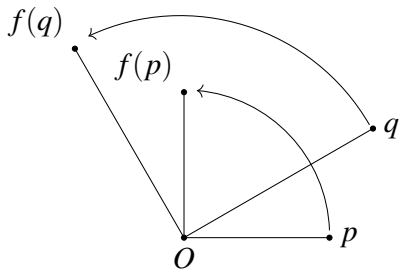
Esempi 2.27. Sia X l'insieme degli studenti presenti in un'aula e sia Y l'insieme $\mathbb{N}$ dei numeri naturali. Associando a ogni studente x il suo numero di matricola, si ottiene una funzione $X \longrightarrow \mathbb{N}$.

In questo esempio, la funzione rappresenta l'assegnazione di un dato a ogni entrata di una lista.

Le funzioni possono essere anche usate per descrivere l'andamento di un dato fenomeno. Si consideri ad esempio un corpo che a un certo istante di tempo parte da un punto O e inizia a muoversi lungo una traiettoria rettilinea. Identificando i punti della retta con i numeri reali con il punto O in corrispondenza dello zero, come descritto nell'Osservazione 2.3, possiamo assegnare a ogni $t \in \mathbb{R}$ la posizione $x(t)$ del corpo sulla retta lungo cui si muove dopo un tempo t dall'istante iniziale (se t è positivo, altrimenti sia il tempo precedente all'istante iniziale). Questo definisce una funzione $\mathbb{R} \longrightarrow \mathbb{R}$ che descrive il moto del corpo.

Ancora, le funzioni possono rappresentare delle operazioni. Ad esempio la funzione $f\colon \mathbb{Z} \longrightarrow \mathbb{Z}$ che assegna a ogni numero intero n il suo quadrato $f(n) = n^2$. Non otteniamo invece una funzione $\mathbb{Z} \longrightarrow \mathbb{Z}$ definendo $f(n) = \sqrt{n}$, perché alcuni elementi del dominio non hanno immagine nel codominio,. Infatti se $n = 2$ o $n = -1$, $f(n) = \sqrt{2}$ e $f(n) = \sqrt{-1}$ non sono numeri interi (il secondo non è neanche un numero reale).

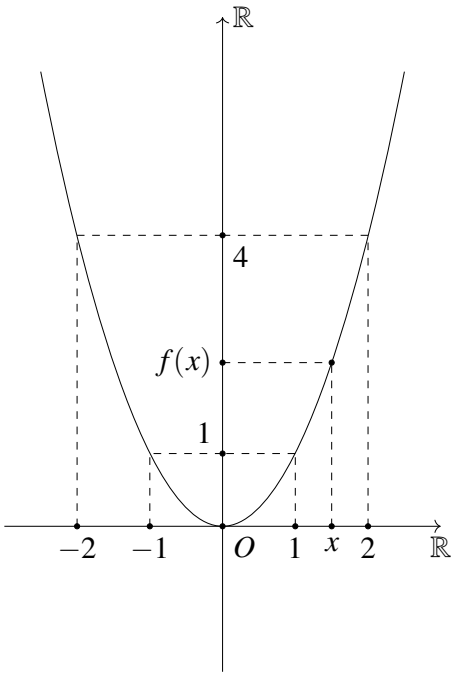
Infine, come ultimo esempio, le funzioni possono rappresentare trasformazioni di tipo geometrico. Se X è l'insieme dei punti del piano, scelto un punto O possiamo definire una funzione $X \longrightarrow X$ assegnando a ogni punto P il punto $f(P)$ che si ottiene ruotando P attorno a O di un angolo di 90 gradi in senso antiorario.



Osservazione 2.28. La definizione di funzione data sopra è sufficiente per gli scopi di questo corso ma non è la definizione rigorosa e formale di funzione, che afferma che una funzione da un insieme X a un insieme Y è un sottoinsieme S del prodotto cartesiano $X \times Y$ che per ogni $x \in X$ contiene una e una sola coppia (x, y) che ha x come prima componente (in simboli[5], $\forall x \in X\ \exists! y \in Y\ (x, y) \in S$)
In pratica, l'idea è di rappresentare l'informazione che a x stiamo assegnando un certo $y \in Y$ tramite la coppia (x, y). La richiesta che per ogni x esista un'unica coppia che ha x come primo elemento traduce il fatto che a ogni x dobbiamo assegnare un ben preciso y (uno e uno solo).

[5] Il quantificatore universale $\exists$, seguito dal punto esclamativo, significa "esiste un unico".

L'idea che una funzione $f\colon X \longrightarrow Y$ sia definita da un sottoinsieme del prodotto cartesiano $X \times Y$ non è tanto sorprendente se si pensa che ogni funzione $f\colon \mathbb{R} \longrightarrow \mathbb{R}$ può essere rappresentata mediante il suo grafico, che può essere pensato proprio come un sottoinsieme del prodotto cartesiano $\mathbb{R} \times \mathbb{R}$. Più precisamente, se come già descritto nell'Osservazione 2.3 consideriamo un sistema di coordinate per cui a ogni punto del piano corrisponde una coppia (x, y) di numeri reali e viceversa (e quindi il piano si identifica con il prodotto cartesiano $\mathbb{R} \times \mathbb{R}$), possiamo rappresentare una funzione $f\colon \mathbb{R} \longrightarrow \mathbb{R}$ disegnando l'insieme dei punti che hanno come prima coordinata x e come seconda coordinata l'immagine $f(x)$ (per ogni $x \in \mathbb{R}$).
Ad esempio, se la funzione fosse $f(x) = x^2$ si avrebbe



L'insieme di punti così ottenuto (nel disegno dato la parabola) è quindi un sottoinsieme del piano, che è come dire un sottoinsieme del prodotto cartesiano $\mathbb{R}\times\mathbb{R}$, che rappresenta la funzione.

Consideriamo ora alcune proprietà delle funzioni.
Iniziamo con l'osservare che la definizione di funzione non impedisce che due elementi diversi del dominio abbiano la stessa immagine. Ad esempio, la funzione quadrato $f\colon \mathbb{Z} \longrightarrow \mathbb{Z}$ considerata sopra che assegna a ogni $n \in \mathbb{Z}$ l'intero $f(n) = n^2$ è tale che $f(2) = f(-2) = 4$.

Questo non accade ad esempio per la funzione $X \longrightarrow \mathbb{N}$ che assegna a ogni studente di una certa aula il suo numero di matricola, in quanto non è possibile che studenti diversi abbiano lo stesso numero di matricola.
Alla luce di questi due esempi, diamo la seguente

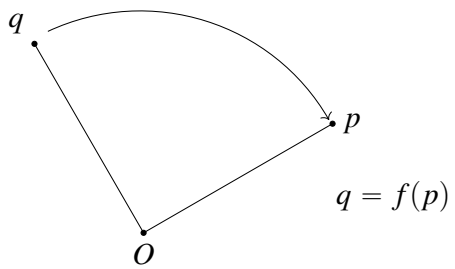
Definizione 2.29. Una funzione $f\colon X \longrightarrow Y$ si dice **iniettiva** se due diversi elementi del dominio hanno sempre immagini diverse. In simboli

$$x \neq x' \Rightarrow f(x) \neq f(x')\,.$$

Quindi, la funzione che assegna a ogni studente il suo numero di matricola è iniettiva, mentre la funzione quadrato da $\mathbb{Z}$ in $\mathbb{Z}$ non è iniettiva.

Osservazione 2.30. Usando il fatto che ogni implicazione è equivalente alla sua contronominale, possiamo anche dire che una funzione f è iniettiva se per ogni $x, x' \in X$ si ha $f(x) = f(x') \Rightarrow x = x'$ (ovvero se l'unica possibilità per cui x e x' abbiano la stessa immagine sia che $x = x'$).
Tale riformulazione è spesso più conveniente quando si vuole dimostrare che una funzione data è iniettiva. Se volessi verificare ad esempio che la funzione $f\colon \mathbb{N} \longrightarrow \mathbb{N}$ data da $f(n) = n + 1$ è iniettiva, dovrei dimostrare l'implicazione $f(n) = f(m) \Rightarrow n = m$, ovvero $n+1 = m+1 \Rightarrow n = m$. Ma questo è quasi immediato in quanto da $n+1 = m+1$, portando l'1 del primo membro a secondo si ottiene subito $n = m + 1 - 1 = m$.

Un'altra particolarità della funzione $X \longrightarrow \mathbb{N}$ che assegna a ogni studente il suo numero di matricola e della funzione quadrato $\mathbb{Z} \longrightarrow \mathbb{Z}$ è che non tutti gli elementi del codominio sono immagine di qualche elemento del dominio. Cioè non tutti i numeri naturali sono numeri di matricola di qualche studente, esattamente come non tutti gli interi sono quadrati di qualche intero di partenza (ad esempio, 5 non è quadrato, ovvero non è della forma n^2 per nessun $n \in \mathbb{Z}$).
Al contrario, questo non accade, ad esempio, per la rotazione che abbiamo definito nell'Esempio 2.27, cioè la funzione $f\colon X \longrightarrow X$ dal piano X in se stesso e che manda ogni punto P nel punto $f(P)$ che si ottiene ruotando P di 90 gradi in senso antiorario. Infatti qualunque punto Q del codominio X è immagine $f(P)$ di qualche punto P del dominio (basta prendere come P il punto che si ottiene ruotando Q in senso *orario*, come si vede nel disegno seguente)



Diamo allora la seguente

Definizione 2.31. Una funzione $f\colon X \longrightarrow Y$ si dice **suriettiva** se per ogni $y \in Y$ esiste un $x \in X$ tale che $y = f(x)$.

Quindi, sia la funzione $X \longrightarrow \mathbb{N}$ che assegna a ogni studente il suo numero di matricola che la funzione quadrato su $\mathbb{Z}$ sono non suriettive, mentre la rotazione del piano in sè è suriettiva.
In effetti, quest'ultima è anche iniettiva. Tali funzioni sono particolarmente importanti.

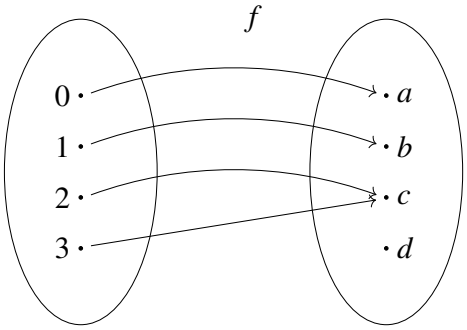
Definizione 2.32. Una funzione $f\colon X \longrightarrow Y$ si dice **biiettiva** se è sia iniettiva che suriettiva.

Una funzione biiettiva si chiama anche *biiezione* o *corrispondenza biunivoca*.

Un modo alternativo di definire l'iniettività è usare il concetto di *controimmagine*. Data una funzione $f\colon X \longrightarrow Y$ e un elemento $y \in Y$, si dice controimmagine di y (e si denota con $f^{-1}(y)$) l'insieme di tutti gli elementi di X a cui F associa y. In simboli

$$f^{-1}(y) = \{x \in X \mid f(x) = y\}\,.$$

Ad esempio, si consideri la funzione $f\colon \{0,1,2,3\} \longrightarrow \{a,b,c,d\}$ rappresentata nel seguente disegno



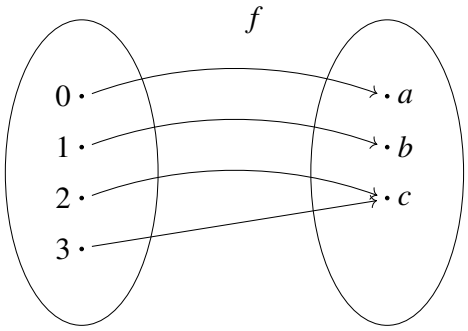
Si ha $f^{-1}(a) = \{0\}$, $f^{-1}(b) = \{1\}$, $f^{-1}(c) = \{2,3\}$, $f^{-1}(d) = \varnothing$.
Come in questo esempio la controimmagine di un elemento di Y potrebbe essere vuota. *In effetti una funzione è suriettiva se e solo se non esistono elementi $y \in Y$ tali che nessun $x \in X$ soddisfa $f(x) = y$.* Chiaramente, una funzione è iniettiva se non succede mai che $f^{-1}(y)$ abbia più di un elemento, perchè in tal caso vorrebbe dire che ci sono elementi diversi di X che vanno a finire in y. In altre parole, possiamo dire che *una funzione $f\colon X \longrightarrow Y$ è iniettiva se e solo se tutte le controimmagini degli elementi di Y hanno al massimo un elemento.*

Anche la suriettività ammette una definizione alternativa, basata sul concetto di *immagine della funzione*. L'immagine di una funzione $f\colon X \longrightarrow Y$ (denotata $Im(f)$) è semplicemente l'insieme di tutte le immagini degli elementi di X, in simboli

$$Im(f) = \{y \in Y \mid y = f(x) \text{ per qualche } x \in X\}$$

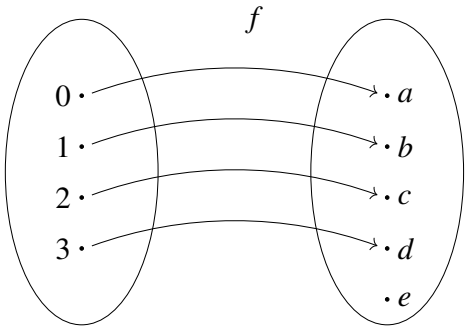
Ad esempio, nella funzione $f\colon \{0,1,2,3\} \longrightarrow \{a,b,c,d\}$ vista sopra si ha $Im(f) = \{a,b,c\}$ (solo d non si scrive come immagine di qualche elemento del dominio). Allora, è chiaro che una funzione è suriettiva solo se $Im(f) = Y$, in quanto questa uguaglianza si verifica esattamente quando ogni elemento $y \in Y$ si scrive come $y = f(x)$ per qualche $x \in X$, che è proprio la definizione di suriettività.

Osserviamo che iniettività e suriettività sono nozioni indipendenti: una funzione può essere iniettiva e non suriettiva, suriettiva e non iniettiva, nessuna delle due o entrambe. Ad esempio, la funzione $f\colon \{0,1,2,3\} \longrightarrow \{a,b,c\}$ rappresentata dal seguente disegno



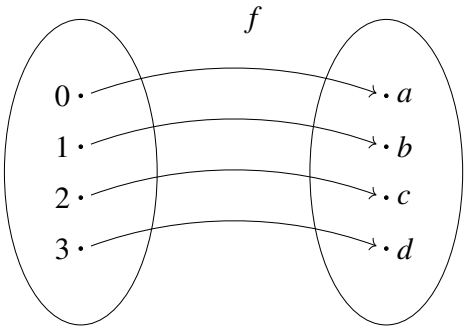
(ovvero $f(0) = a$, $f(1) = b$, $f(2) = c$, $f(3) = c$) è suriettiva (in quanto ogni elemento a,b,c del codominio è immagine di qualche elemento del dominio) ma non iniettiva in quanto esistono due elementi diversi del dominio (ossia 2 e 3) che hanno la stessa immagine.

Invece, la funzione $f\colon \{0,1,2,3\} \longrightarrow \{a,b,c,d,e\}$ rappresentata dal seguente disegno



è iniettiva ma non suriettiva (c'è un elemento del codomino, ovvero e, che non è immagine di nessun elemento del dominio).

Infine, la funzione $f\colon \{0,1,2,3\} \longrightarrow \{a,b,c,d\}$ rappresentata dal seguente disegno



è sia iniettiva che suriettiva, ovvero biiettiva.

Osserviamo che non c'è nessuna possibile scelta delle immagini di $0,1,2,3$ nella prima delle tre funzioni precedenti che la renda una funzione iniettiva. Infatti, essendoci più elementi nel dominio che nel codominio, almeno due elementi dovranno avere per forza la stessa immagine. In generale, dati due insiemi finiti X e Y, il primo con $|X|$ elementi e il secondo con $|Y|$ elementi, abbiamo che

$$|X| > |Y| \Rightarrow f\colon X \longrightarrow Y \text{ non è iniettiva} \tag{2.6}$$

o equivalentemente, usando la contronominale,

$$f\colon X \longrightarrow Y \text{ è iniettiva} \Rightarrow |X| \leqslant |Y|\,. \tag{2.7}$$

Osservazione 2.33. La (2.6) si chiama anche *principio della piccionaia* per il seguente motivo. Supponiamo che X sia un insieme di piccioni e Y un insieme di gabbie. Se dobbiamo mettere ogni piccione in una gabbia e ci sono più piccioni che gabbie almeno due piccioni dovranno essere messi nella stessa gabbia. A dispetto della sua elementarità può essere usato per dimostrare fatti a prima vista non banali.

Ad esempio, usando tale principio possiamo dimostrare che nel mondo esistono almeno due persone con lo stesso numero di capelli. Infatti, sia X l'insieme delle persone del mondo, che quindi diciamo contiene 8 miliardi di elementi, e consideriamo la funzione f che assegna a ogni persona il numero dei suoi capelli. Stimando che il numero massimo N di capelli che possono essere contenuti in una testa sia sicuramente minore di 8 miliardi, la funzione f sarà una funzione $X \longrightarrow Y = \{0,1,2,3,\ldots,N\}$ tale che $|X| > |Y|$. In base al principio enunciato f non può essere iniettiva, ovvero esistono almeno due elementi del dominio (due persone) con la stessa immagine (lo stesso numero di capelli).

Analogamente, per quanto riguarda la seconda delle tre funzioni rappresentate nei disegni sopra, ovvero la $f\colon \{0,1,2,3\} \longrightarrow \{a,b,c,d,e\}$, non c'è nessuna possibile scelta alternativa delle immagini che la renda una funzione suriettiva. Infatti, essendoci nel dominio meno elementi che nel codominio, non è possibile raggiungere tutti gli elementi del codominio facendo partire una freccia da ciascun elemento del dominio. Analogamente a sopra, possiamo dire che in generale dati due insiemi finiti X e Y, il primo con $|X|$ elementi e il secondo con $|Y|$ elementi, abbiamo che

$$|X| < |Y| \Rightarrow f\colon X \longrightarrow Y \text{ non è suriettiva} \tag{2.8}$$

o equivalentemente, usando la contronominale,

$$f\colon X \longrightarrow Y \text{ è suriettiva} \Rightarrow |X| \geqslant |Y|\,. \tag{2.9}$$

Dal momento che le funzioni biiettive sono sia iniettive che suriettive, combinando la (2.7) e la (2.9) otteniamo

$$f\colon X \longrightarrow Y \text{ è biiettiva} \Rightarrow |X| = |Y|\,. \tag{2.10}$$

Osservazione 2.34. La (2.10) ci dice che se esiste una funzione biiettiva tra due insiemi *finiti*, allora essi hanno lo stesso numero di elementi. Non è difficile verificare che vale il viceversa: se il numero di elementi di X e Y è lo stesso, allora non è difficile costruire una funzione biiettiva $X \longrightarrow Y$.

Nel caso in cui X e Y siano insiemi infiniti, parlare di numero di elementi sembra non avere più senso, o al più verrebbe spontaneo dire che due insiemi infiniti hanno lo stesso numero di elementi (infinito, appunto). Tuttavia, se decidiamo, procedendo per analogia con gli insiemi finiti, di dire anche per due insiemi infiniti X e Y che essi hanno lo stesso numero di elementi se esiste una funzione biiettiva $X \longrightarrow Y$, allora scopriremo che *non tutti gli insiemi infiniti hanno lo stesso numero di elementi* (detto anche *cardinalità*).

Per illustrare tale affermazione, usiamo come esempi gli insiemi numerici $\mathbb{N}$ (i numeri naturali), $\mathbb{Z}$ (i numeri interi), $\mathbb{Q}$ (i numeri razionali), $\mathbb{R}$ (i numeri reali)[6], che sono inclusi uno nel successivo:

$$\mathbb{N} \subset \mathbb{Z} \subset \mathbb{Q} \subset \mathbb{R}\,.$$

Iniziamo con il chiederci se esiste una funzione biiettiva $\mathbb{N} \longrightarrow \mathbb{Z}$. Questo significherebbe che è possibile associare a ogni naturale un intero in modo da non lasciare scoperto nessun intero e da non ripetere due volte lo stesso intero, o in altre parole mettere tutti gli interi in una successione infinita $a_0, a_1, a_2 \ldots$ senza ripetizioni.

[6] Pur non avendo definito formalmente questi insiemi numerici, possiamo portare avanti questa discussione ricordandoci che non ha la pretesa di essere rigorosa. L'insieme dei numeri naturali sarà definito rigorosamente nella Sezione 2.7.

La risposta è affermativa, ad esempio gli interi possono essere messi in sequenza come segue:

$$\begin{array}{cccccccc} 0 & 1 & 2 & 3 & 4 & 5 & 6 & \dots \\ \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \dots \\ 0 & 1 & -1 & 2 & -2 & 3 & -3 & \dots \end{array}$$

Dunque esiste una corrispondenza biunivoca tra $\mathbb{N}$ e $\mathbb{Z}$, ovvero possiamo dire che i naturali e gli interi hanno "lo stesso numero di elementi" o, più precisamente, la stessa cardinalità.
Mentre da una parte questo può risultare banale (si potrebbe pensare che avendo entrambi infiniti elementi, necessariamente questi debbano poter essere messi in corrispondenza biunivoca), dall'altra il risultato ha un aspetto sorprendente. Specificatamente, dal momento che $\mathbb{N}$ si identifica con un sottoinsieme di $\mathbb{Z}$ (quello degli interi non negativi), stiamo dicendo che $\mathbb{Z}$ ha lo stesso numero di elementi di un suo sottoinsieme proprio.[7]
Questo non è possibile per insiemi finiti: in effetti questa proprietà caratterizza gli insiemi infiniti ed è spesso assunto come definizione di insieme infinito. Ovvero, un insieme si dice infinito se può essere messo in corrispondenza biunivoca con un suo sottoinsieme proprio.
Inoltre possiamo mostrare la meno intuitiva esistenza di una biiezione tra i razionali $\mathbb{Q}$ e i naturali $\mathbb{N}$. Per mostrare come, consideriamo per facilità solo i razionali positivi (il ragionamento si modifica successivamente per accomodare tutti i razionali), che si scrivono come quoziente n/m di due naturali n, m (con $m \neq 0$).
Consideriamo la seguente tabella infinita di frazioni dove, nella riga n-esima, scriviamo le frazioni con numeratore uguale a n:

$$\begin{array}{c|cccccc} & 1 & 2 & 3 & 4 & 5 & \cdots \\ \hline 1 & 1/1 & 1/2 & 1/3 & 1/4 & 1/5 & \dots \\ 2 & 2/1 & 2/2 & 2/3 & 2/4 & 2/5 & \dots \\ 3 & 3/1 & 3/2 & 3/3 & 3/4 & 3/5 & \dots \\ 4 & 4/1 & 4/2 & 4/3 & 4/4 & 4/5 & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{array}$$

Chiaramente in questa tabella alcune frazioni rappresentano lo stesso razionale (1/2 e 2/4, oppure 1/1 e 2/2)) ma se riusciamo a mostrare che è possibile realizzare una corrispondenza biunivoca tra le entrate della tabella e i naturali, a maggior ragione questo sarà vero per i razionali che corrispondono alle entrate della tabella senza ripetizioni.
La corrispondenza cercata con le frazioni della tabella si può per esempio realizzare scrivendo in sequenza tutte le diagonali della tabella a partire dall'angolo in alto a sini-

[7] Si dice sottoinsieme proprio di un insieme X un sottoinsieme che non sia X stesso (in base alla definizione di sottoinsieme, vale sempre l'inclusione $X \subseteq X$, ovvero ogni insieme è sottoinsieme di se stesso).

stra: prima $1/1$, poi la diagonale adiacente $1/2, 2/1$, poi continuando a scendere verso destra la terza diagonale $1/3, 2/2, 3/1$ e così via.

$$1/1, 1/2, 2/1, 1/3, 2/2, 3/1, 1/4, 2/3, 3/2, 4/1, \ldots$$

In questo modo, si riesce a mettere in una sequenza (ovvero in corrispondenza con i numeri naturali) tutti i razionali, e realizzare una funzione biiettiva $\mathbb{N} \longrightarrow \mathbb{Q}$. Possiamo concludere che anche $\mathbb{Q}$ ha la stessa cardinalità di $\mathbb{N}$.

Questi esempi potrebbero generare l'idea che qualunque insieme infinito possa essere messo in corrispondenza biunivoca con l'insieme dei naturali. In realtà questo è falso: si dimostra che per l'insieme $\mathbb{R}$ dei numeri reali questa corrispondenza non esiste.
Un'idea della dimostrazione è la seguente: supponiamo per assurdo di poter mettere in una sequenza in corrispondenza biunivoca con i numeri naturali tutti i numeri reali: $x_0, x_1, x_2, x_3, \ldots$.
Rappresentando ogni reale con la sua rappresentazione decimale, possiamo riscrivere tale sequenza come

$$\begin{aligned} x_0 &= a_0, b_0 c_0 d_0 \ldots \\ x_1 &= a_1, b_1 c_1 d_1 \ldots \\ x_2 &= a_2, b_2 c_2 d_2 \ldots \\ x_3 &= a_3, b_3 c_3 d_3 \ldots \\ &\ldots \end{aligned}$$

Ora, mostreremo che in realtà una tale sequenza non può mai contenere tutti i numeri reali, e lo faremo costruendo esplicitamente un numero reale che non è contenuto nella sequenza, il numero

$$x = 0, bcd \ldots$$

in cui $b, c, d, \ldots$ sono definiti come segue. Poniamo $b = 1$ se $b_0 = 0$ e $b = 0$ se $b_0 \neq 0$ (questo garantisce già che x non sia il primo numero della sequenza x_0). Successivamente definiamo $c = 1$ se $c_1 = 0$ e $c = 0$ se $c_1 \neq 0$ (questo garantisce che x non sia il secondo numero della sequenza x_1). Ancora poniamo $d = 1$ se $d_2 = 0$ e $d = 0$ se $d_2 \neq 0$ (questo garantisce che x non sia il terzo numero della sequenza x_2). Ripetendo questo ragionamento per ogni decimale otterremo il numero desiderato.

Dunque, $\mathbb{R}$ non ha lo stesso numero di elementi di $\mathbb{N}$ (pur essendo entrambi gli insiemi infiniti), e poichè $\mathbb{R}$ contiene $\mathbb{N}$ possiamo dire che $\mathbb{R}$ ha una cardinalità strettamente maggiore di quella di $\mathbb{N}$.
Non esiste dunque un unico "infinito", ma infiniti di tipo diverso. È lecito chiedersi se $\mathbb{R}$ sia il maggiore infinito possibile. La risposta è no. Infatti, dato un qualunque insieme

infinito, ne esiste sicuramente sempre uno di cardinalità più grande. Più precisamente, Cantor (al quale si deve la teoria degli insiemi infiniti che stiamo illustrando) mostrò che dato un insieme X, l'insieme potenza $P(X)$ di X (cf. Esempio 2.18) ha sempre cardinalità maggiore di X. Quindi, ad esempio, se vogliamo un insieme che abbia cardinalità maggiore di $\mathbb{R}$, basta prendere l'insieme potenza $P(\mathbb{R})$ di $\mathbb{R}$.

Concludiamo dicendo che si dimostra che la cardinalità di $\mathbb{R}$ coincide con quella dell'insieme potenza di $\mathbb{N}$.

L'affermazione che non esistono insiemi di cardinalità compresa tra quella di $\mathbb{N}$ e quella di $\mathbb{R}$ si chiama *ipotesi del continuo*. Si può dimostrare che tale affermazione non è né dimostrabile né confutabile nella teoria assiomatica degli insiemi.

Teorema 2.35. *Dato un qualunque insieme X, il suo insieme potenza $P(X)$ ha cardinalità strettamente maggiore di quella di X.*

Dimostrazione. Dobbiamo mostrare che esiste una funzione iniettiva $X \longrightarrow P(X)$ ma non esiste una funzione biiettiva $X \longrightarrow P(X)$. Per quello che riguarda la prima affermazione, un esempio di funzione iniettiva da X a $P(X)$ è dato dalla funzione che manda ogni $x \in X$ nel sottoinsieme $\{x\}$ che contiene solamente x. L'iniettività di tale funzione è immediata in quanto è chiaro che se $x \neq x'$ allora i due sottoinsiemi $\{x\}$ e $\{x'\}$ sono necessariamente diversi.

Dimostriamo ora la seconda affermazione per assurdo, ovvero supponiamo che esista una funzione $f\colon X \longrightarrow P(X)$ biiettiva e mostriamo che questo porta a una contraddizione. Tale funzione associa a ogni $x \in X$ un sottoinsieme $f(x)$ di X. Questo sottoinsieme $f(x)$ potrebbe contenere x o meno. Consideriamo allora il sottoinsieme

$$A = \{x \in X \mid x \notin f(x)\}$$

costituito da tutti gli elementi di X che non appartengono al sottoinsieme $f(x)$ che viene loro associato mediante la f.

Dal momento che A è un sottoinsieme di X, ovvero appartiene a $P(X)$, e la funzione $f\colon X \longrightarrow P(X)$ è biiettiva e in particolare suriettiva, deve esistere un elemento $a \in X$ che ha come immagine proprio A, ovvero $f(a) = A$.

Ma a questo punto otteniamo la contraddizione cercata chiedendoci se a appartiene a A. Infatti, se $a \in A$, allora essendo A l'insieme degli elementi x caratterizzati dalla proprietà $x \notin f(x)$, sarebbe $a \notin f(a)$, che essendo $f(a) = A$ significa $a \notin A$. Viceversa, se $a \notin A$ questo significa che $a \notin f(a)$ e quindi a gode della proprietà $x \notin f(x)$ che definisce gli elementi di A ovvero $a \in A$. Vediamo quindi che $a \in A$ se e solo se $a \notin A$.

Deduciamo quindi che l'ipotesi che $f\colon X \longrightarrow P(X)$ fosse biiettiva era assurda, ovvero non esiste nessuna funzione biiettiva da X a $P(X)$. □

2.6 Composizione di funzioni e funzioni invertibili

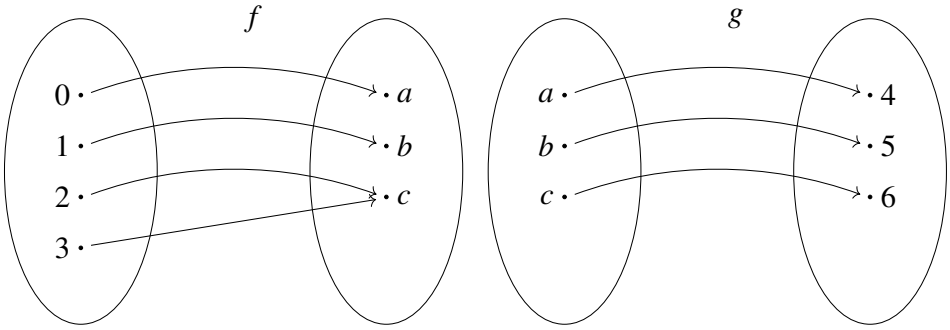
Cosí come negli insiemi numerici esistono operazioni (somma e prodotto) che dati due numeri ce ne danno un terzo, e nell'algebra di Boole esistono operazioni (le congiunzioni "e" e "o") che date due proposizioni ne danno una terza, anche per le funzioni, sotto ipotesi opportune, esiste un'operazione che produce una funzione a partire da due funzioni date: tale operazione si chiama **composizione**.

Più precisamente, supponiamo di avere due funzioni $f\colon X \longrightarrow Y$ e $g\colon Y \longrightarrow Z$, tali che *il codominio della prima coincida con il dominio della seconda*. Allora, per ogni elemento $x \in X$, possiamo applicare prima f ottenendo $f(x) \in Y$, e poi dal momento che Y è anche il dominio della g possiamo applicare la g a $f(x)$, ottenendo $g(f(x))$. In questo modo otteniamo una nuova funzione che associa a ogni elemento di X un elemento di Z:

$$\begin{aligned} g \circ f\colon X &\longrightarrow Y \\ x &\mapsto g(f(x)) \end{aligned}$$

A causa dell'ordine in cui appaiono le funzioni in $g(f(x))$, questa nuova funzione si denota con $g \circ f$ (che si legge "f composto g" a causa dell'ordine in cui vengono applicate).

Esempio 2.36. Consideriamo le funzioni $f\colon \{0,1,2,3\} \longrightarrow \{a,b,c\}$ e $g\colon \{a,b,c\} \longrightarrow \{4,5,6\}$ rappresentate nel seguente disegno



Essendo il codominio di f uguale al dominio di g, si può costruire la composizione $g \circ f$ applicando prima f e poi g ad ogni elemento di $\{0,1,2,3\}$. Otteniamo quindi $g \circ f\colon \{0,1,2,3\} \longrightarrow \{a,b,c\}$,

$$\begin{aligned} (g \circ f)(0) &= g(f(0)) = g(a) = 4 \\ (g \circ f)(1) &= g(f(1)) = g(b) = 5 \end{aligned}$$

$$(g \circ f)(2) = g(f(2)) = g(c) = 6$$
$$(g \circ f)(3) = g(f(3)) = g(c) = 6$$

Si noti che, in questo esempio, la composizione $f \circ g$ non è definita in quanto il codominio di g non coincide[8] con il dominio di f. Infatti non ha senso provare a calcolare, per esempio, $f(g(a)) = f(4)$.
In alcuni casi hanno senso invece tutte e due le composizioni, ad esempio se $f\colon X \longrightarrow Y$ e $g\colon Y \longrightarrow X$ allora hanno senso sia $g \circ f\colon X \longrightarrow X$ che $f \circ g\colon Y \longrightarrow Y$. Chiaramente, in generale, si tratta di due funzioni diverse, ovvero si avrà $g \circ f \neq f \circ g$. Per cui possiamo dire che la composizione di funzioni non gode sicuramente della proprietà commutativa.
Questo è vero anche nel caso specifico in cui componiamo due funzioni $f\colon X \longrightarrow X$ e $g\colon X \longrightarrow X$, per le quali si possono considerare sia $g \circ f$ che $f \circ g$ ed entrambe sono funzioni $X \longrightarrow X$.

Esempio 2.37. Si considerino le due funzioni

$$f\colon \mathbb{R} \longrightarrow \mathbb{R}, \quad f(x) = x^2$$
$$g\colon \mathbb{R} \longrightarrow \mathbb{R}, \quad g(x) = x + 1$$

Si ha allora

$$(g \circ f)(x) = g(f(x)) = g(x^2) = x^2 + 1$$

mentre

$$(f \circ g)(x) = f(g(x)) = f(x + 1) = (x + 1)^2 = x^2 + 2x + 1.$$

La composizione di funzioni gode invece sempre della proprietà associativa. Più precisamente, date tre funzioni $f\colon X \longrightarrow Y$, $g\colon Y \longrightarrow Z$, $h\colon Z \longrightarrow W$, si ha

$$(h \circ g) \circ f = h \circ (g \circ f)\,. \tag{2.11}$$

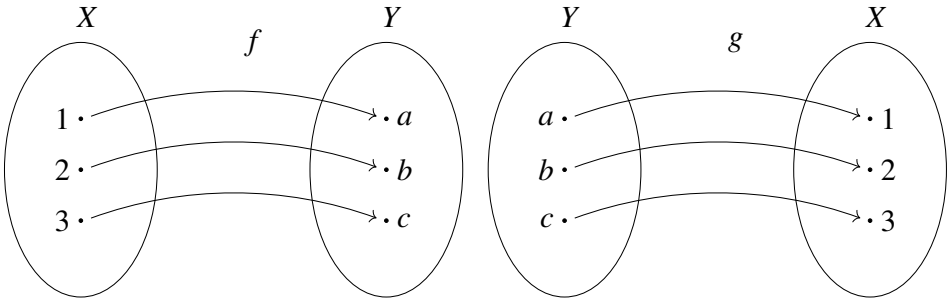
Infatti, si verifica subito che entrambi i membri di questa uguaglianza sono funzioni $X \longrightarrow W$; inoltre, per ogni $x \in X$, per definizione di composizione si ha

$$[(h \circ g) \circ f](x) = (h \circ g)(f(x)) = h(g(f(x))) \text{ e}$$
$$[h \circ (g \circ f)](x) = h((g \circ f)(x)) = h(g(f(x)))\,.$$

[8] In realtà per poter comporre due funzioni $f\colon X \longrightarrow Y$ e $g\colon Y' \longrightarrow Z$ basterebbe anche solo che il codominio Y della prima fosse *contenuto* nel dominio Y' della seconda. Infatti, se $Y \subseteq Y'$, allora ogni elemento di Y è anche un elemento di Y', quindi $f(x) \in Y$ sarebbe anche un elemento di Y' e potremmo applicargli la g ottenendo $g(f(x))$. Tuttavia, per semplicità supporremmo sempre che il codominio della prima funzione coincida col dominio della seconda.

Grazie a questa proprietà, quando dobbiamo comporre tre (o più) funzioni possiamo scrivere semplicemente $h \circ g \circ f$, omettendo le parentesi senza che l'espressione risulti ambigua.

Consideriamo ora $X = \{1,2,3\}$ e $Y = \{a,b,c\}$ e le funzioni $f\colon X \longrightarrow Y$ e $g\colon Y \longrightarrow X$ rappresentate nel disegno seguente.



Come si vede subito, si ha

$$\begin{aligned}(g \circ f)(1) &= g(f(1)) = g(a) = 1,\\ (g \circ f)(2) &= g(f(2)) = g(b) = 2,\\ (g \circ f)(3) &= g(f(3)) = g(c) = 3,\end{aligned}$$

ovvero $g \circ f\colon \{1,2,3\} \longrightarrow \{1,2,3\}$ è la funzione che manda ogni elemento dell'insieme $X = \{1,2,3\}$ in se stesso. Tale particolare funzione si chiama **funzione identica** di X e si denota con id_X. Possiamo quindi scrivere

$$g \circ f = id_X\,. \tag{2.12}$$

Analogamente, si ha

$$\begin{aligned}(f \circ g)(a) &= f(g(a)) = f(1) = a,\\ (f \circ g)(b) &= f(g(b)) = f(2) = b,\\ (f \circ g)(c) &= f(g(c)) = f(3) = c,\end{aligned}$$

ovvero $f \circ g\colon \{a,b,c\} \longrightarrow \{a,b,c\}$ è la funzione che manda ogni elemento dell'insieme $Y = \{a,b,c\}$ in se stesso, cioè

$$f \circ g = id_Y\,. \tag{2.13}$$

Le uguaglianze (2.12) e (2.13) sono analoghe alle uguaglianze

$$a \cdot b = 1 \text{ e } b \cdot a = 1$$

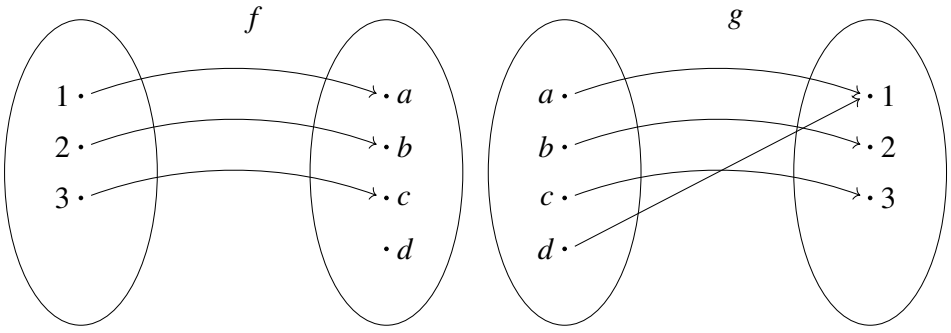
che sussistono quando a e b sono due numeri, uno l'inverso dell'altro (ad esempio, $a = 2$ e $b = \frac{1}{2}$). In effetti, per convincersi ulteriormente di quanto questa analogia sia appropriata, osserviamo che, così come il numero 1 ha la proprietà di fungere da cosiddetto "elemento neutro per la moltiplicazione" (cioè soddisfa $x \cdot 1 = 1 \cdot x = x$ per ogni numero reale x), anche le funzioni identiche fungono da elementi neutri per la composizione. Più precisamente, se abbiamo una funzione $f\colon X \longrightarrow Y$, allora $f \circ id_X = f$ e $id_Y \circ f = f$ in quanto, dal momento che la funzione identica manda ogni elemento in se stesso, si ha

$$(f \circ id_X)(x) = f(id_X(x)) = f(x),$$
$$(id_Y \circ f)(x) = id_Y(f(x)) = f(x)\,.$$

Questa analogia con quello che accade per l'inverso di un numero ci suggerisce la seguente, importante

Definizione 2.38. Una funzione $f\colon X \longrightarrow Y$ si dice **invertibile** se esiste una funzione $g\colon Y \longrightarrow X$ per cui $g \circ f = id_X$ e $f \circ g = id_Y$. In tal caso la funzione g si dice **inversa** di f e la si denota con f^{-1}.

Mentre nell'insieme dei numeri reali tutti i numeri sono invertibili tranne lo zero, per le funzioni e la composizione la situazione è un po' più complessa. Infatti esistono funzioni f per cui vale la $g \circ f = id_X$ ma $f \circ g \neq id_Y$, e in tal caso si dice che g è un'**inversa sinistra** di f. Al contempo esistono funzioni f per cui viceversa si ha $f \circ g = id_Y$ ma $g \circ f \neq id_X$, e in tal caso si dice che g è un'**inversa destra** di f.
Per un esempio di funzione del primo tipo, si prendano $f\colon \{1,2,3\} \longrightarrow \{a,b,c,d\}$ e $g\colon \{a,b,c,d\} \longrightarrow \{1,2,3\}$ definite come nel seguente disegno.

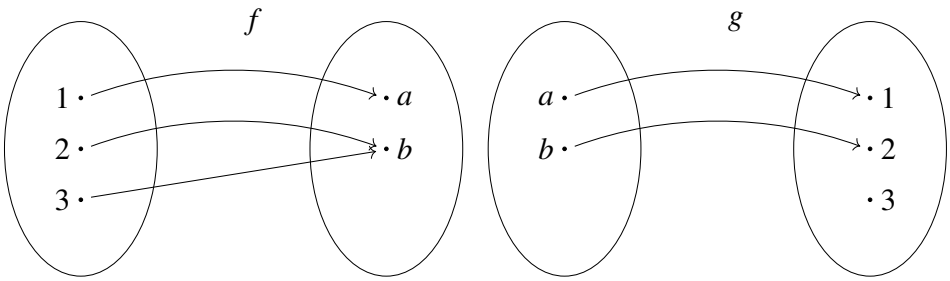


Si vede subito che $g \circ f$ è la funzione identica di $\{1,2,3\}$, ma $f \circ g$ non è la funzione identica di $\{a,b,c,d\}$. Infatti, pur essendo $f(g(a)) = f(1) = a$, $f(g(b)) = f(2) = b$ e $f(g(c)) = f(3) = c$, si ha $f(g(d)) = f(1) = a$, ovvero $f \circ g$ non manda d in se stesso.
Si osservi che non c'è alcun modo di modificare g in modo che sia anche un'inversa destra per f. Qualunque valore assegniamo a d non sarà mai $f(g(d)) = d$, perchè

$g(d)$ dovrebbe essere un elemento mandato da f in d, ma non esiste nessun elemento di $\{1,2,3\}$ che viene mandato da f in d.
In altre parole, il motivo per cui non esiste un'inversa destra di f è dovuto alla non suriettività di f. Questo esempio illustra la validità della seguente proposizione, di cui omettiamo la dimostrazione rigorosa.

Proposizione 2.39. *Una funzione $f\colon X \longrightarrow Y$ ammette un'inversa destra se e solo se è suriettiva*

Analogamente, per un esempio di funzione che ammette un'inversa destra ma non sinistra[9], si prendano $f\colon \{1,2,3\} \longrightarrow \{a,b\}$ e $g\colon \{a,b\} \longrightarrow \{1,2,3\}$ definite dal seguente disegno.



Si nota immediatamente che $f \circ g$ è la funzione identica di $\{a,b\}$, ma $g \circ f$ non è la funzione identica di $\{1,2,3\}$. Di fatto, pur essendo $g(f(1)) = g(a) = 1$ e $g(f(2)) = g(b) = 2$, si ha $g(f(3)) = g(b) = 2$, ovvero $g \circ f$ non manda 3 in se stesso.
Si osservi che anche qui non c'è alcun modo di modificare g in modo che sia anche un'inversa sinistra per f. Se avessimo posto $g(b) = 3$ avremmo sì ottenuto $g(f(3)) = g(b) = 3$ ma stavolta sarebbe stato $g(f(2)) = g(b) = 3$, ovvero $g \circ f$ non avrebbe mandato 2 in se stesso. Come si vede, il problema stavolta è che f manda entrambi gli elementi 2 e 3 in b, quindi necessariamente $g(f(2))$ e $g(f(3))$ coincideranno e non potranno mai essere il primo 2 e il secondo 3.
In altre parole, il motivo per cui non esiste un'inversa sinistra di f è dovuto alla non iniettività di f. Questo esempio illustra la validità della seguente proposizione, di cui omettiamo una dimostrazione rigorosa.

Proposizione 2.40. *Una funzione $f\colon X \longrightarrow Y$ ammette un'inversa sinistra se e solo se è iniettiva*

Poiché, per la Definizione 2.38, dire che f ha un'inversa g significa dire che g è sia inversa sinistra che inversa destra, possiamo combinare la Proposizione 2.39 e la Proposizione 2.40 ottenendo la

[9] In realtà basterebbe l'esempio già visto, in cui è g ad avere f come inversa destra (ma non sinistra) ma per una maggiore chiarezza facciamo un ulteriore esempio.

Proposizione 2.41. *Una funzione $f\colon X \longrightarrow Y$ ammette un'inversa se e solo se è sia iniettiva che suriettiva (ovvero biiettiva).*

2.7 Numeri naturali

In questa sezione useremo la nozione di funzione biiettiva e le sue proprietà per definire rigorosamente i concetti elementari di "numero" e "contare".

Definizione 2.42. Un insieme X si dice **equipotente** a un insieme Y se esiste una funzione biiettiva $f\colon X \longrightarrow Y$.

La Definizione 2.42 definisce una relazione tra insiemi, che denotiamo in questo capitolo con il simbolo $\sim$.

Proposizione 2.43. *La relazione di equipotenza definita nella Definizione 2.42 è una relazione di equivalenza.*

Prima di dimostrare la proposizione enunciamo un lemma indipendentemente per la sua importanza e per referenza futura.

Lemma 2.44. *Siano $f\colon X \longrightarrow Y$ e $g\colon Y \longrightarrow Z$ due funzioni. Allora*

(1) se f e g sono iniettive, anche $g \circ f$ è iniettiva

(2) se f e g sono suriettive, anche $g \circ f$ è suriettiva

In particolare, se f e g sono biiettive, anche $g \circ f$ è biiettiva.

Dimostrazione. Per dimostrare che $g \circ f\colon X \longrightarrow Z$ è iniettiva, dobbiamo mostrare che per ogni $x_1, x_2 \in X$, se $(g \circ f)(x_1) = (g \circ f)(x_2)$ allora $x_1 = x_2$. Ma, per definizione di composizione, $(g \circ f)(x_1) = (g \circ f)(x_2)$ significa $g(f(x_1)) = g(f(x_2))$. Da questa uguaglianza, siccome g è iniettiva segue che $f(x_1) = f(x_2)$, e siccome f è iniettiva a sua volta da quest'ultima segue $x_1 = x_2$, come volevamo.
Per dimostrare che $g \circ f\colon X \longrightarrow Z$ è suriettiva dobbiamo dimostrare che per ogni $z \in Z$ esiste un $x \in X$ tale che $(g \circ f)(x) = z$. Ora, dal momento che $g\colon Y \longrightarrow Z$ è suriettiva, per ogni $z \in Z$ esiste un $y \in Y$ tale che $z = g(y)$. A sua volta, dal momento che y appartiene al codominio di $f\colon X \longrightarrow Y$ che è anch'essa per ipotesi suriettiva, esiste un $x \in X$ tale che $y = f(x)$. Sostituendo allora $y = f(x)$ in $z = g(y)$ si ottiene $z = g(f(x))$, ovvero per definizione di composizione $z = (g \circ f)(x)$, come volevamo.
L'ultima parte del lemma è un corollario immediato delle prime due. □

Siamo ora pronti per la

Dimostrazione della Proposizione 2.43. Per dimostrare che $\sim$ è una relazione di equivalenza dobbiamo mostrare che è riflessiva, simmetrica e transitiva.

i) **Riflessività.** Dato un qualunque insieme X, la funzione identica $id_X : X \longrightarrow X$ che manda ogni elemento di X in se stesso è una funzione biiettiva (è chiaramente iniettiva in quanto due elementi x, x' diversi hanno come immagini x, x' stessi, quindi immagini diverse; è chiaramente suriettiva in quanto ogni elemento x del codominio è immagine di se stesso nel dominio).

ii) **Simmetria.** Supponiamo che sia $X \sim Y$, ossia che esista una funzione biiettiva $f\colon X \longrightarrow Y$. Per mostrare che $Y \sim X$ (e quindi la simmetria) ci basta trovare una funzione biiettiva da Y a X. A questo scopo, basta ricordare che le funzioni biiettive sono caratterizzate dal fatto di essere invertibili e prendere la funzione $f^{-1} : Y \longrightarrow X$ inversa di f. Essendo f^{-1} l'inversa di f, si ha $f \circ f^{-1} = id_Y$ e $f^{-1} \circ f = id_X$. Ma queste due uguaglianze ci dicono che f è a sua volta l'inversa (sia destra che sinistra) di f^{-1}. Quindi f^{-1} è invertibile e, per quanto appena ricordato, biiettiva. Abbiamo dunque trovato una funzione biiettiva da Y a X. Questo implica che $Y \sim X$, e possiamo concludere che la relazione è simmetrica.

iii) **Transitività.** Supponiamo che sia $X \sim Y$ e $Y \sim Z$, ovvero che esista una funzione $f\colon X \longrightarrow Y$ biiettiva da X a Y e una funzione $g\colon Y \longrightarrow Z$ biiettiva da Y a Z. Se riusciamo a mostrare che la composizione $g \circ f\colon X \longrightarrow Z$ è anch'essa biiettiva avremo dimostrato che esiste una funzione biiettiva da X a Z e quindi che $X \sim Z$, da cui la transitività. A questo scopo, dimostriamo separatamente che $g \circ f\colon X \longrightarrow Z$ è iniettiva e che è suriettiva.

   Per quanto riguarda l'iniettività, siano $x, x' \in X$, con $x \neq x'$. Vogliamo dimostrare che $(g \circ f)(x) \neq (g \circ f)(x')$. Iniziamo con l'osservare che, dal momento che f è biiettiva e quindi in particolare iniettiva, abbiamo $f(x) \neq f(x')$. Essendo $f(x)$ e $f(x')$ elementi diversi di Y, che è il codominio di f ma anche il dominio di $g\colon Y \longrightarrow Z$, possiamo applicare loro la g e, essendo anche quest'ultima biiettiva e in particolare iniettiva, si avrà $g(f(x)) \neq g(f(x'))$. Per definizione di composta questo significa proprio $(g \circ f)(x) \neq (g \circ f)(x')$, che è quello che volevamo.

   Per la suriettività di $g \circ f\colon X \longrightarrow Z$, dobbiamo dimostrare che per ogni $z \in Z$ esiste un $x \in X$ tale che $(g \circ f)(x) = z$. Iniziamo con l'osservare che, dal momento che $g\colon Y \longrightarrow Z$ è biiettiva e in particolare suriettiva, esiste (almeno) un $y \in Y$ tale che $g(y) = z$. Ma dal momento che $f\colon X \longrightarrow Y$ è biiettiva e in particolare suriettiva, a sua volta esiste un $x \in X$ tale che $f(x) = y$. Ma allora possiamo riscrivere la $g(y) = z$ come $g(f(x)) = z$. Per definizione di composizione questo significa proprio $(g \circ f)(x) = z$, che è quello che volevamo dimostrare.

□

Essendo l'equipotenza una relazione di equivalenza, possiamo considerare delle sue classi di equivalenza. Ebbene, i numeri naturali, che ci accingiamo a definire in modo

rigoroso, sono esattamente il modo in cui noi identifichiamo le classi di equivalenza degli insiemi finiti. Infatti, se possiamo dire per esempio che l'insieme $X = \{a, b, c, d\}$ ha 4 elementi, è perché abbiamo *contato* i suoi elementi. Cioè li abbiamo messi in corrispondenza con gli elementi dell'insieme $\{1, 2, 3, 4\}$ mediante una funzione biettiva (ad esempio, a 1 associamo a, a 2 associamo b, a 3 associamo c, a 4 associamo d). Chiaramente avremmo contato male se avessimo contato a due volte, associando sia a 1 che a 2 l'elemento a, e anche se dimenticassimo di contare qualche elemento di X, ad esempio d. Il primo errore equivarrebbe alla costruzione di una corrispondenza $\{1, 2, 3, 4\} \longrightarrow \{a, b, c, d\}$ non iniettiva (in cui gli elementi diversi 1 e 2 del dominio hanno la stessa immagine, a, nel codominio). Il secondo errore significherebbe aver costruito una corrispondenza non suriettiva (in cui l'elemento d, non essendo stato contato, non è immagine di nessuno degli elementi $1, 2, 3, 4$ del dominio).
In generale, *contare* gli elementi di un insieme X e concludere che questo ha n elementi significa costruire una funzione biiettiva tra X e l'insieme dei numeri naturali che vanno da 1 a n.
Per rendere rigoroso cosa significa "l'insieme dei numeri naturali da 1 a n" dobbiamo definire i numeri naturali e chiarire da dove nasce la successione naturale in cui li pensiamo quando contiamo. La definizione che daremo ora è assiomatica, ovvero definiremo l'insieme $\mathbb{N}$ dei numeri naturali mediante assiomi che, pur non dicendoci cosa siano i numeri naturali, ci dicono però le relazioni che intercorrono tra di essi e li definiscono così implicitamente.[10]
Enunciamo allora gli assiomi che implicitamente definiscono i numeri naturali, detti **assiomi di Peano** (dal nome del matematico italiano Giuseppe Peano, al quale sono dovuti).

(1) Lo zero 0 è un numero naturale.

    Questo assioma afferma che l'insieme dei numeri naturali è un insieme non vuoto, nel quale c'è almeno un elemento che chiamiamo "zero" (i successivi assiomi chiariranno la particolarità di questo elemento rispetto agli altri). Si noti che se ci fermassimo a questo assioma, l'insieme dei numeri naturali potrebbe anche contenere solo 0, ovvero essere $\mathbb{N} = \{0\}$.

(2) Per ogni numero naturale n, ne esiste un altro n' detto il suo *successore*.

    Questo assioma introduce in pratica una funzione $s : \mathbb{N} \longrightarrow \mathbb{N}$ dall'insieme $\mathbb{N}$ dei numeri naturali in se stesso, detta appunto funzione successore.

(3) Due numeri naturali diversi hanno successori diversi.

[10] Esattamente come nella costruzione mediante assiomi della geometria euclidea non diciamo cosa siano punto e retta ma specifichiamo mediante gli assiomi le relazioni che li legano - ad esempio "dati due punti distinti esiste ed è unica la retta che li contiene".

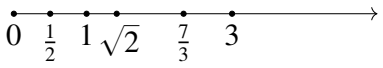
Questo assioma ci sta dicendo che la funzione successore è iniettiva.

(4) Lo zero 0 non è successore di nessun numero.

Prima di enunciare il quinto e ultimo assioma, fermiamoci a riflettere sulle prime conseguenze degli assiomi (1), (2), (3), (4). L'assioma (4) ci dice che il successore di 0 non può essere 0 stesso.[11] Chiamando "uno" il successore di zero e denotandolo 1, vediamo che il successore di 1 (che esiste per il secondo assioma) non può essere né zero (per il quarto assioma) nè nuovamente 1, altrimenti contrariamente a quello che afferma il terzo assioma la funzione successore non sarebbe iniettiva. Quindi abbiamo che il successore di 1 è un nuovo numero naturale, diverso da 0 e 1, che denotiamo con 2, e così via.

Questi primi quattro assiomi impediscono quindi ad esempio che i numeri naturali siano ad esempio l'insieme finito $\{0,1,2,3\}$ con successore di 0 dato da 1, successore di 1 dato da 2, successore di 2 dato da 3, e successore di 3 dato da 0 (o da un altro degli elementi $1,2,3$). Ovvero non è possibile che la funzione successore "permuti ciclicamente" gli elementi. Gli assiomi (1), (2), (3), (4) garantiscono quindi che l'insieme dei naturali sia un insieme infinito.[12]

Siamo ora pronti a enunciare il quinto e ultimo assioma di Peano. Allo scopo di comprendere meglio l'importanza di tale assioma, osserviamo che i primi 4 assiomi risultano soddisfatti se consideriamo ad esempio l'insieme $\mathbb{R}$ di tutti i numeri reali (cioè tutti i decimali, limitati o illimitati, periodici e non, con segno + o -) maggiori o uguali di zero, rappresentati lungo una semiretta come nel disegno.



Come funzione successore consideriamo la funzione che manda un qualsiasi numero x nel numero $x+1$. Questo ci fa capire che i primi quattro assiomi non sono sufficienti a definire in modo univoco i numeri naturali così come siamo abituati a pensarli. Come questo esempio illustra bene, i primi quattro assiomi non impediscono che oltre allo zero e ai numeri che ottengo applicando ripetutamente la funzione successore ci siano altri elementi. Per evitare questo e fare di $\mathbb{N}$ *il più piccolo insieme che soddisfi primi quattro assiomi*, serve il seguente, quinto e ultimo assioma.

[11] Si noti che questo non è impedito dagli altri assiomi: l'insieme $\{0\}$ costituito da un solo elemento denotato 0, con la funzione successore che associa a 0 se stesso, verifica i primi tre assiomi . Per l'iniettività, si osservi che l'implicazione che definisce una funzione iniettiva, cioè $x \neq x' \Rightarrow f(x) \neq f(x')$, è vera in quanto, non essendoci due elementi diversi, l'antecedente dell'implicazione è sempre falso, il che come sappiamo rende l'implicazione automaticamente vera.

[12] Daremo più avanti la definizione rigorosa di insieme infinito, e usiamo per il momento questa parola con il suo significato intuitivo nel linguaggio comune.

(5) Se un sottoinsieme U dell'insieme dei numeri naturali contiene 0, e per ogni numero n che sta in U appartiene a U anche il successore n' di n, allora U coincide con $\mathbb{N}$.

Questo assioma garantisce che si ottengano tutti i numeri naturali partendo dallo zero e applicando ripetutamente la funzione successore, e che non ci siano altri elementi in $\mathbb{N}$ oltre a quelli che si possono ottenere in questo modo.

Come abbiamo accennato sopra, gli assiomi di Peano caratterizzano completamente i numeri naturali così come li intendiamo comunemente, e a partire da essi è possibile costruire tutta l'aritmetica usuale. Ad esempio è possibile definire le operazioni di somma e prodotto tra naturali e verificarne le varie proprietà. In base a tali definizioni, che non diamo, il successore di un numero naturale n risulta $n + 1$, ed è così che lo denoteremo da ora in poi.
Tornando al punto da cui siamo partiti, gli assiomi chiariscono in che modo i naturali sono i numeri che "ci servono per contare". Quando stiamo contando gli elementi di un insieme, partiamo da uno e applichiamo ripetutamente la funzione successore, fino a che non avremo esaurito gli elementi dell'insieme che stiamo contando. Il numero a cui arriviamo ci dice "quanti elementi ha l'insieme dato" ovvero identifica la classe di equivalenza della relazione di equipotenza. Quando diciamo che due insiemi X e Y hanno entrambi 4 elementi stiamo dicendo che esiste una funzione biiettiva dall'insieme $\{1, 2, 3, 4\}$ in X e una funzione biiettiva dall'insieme $\{1, 2, 3, 4\}$ in Y, e quindi che, dato che la relazione di equipotenza è una relazione di equivalenza, esiste una funzione biiettiva da X a Y. Possiamo dire che X e Y stanno nella stessa classe di equivalenza, determinata dal numero naturale 4, e che "hanno lo stesso numero di elementi". Si noti che nel contare partiamo da 1: il numero zero è usato per denotare la classe di equivalenza dell'insieme vuoto rispetto alla relazione di equipotenza.

2.8 Applicazioni del quinto assioma (il principio d'induzione)

Il quinto assioma di Peano (detto anche *principio d'induzione*) è di fondamentale importanza quando si vuole dimostrare che una certa proprietà, un'uguaglianza o una formula vale per ogni numero naturale. Infatti, basterà dimostrare che tale proprietà/formula/uguaglianza etc. vale per 0 (questa prima parte si chiama spesso *passo base*), e poi che se essa vale per un numero naturale m allora vale anche per il suo successore $m + 1$ (questa seconda parte è spesso chiamata *passo induttivo*). In base al quinto assioma, concluderemo allora che l'insieme U dei naturali per i quali vale la nostra formula coincide con l'insieme di tutti i numeri naturali.

Esempi 2.45. Vediamo subito alcuni semplici ma significativi esempi.

(1) Supponiamo di voler dimostrare applicando il principio di induzione che la disuguaglianza

$$2^n \geqslant n+1 \tag{2.14}$$

vale per ogni numero naturale n.

A questo scopo, osserviamo innanzitutto che la disuguaglianza vale per $n = 0$, in quanto per tale valore primo e secondo membro sono uguali ($2^0 = 1$ e $0+1 = 1$). Il cosiddetto passo base è concluso.

Per applicare il cosiddetto passo induttivo, supponiamo ora che la disuguaglianza si vera per un certo numero naturale arbitrario m (questa ipotesi è detta *ipotesi induttiva*, cioè che $2^m \geqslant m+1$. Mostriamo che, *sotto tale ipotesi*, essa vale anche per il suo successore $m+1$, ovvero che $2^{m+1} \geqslant (m+1)+1$. Infatti, $2^{m+1} = 2 \cdot 2^m$: ma essendo per ipotesi induttiva $2^m \geqslant m+1$ sarà $2 \cdot 2^m \geqslant 2 \cdot (m+1)$ (una disuguaglianza rimane valida se moltiplichiamo primo e secondo membro per uno stesso numero positivo). Quindi

$$2^{m+1} \geqslant 2(m+1) = 2m+2$$

e per concludere la disuguaglianza che serva a noi, ovvero che $2^{m+1} \geqslant (m+1)+1$, ci basterà allora mostrare che $2m+2 \geqslant (m+1)+1$. Ma quest'ultima disuguaglianza significa semplicemente $2m+2 \geqslant m+2$, che semplificando il 2 da entrambi i membri e portando a primo membro l'm che compare a secondo ci dà $m \geqslant 0$, cioè essa è vera qualunque sia m.

Riassumendo, abbiamo dimostrato che la (2.14) è vera per $n = 0$ e che se essa è vera per m allora è vera per $m+1$. In base al principio d'induzione essa è vera per qualunque numero naturale.

(2) La somma dei numeri naturali da 0 a n vale $\frac{n(n+1)}{2}$, ovvero, usando il simbolo di sommatoria,

$$\sum_{k=0}^{n} k = \frac{n(n+1)}{2} \tag{2.15}$$

Per dimostrare per induzione che questa formula vale per ogni numero naturale $n \in \mathbb{N}$, verifichiamo prima che essa vale per $n = 0$: infatti, in tal caso la sommatoria contiene solo 0, mentre dall'altra si ha $\frac{0(0+1)}{2} = 0$, quindi l'uguaglianza è verificata.

Ora, supponiamo di sapere che la formula (2.15) valga per n, e dimostriamo che essa vale per $n+1$, ovvero dimostriamo che

$$\sum_{k=0}^{n+1} k = \frac{(n+1)[(n+1)+1]}{2} = \frac{(n+1)(n+2)}{2}\,. \tag{2.16}$$

Si ha chiaramente

$$\sum_{k=0}^{n+1} k = \sum_{k=0}^{n} k + (n+1)\,. \tag{2.17}$$

Ma poiché stiamo supponendo che la formula (2.15) valga per n, possiamo sostituire $\sum_{k=0}^{n} k$ con $\frac{n(n+1)}{2}$, e quindi la (2.17) si scrive

$$\sum_{k=0}^{n+1} k = \frac{n(n+1)}{2} + (n+1)\,.$$

Ma, svolgendo i conti, il secondo membro è uguale a

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

e quindi abbiamo dimostrato la (2.16), come volevamo.

Osservazione 2.46. Il metodo si applica anche quando si vuole dimostrare che una certa proprietà P vale non per ogni naturale ma a partire da un $n_0 \in \mathbb{N}$ dato. In tal caso, bisogna dimostrare che

(1) la proprietà P vale per n_0

(2) ogniqualvolta la P vale per un certo numero n, allora vale per il suo successivo $n+1$ (scriveremo $P(n) \Rightarrow P(n+1)$)

Esempio 2.47. Per ogni numero naturale $n \geqslant 1$, si ha

$$2^{n-1} \leqslant n! \tag{2.18}$$

Dimostriamo anche questa formula per induzione, nella versione vista nell'Osservazione 2.46. Poichè vogliamo mostrare che essa vale per ogni $n \geqslant 1$, iniziamo con il dimostrare che essa vale per $n = 1$. Sostituendo $n = 1$ in (2.18) si trova da un lato $2^{1-1} = 2^0 = 1$ e dall'altro $1! = 1$, ovvero $1 \leqslant 1$, che è vera.
Ora, come prevede il metodo di induzione, supponiamo che la (2.18) sia valida per un certo n e dimostriamo che essa vale per $n+1$, ovvero che

$$2^{(n+1)-1} \leqslant (n+1)!\,. \tag{2.19}$$

Si ha chiaramente $2^{(n+1)-1} = 2^n = 2 \cdot 2^{n-1}$, e allora poiché stiamo supponendo che valga la $2^{n-1} \leqslant n!$ possiamo scrivere

$$2^{(n+1)-1} = 2 \cdot 2^{n-1} \leqslant 2 \cdot n! \leqslant (n+1) \cdot n! = (n+1)!$$

dove nell'ultima disuguaglianza abbiamo sfruttato il fatto che per ogni $n \geqslant 1$ si ha $2 \leqslant n+1$. La (2.19) è dimostrata.

Osservazione 2.48. Quando si dimostra per induzione che una certa proprietà P vale per ogni numero naturale maggiore o uguale di un numero di partenza n_0, dopo aver dimostrato che $P(n_0)$ è vera, bisogna stare attenti che l'implicazione $P(n) \Rightarrow P(n+1)$ sia valida effettivamente per ogni $n \geqslant n_0$.
Ad esempio, consideriamo la disuguaglianza

$$2^n \geqslant n^2 \tag{2.20}$$

e supponiamo di voler dimostrare che essa è vera per tutti i numeri naturali.
Chiaramente, tale disuguaglianza è vera per $n = 0$ in quanto si riduce a $2^0 = 1 \geqslant 0^2 = 0$.
Ora, come prevede l'induzione, supponiamo che la formula sia vera per n (ipotesi induttiva) e dimostriamo che essa vale per $n+1$, ovvero che

$$2^{(n+1)} \geqslant (n+1)^2 . \tag{2.21}$$

Usando l'ipotesi induttiva $2^n \geqslant n^2$, si ha

$$2^{(n+1)} = 2 \cdot 2^n \geqslant 2 \cdot n^2$$

e quindi per dimostrare la (2.21) basterebbe mostrare che $2n^2 \geqslant (n+1)^2$, ovvero $2n^2 \geqslant n^2 + 2n + 1$, che portando tutto a primo membro equivale alla disuguaglianza

$$n^2 - 2n - 1 \geqslant 0 . \tag{2.22}$$

Ora, ricordiamo che una generica disuguaglianza di secondo grado $ax^2 + bx + c \geqslant 0$ è verificata per $x \geqslant x_1$ e $x \leqslant x_2$ se a è positivo, e per $x_1 \leqslant x \leqslant x_2$ se a è negativo, dove x_1 e x_2 sono le soluzioni di $ax^2 + bx + c = 0$.
Nel nostro caso, usando la nota formula risolutiva $x_{1,2} = \frac{-b \pm \sqrt{b^2-4ac}}{2a}$, si vede che $n^2 - 2n - 1 = 0$ ha come soluzioni $n = 1 - \sqrt{2}$ e $n = 1 + \sqrt{2}$. Quindi la nostra disuguaglianza (2.22) è verificata per $n \geqslant 1 + \sqrt{2} \sim 2,4$ e $n \leqslant 1 - \sqrt{2} \sim -0,4$. Siccome stiamo lavorando nei naturali, il primo numero naturale per cui la disuguaglianza è verificata è 3. Ne consegue che l'implicazione $P(n) \Rightarrow P(n+1)$ che vogliamo dimostrare è vera solo per $n \geqslant 3$. In conclusione, pur essendo $P(0)$ vera non possiamo partire da 0 per innescare il meccanismo dell'induzione. Come minimo, dobbiamo partire da 3, tuttavia, $P(3)$ non è vera. Infatti, sostituendo $n = 3$ in (2.20) si ottiene $2^3 \geqslant 3^2$, cioè $8 \geqslant 9$.

Sostituendo $n = 4$ vediamo invece che $2^4 \geqslant 4^2$ (cioè $16 \geqslant 16$), ovvero $P(4)$ è vera. Quindi non solo possiamo usare 4 come punto di partenza dell'induzione, ma sapendo anche che per $n \geqslant 4$ la $P(n) \Rightarrow P(n+1)$ è vera, la formula risulta dimostrata per induzione per $n \geqslant 4$.

Il principio d'induzione viene a volta enunciato e usato anche nella seguente forma, detta **principio d'induzione forte**.

Se una proprietà

(i) *vale per* $n = 0$

(ii) *vale per* $n+1$ *sotto l'ipotesi che valga per ogni numero naturale* k *compreso tra* 0 *e* n *(cioè* $0 \leqslant k \leqslant n$*)*

allora essa vale per tutti i numeri naturali.

In questa forma, il passo base (i) è identico al passo base del principio d'induzione già enunciato sopra mentre il passo induttivo si differenzia invece dal passo induttivo descritto prima. Infatti non assumiamo più che la proprietà valga solo per n per concludere poi che vale per $n+1$ (cioè che vale per un numero ogniqualvolta essa vale per quello immediatamente precedente), ma assumiamo che valga per tutti i numeri tra 0 e n (cioè che vale per un numero ogniqualvolta essa vale per *tutti* quelli precedenti).
Benchè in questa forma l'ipotesi induttiva sembri più restrittiva (poiché stiamo chiedendo qualcosa in più), si può dimostrare che queste due formulazioni diverse sono in realtà del tutto equivalenti. In altre parole, se sostituissimo il principio d'induzione forte al posto del quinto assioma di Peano così come l'abbiamo enunciato sopra si otterrebbe sempre lo stesso insieme $\mathbb{N}$ dei numeri naturali.
Tuttavia a volte l'uso del principio d'induzione nella forma forte si rivela più efficace, come vediamo nel seguente

Esempio 2.49. Un numero naturale[13] $n > 1$ si dice *primo* se gli unici modi di decomporlo come prodotto di due numeri naturali sono $n = n \cdot 1$ e $n = 1 \cdot n$. Ad esempio, 5 è primo perché non ci sono altre decomposizioni a parte $5 = 5 \cdot 1$ e $5 = 1 \cdot 5$, mentre 6 non è primo perché oltre a $6 = 6 \cdot 1$ e $6 = 1 \cdot 6$ abbiamo ad esempio anche $6 = 2 \cdot 3$. Dimostriamo ora, usando il principio d'induzione forte, che ogni numero naturale $n \geqslant 2$ si può decomporre come prodotto (con uno, due o più fattori) di primi. Per $n = 2$ questo è banalmente vero in quanto essendo esso stesso primo la decomposizione in fattori primi è data da 2 stesso (decomposizione con un solo fattore). Ora, supponiamo che questo valga per ogni numero naturale $2 \leqslant k \leqslant n$, cioè che ogni numero naturale compreso tra 2 e n si possa esprimere come prodotto di primi (ipotesi induttiva) e mostriamo che sotto tale ipotesi la stessa cosa vale per $n+1$, cioè che $n+1$ può essere scritto come

[13] In realtà questa definizione vale sull'insieme dei numeri interi, dei quali ci occuperemo più diffusamente nel prossimo capitolo.

prodotto di primi. Si hanno due possibilità: o $n+1$ è già primo, e in tal caso, esattamente come abbiamo fatto per 2, la decomposizione cercata è data da $n+1$ stesso (un solo fattore), oppure $n+1$ non è primo. In questo secondo esisterà una decomposizione $n+1 = a \cdot b$ di $n+1$ come prodotto in cui i due fattori non sono 1 e quindi sono necessariamente compresi tra 2 e n. In base all'ipotesi induttiva fatta, per cui ogni numero compreso tra 2 e n può essere decomposto in fattori primi, sia a che b possono essere decomposti in fattori primi, ovvero diciamo $a = p_1 p_2 \cdots p_m$ e $a = p'_1 p'_2 \cdots p'_l$. Ma allora $n+1 = ab = p_1 p_2 \cdots p_m p'_1 p'_2 \cdots p'_l$, e questa è proprio una decomposizione di $n+1$ in fattori primi, della quale dovevamo dimostrare l'esistenza. Questo completa la verifica delle due condizioni espresse dal principio d'induzione forte e quindi in base a tale principio mostra che tale affermazione vale per tutti i numeri naturali, come voluto.

Sottolineiamo come in questa dimostrazione il principio d'induzione nella forma classica sarebbe stato del tutto inutile. Infatti, non possiamo dedurre che $n+1$ ha una decomposizione in primi dal fatto che n ce l'ha. Ad esempio non possiamo ottenere una decomposizione in fattori primi di $n+1 = 60$ usando la decomposizione in fattori primi del suo predecessore $n = 59$ (le due composizioni in generale non hanno nulla a che vedere l'una con l'altra). Però, notando che $60 = 6 \cdot 10$ e conoscendo le decomposizioni in fattori primi di tutti i numeri che precedono 60, quindi in particolare anche di $6 = 2 \cdot 3$ e di $10 = 2 \cdot 5$, ottengo subito la decomposizione $60 = 2 \cdot 3 \cdot 2 \cdot 5$ di 60 in fattori primi.

Un'altra importante applicazione del principio d'induzione (sia nella forma classica che in quella forte) sono le cosiddette *definizioni ricorsive*. Tali definizioni consentono ad esempio di definire una successione di numeri $a_0, a_1, a_2, \ldots$ senza dare l'espressione esplicita del termine generico a_n.
Ad esempio, la successione $a_n = 2^n$ è tale che il primo termine $a_0 = 1$ e i successivi termini sono $a_1 = 2, a_2 = 4, a_3 = 8$ e così via. Notiamo che ogni termine viene ottenuto moltiplicando per 2 il precedente. Possiamo sfruttare questa osservazione per scrivere la successione come

$$\begin{cases} a_0 = 1 \\ a_{n+1} = 2a_n \,. \end{cases}$$

Il fatto che queste due uguaglianze siano sufficienti a ottenere tutti i termini a_n e definire così completamente la successione è garantito proprio dal principio d'induzione (in questo caso nella forma classica). Difatti, detto U l'insieme dei numeri naturali per cui la successione è definita, la $a_0 = 1$ ci dice che la successione è definita per $n = 0$, ovvero $0 \in U$, La seconda condizione $a_{n+1} = 2a_n$ implica che se per un certo numero naturale n il termine a_n è definito (cioè $n \in U$), allora lo è anche a_{n+1} (che si ottiene infatti semplicemente moltiplicando per 2 il termine a_n), ovvero $n+1 \in U$. Quindi sono verificate le due condizioni del principio d'induzione che garantiscono che $U = \mathbb{N}$, cioè la successione è definita per qualunque $n \in \mathbb{N}$.

Un altro esempio di successione definita in questo modo é il fattoriale $a_n = n!$, che può essere definito ricorsivamente da

$$\begin{cases} a_0 = 1 \\ a_{n+1} = (n+1)a_n \end{cases}$$

come è facile verificare.
Una definizione ricorsiva può sfruttare anche il principio d'induzione forte: questo succede quando nella definizione del termine a_{n+1} della successione non si utilizza solo il termine immediatamente precedente ma anche gli altri (o qualcuno degli altri). Un esempio significativo è dato dalla successione

$$\begin{cases} a_0 = 1 \\ a_1 = 1 \\ a_{n+1} = a_n + a_{n-1} \end{cases} \tag{2.23}$$

dove il termine a_{n+1} risulta definito da a_n e a_{n-1}. I primi termini della successione sono

$$a_0 = 1,\ a_1 = 1,\ a_2 = 2,\ a_3 = 3,\ a_4 = 5,\ a_5 = 8,\ a_6 = 13,\ \ldots$$

Si noti che è necessario dare il valore non solo di a_0 ma anche di a_1 in quanto la $a_{n+1} = a_n + a_{n-1}$, per $n = 1$, ci dà $a_1 = a_0 + a_{-1}$, e non essendo a_{-1} definito (-1 non è un numero naturale), non è possibile ottenere a_1 solo a partire da a_0. A partire da $n = 1$ invece gli indici a secondo membro della $a_{n+1} = a_n + a_{n-1}$ sono sempre numeri naturali, questo problema non si presenta. Quindi solo a partire da $n = 1$ la $a_{n+1} = a_n + a_{n-1}$ ci consente sempre di definire a_{n+1} una volta che sono definiti i precedenti, come richiede l'induzione forte.[14]
La successione (2.23) è un significativo esempio di come definire una successione per ricorsione possa essere molto più semplice che dare l'espressione esplicita del termine a_n. Inoltre la (2.23) fornisce un'illustrazione di come possa essere tutt'altro che immediato capire quale sia tale espressione. Più precisamente, si può vedere che il generico termine a_n definito dalla (2.23) è

$$a_n = \frac{1}{\sqrt{5}}\left(\frac{1+\sqrt{5}}{2}\right)^{n+1} - \frac{1}{\sqrt{5}}\left(\frac{1-\sqrt{5}}{2}\right)^{n+1}.$$

Osservazione 2.50. La successione (2.23) prende il nome di *successione di Fibonacci*, dal nome del matematico italiano del tredicesimo secolo che risolse tramite di essa

[14] In altre parole, il punto è che l'implicazione "se a_k è definito per ogni $0 \leqslant k \leqslant n$ allora a_{n+1} è definito" deve essere verificata per n generico, ma non lo è per $n = 0$. Di fatto per $n = 0$ essa diventa "se a_k è definito per ogni $0 \leqslant k \leqslant 0$ allora a_1 è definito", ovvero "se a_0 è definito allora a_1 è definito",. Questo non è vero proprio perché, come abbiamo osservato, dalla $a_{n+1} = a_n + a_{n-1}$ e conoscendo solo a_0 non riusciamo a ricavare a_1.

il seguente problema (proposto dall'imperatore Federico II). Supponiamo di partire al momento 0 da una coppia di conigli, che al mese 1 diventa fertile e al mese 2 genera una nuova coppia. Supponiamo inoltre che ogni nuova coppia si comporti nello stesso modo (cioè dal momento della nascita impiega un mese a diventare fertile e un altro mese per generare un'altra coppia). Quante coppie di conigli avremo all'n-esimo mese?
Il fatto che sia la successione (2.23) a dirci il numero di coppie di conigli all'n-esimo mese è il ragionamento seguente. Al mese 0 abbiamo la coppia di partenza ($a_0 = 1$). Al mese 1 questa coppia diventa fertile ma non ha ancora generato (quindi ancora $a_1 = 1$). Al mese 2 avremo tante coppie quante ne avevamo al mese 1 (stiamo ovviamente supponendo che le coppie non muoiano) più una per ogni coppia presente al mese 0 (le coppie dell'ultimo mese sono fertili ma non generano ancora). Al mese 3 avremo tante coppie quante ne avevamo al mese 2 più una nuova per ogni coppia presente al mese 1 (le coppie presenti al mese 2 ma non al mese 1 sono fertili ma non generano ancora). Iterando questo ragionamento vediamo che al mese $(n+1)$-esimo abbiamo tante coppie a_{n+1} quante sono quelle a_n del mese precedente più le nuove coppie generate da quelle a_{n-1} del mese precedente ancora (le coppie generate al mese n-esimo sono fertili ma non hanno ancora generato nel mese $n+1$).

2.9 Calcolo combinatorio

Ora che abbiamo visto cosa significa che due insiemi (finiti o infiniti) hanno lo stesso numero di elementi e abbiamo dato un preciso significato teorico al verbo "contare", vogliamo risolvere problemi pratici che consistono proprio nel dire quanti elementi ci sono in un insieme dato o nel contare quanti possibili configurazioni diverse può assumere un sistema dato sotto certe condizioni. Problemi di questo tipo sono l'oggetto di studio di quel ramo della matematica detta *combinatoria*.

Ci concentreremo inizialmente sulla seguente

Domanda: *In quanti modi è possibile permutare (ovvero cambiar d'ordine) n oggetti distinti?*

La risposta è data dal ragionamento seguente: dare una disposizione significa dire chi è il primo elemento, chi il secondo e così via. Ora, essendoci n elementi, il primo che scegliamo potrebbe essere uno qualunque di essi, quindi abbiamo n possibili scelte. Per ognuna di queste n scelte, il secondo elemento può essere scelto tra tutti gli elementi diversi da quello che abbiamo già scelto come primo, quindi abbiamo $n-1$ possibili scelte per ognuna delle n scelte precedenti. In tutto $n(n-1)$ scelte per i primi due elementi. Per ognuna di queste $n(n-1)$ scelte, il terzo elemento può essere scelto tra gli $n-2$ rimasti (tutti tranne i due già scelti come primo e secondo). Quindi abbiamo in

tutto $n(n-1)(n-2)$ scelte, etc., fino a che non esauriamo tutti gli elementi. Il numero di possibili scelte è quindi

$$n(n-1)(n-2)(n-3)\cdots 1 \tag{2.24}$$

ovvero n moltiplicato per tutti i numeri naturali precedenti escluso lo zero.
Il numero che compare nella (2.24) si dice *fattoriale di n* e si denota con il simbolo $n!$.
A scopo esemplificativo, scriviamo di seguito i primi fattoriali:

$$\begin{aligned}
1! &= 1\\
2! &= 2\cdot 1 = 2\\
3! &= 3\cdot 2\cdot 1 = 6\\
4! &= 4\cdot 3\cdot 2\cdot 1 = 24\\
5! &= 5\cdot 4\cdot 3\cdot 2\cdot 1 = 120\\
&\vdots
\end{aligned}$$

Per convenzione, si pone inoltre $0! = 1$.
Come si vede, il fattoriale di n cresce molto rapidamente all'aumentare di n.
Ad esempio, elenchiamo esplicitamente tutte le $3! = 6$ permutazioni di 3 elementi a, b, c:

$$abc,\ bac,\ acb,\ cba,\ bca, cab\,.$$

Il calcolo del numero di permutazioni di un insieme con n elementi fa parte del cosiddetto *calcolo combinatorio*, che si occupa di contare i modi per ordinare, raggruppare o scegliere gli elementi di insiemi finiti, sotto condizioni date.
Vediamo alcuni altri problemi tipici del calcolo combinatorio, nei quali il fattoriale ha un ruolo fondamentale.

(1) Supponiamo di voler contare quante sono le possili permutazioni di n elementi (possibilmente non distinti).

Un caso familiare di tale situazione si ha quando si vogliono considerare tutti i possibili anagrammi di una parola in cui alcune lettere si ripetono. Si consideri, per esempio, la parola "PAPPA": in tale parola compaiono due lettere di cui una, la A, si ripete due volte e l'altra, la P, si ripete tre volte, per un totale di 5 elementi non distinti tra loro. Quanti sono gli anagrammi di tale parola?

Iniziamo con l'osservare che, per quanto visto sopra, il numero totale di permutazioni possibili delle 5 lettere di tale parola è $5! = 120$. Tuttavia, questo non

significa che esistono 120 possibili anagrammi di tale parola. Infatti, in queste 120 permutazioni troveremo ogni anagramma di PAPPA ripetuto tante volte quante sono le permutazioni che scambiano solo le A tra loro o solo le P tra loro, perché tali permutazioni non modificano l'anagramma. Quante sono per ogni anagramma dato queste permutazioni che non lo modificano? Abbiamo $2! = 2$ permutazioni delle due A tra loro e, per ciascuna di queste, $3! = 6$ permutazioni delle tre P tra loro. In totale $2!3! = 12$ permutazioni che non cambiano un anagramma dato.

Quindi, il numero totale di permutazioni trovate, 120, va diviso per il numero di queste permutazioni che non cambiano l'anagramma corrispondente, cioè 12. Ne deriva che il numero di anagrammi distinti della parola PAPPA è $120/12 = 10$:

PAPPA, AAPPP, PPPAA, APAPP, PAAPP, PAPAP, APPAP, PPAAP, APPPA, PPAPA.

In generale, supponiamo di avere s elementi distinti in cui il primo si ripete k_1 volte, il secondo k_2 volte, e così via fino all'ultimo che si ripete k_s volte, per un totale di $k_1 + k_2 + \cdots + k_s = n$ elementi non distinti tra loro (nell'esempio precedente sarebbe $s = 2$, $k_1 = 2$, $k_2 = 3$ e $n = k_1 + k_2 = 5$). Esattamente come per il ragionamento fatto sopra, le possibili permutazioni di questi n elementi distinti sono $n!$, ma questo numero va diviso per il numero di permutazioni che non cambia veramente la configurazione trovata, che sono le permutazioni che scambiano i k_1 elementi che si ripetono tra loro, i k_2 che si ripetono tra loro, e così via fino all'ultimo gruppo di k_s elementi. Esattamente come abbiamo visto nel nostro esempio, tali permutazioni sono $k_1!k_2!\cdots k_s!$ (per ognuna delle $k_1!$ permutazioni del primo gruppo di elementi uguali, ne abbiamo $k_2!$ del secondo gruppo, $k_3!$ del terzo etc.).

In conclusione, il numero di permutazioni di n elementi in cui c'è un gruppo di k_1 elementi uguali, un altro di k_2 elementi uguali, e così via fino a k_s elementi uguali, è

$$\frac{n!}{k_1!k_2!\cdots k_s!} \tag{2.25}$$

dove $k_1 + k_2 + \cdots + k_s = n$. Queste permutazioni sono dette **permutazioni con ripetizioni** per distinguerle dalle permutazioni usuali in cui gli elementi sono tutti distinti, dette anche **permutazioni semplici**. Questo numero è anche detto **coefficiente polinomiale**.[15]

[15] Il nome deriva dal fatto che questo è il coefficiente del temine $x_1^{k_1}\cdots x_s^{k_s}$ nel polinomio $(x_1+\cdots+x_s)^n$.

Si noti che se gli elementi fossero tutti distinti senza ripetizioni, allora $k_1! = k_2! = \cdots = k_s! = 1$ e nella (2.25) rimarrebbe $n!$, che coincide col numero di permutazioni di n elementi distinti.

(2) Un'altra importante domanda a cui il calcolo combinatorio risponde è la seguente: dato un insieme X di n elementi e scelto un $k \leqslant n$, quanti sono i sottoinsiemi di X formati da k elementi?

Ad esempio, se $X = \{a, b, c\}$ (quindi $n = 3$) e $k = 2$, i sottoinsiemi di X con 2 elementi sono

$$\{a, b\},\ \{a, c\},\ \text{e}\ \{b, c\}\,.$$

Si noti che dal momento che un sottoinsieme è determinato dai suoi elementi indipendentemente dall'ordine in cui li mettiamo, il problema può essere riformulato nel modo seguente: in quanti modi diversi possiamo scegliere k elementi da un insieme di n (senza ripetizioni e senza tener conto dell'ordine)? Con una terminologia tipica del calcolo combinatorio, si dice anche che vogliamo sapere quante sono le **combinazioni semplici di n elementi di classe k.**

Rispondiamo ora alla domanda. Se X ha n elementi, e dobbiamo scegliere k elementi distinti tra questi n per formare un sottoinsieme S, per il primo elemento di S abbiamo n possibili scelte (uno qualunque degli elementi di X). Per ognuna di queste n scelte, il secondo elemento può essere scelto tra $n-1$ (tutti gli elementi di X tranne quello già scelto). Quindi per i primi due elementi abbiamo $n(n-1)$ possibili scelte. Per ognuna di queste, abbiamo $n-2$ possibilità per il terzo elemento (tutti gli elementi di X tranne i 2 già scelti), quindi $n(n-1)(n-2)$ possibilità per i primi 3 elementi. Iterando questo ragionamento risulta evidente che avremo $n(n-1)(n-2)(n-3)$ possibili scelte per i primi 4 elementi, e in generale saranno $n(n-1)(n-2)\cdots(n-k+1)$ scelte per k elementi.

Tuttavia, questo numero non è ancora la risposta corretta: ad esempio, per $n = 3$ e $k = 2$ come nell'esempio di sopra, otterremmo $n(n-1) = 3 \cdot 2 = 6$, mentre abbiamo visto che ci sono solo tre sottoinsiemi!

Questo perché, nel ragionamento appena fatto, ci sono scelte diverse del primo, del secondo etc. fino al k-esimo elemento che risultano nello stesso sottoinsieme. Ad esempio se in $\{a, b, c\}$ scegliamo a come primo e b come secondo, oppure b come primo a a come secondo, otteniamo chiaramente lo stesso sottoinsieme di due elementi $\{a, b\}$. In questo caso non ci sono altre scelte che danno questo stesso sottoinsieme, perché sono solo 2 le permutazioni possibili dei suoi due elementi.

In generale, lo stesso sottoinsieme di k elementi può essere ottenuto mediante esattamente $k!$ scelte diverse, ovvero tante quante sono le possibili permutazioni dei suoi elementi.

Concludiamo che, per avere il numero effettivo di sottoinsiemi di k elementi da un insieme di n, il numero $n(n-1)(n-2)\cdots(n-k+1)$ delle scelte che si possono fare va diviso per il numero $k!$ di scelte che in realtà danno gli stessi elementi disposti in ordine diverso. Quindi questo numero è

$$\frac{n(n-1)(n-2)\cdots(n-k+1)}{k!}.$$

Tale numero può essere riscritto in maniera compatta come segue. Moltiplicando numeratore e denominatore per $(n-k)(n-k-1)\cdots 2\cdot 1$, ovvero $(n-k)!$, si trova

$$\frac{n(n-1)(n-2)\cdots(n-k+1)(n-k)(n-k-1)\cdots 2\cdot 1}{k!(n-k)!}$$

Ma ora il numeratore è il prodotto di tutti i naturali da n fino ad arrivare a 1, ovvero $n!$. Quindi otteniamo

$$\frac{n!}{k!(n-k)!}. \tag{2.26}$$

L'espressione (2.26) appena scritta si denota con $\binom{n}{k}$, si chiama **coefficiente binomiale**[16] e si legge *n su k*.

Ad esempio, se $n=3$ e $k=2$, troviamo $\binom{n}{k}=\frac{3!}{2!1!}=\frac{6}{2}=3$, in accordo con l'esempio di sopra in cui abbiamo trovato esattamente 3 sottoinsiemi di due elementi in un insieme con 3 elementi.

Il coefficiente binomiale compare in molte importanti formule della matematica, quindi è importante conoscerne le proprietà. A tal fine dimostriamone due che, come vedremo tra poco, ci permettono di calcolare rapidamente i coefficienti binomiali al crescere di n senza usare la definizione.

Lemma 2.51. *I coefficienti binimiali soddisfano le seguenti proprietà:*

$$\binom{n}{k}=\binom{n}{n-k}, \tag{2.27}$$

$$\binom{n}{k}=\binom{n-1}{k}+\binom{n-1}{k-1}. \tag{2.28}$$

[16] In effetti, la (2.26) è un particolare caso del coefficiente polinomiale (2.25) quando $s=2$ (da cui binomiale). Infatti, in tal caso la (2.25) diventa $\frac{n!}{k_1!k_2!}$ con $k_1+k_2=n$, ovvero $k_2=n-k_1$. Quindi possiamo riscriverlo come $\frac{n!}{k_1!(n-k_1)!}$ che è, a parte il k_1 al posto di k, proprio la (2.26).

Dimostrazione. Per dimostrare la (2.27), basta scrivere il secondo membro in base alla definizione stessa di binomiale, cioè la (2.26) con $n-k$ al posto di k:

$$\binom{n}{n-k} = \frac{n!}{(n-k)![n-(n-k)]!} = \frac{n!}{(n-k)!(n-n+k)!} = \frac{n!}{(n-k)!k!} = \binom{n}{k}$$

Dimostriamo ora la (2.28). Per definizione di binomiale si ha

$$\begin{aligned}\binom{n-1}{k} + \binom{n-1}{k-1} &= \frac{(n-1)!}{k!(n-1-k)!} + \frac{(n-1)!}{(k-1)![(n-1)-(k-1)]!} = \\ &= \frac{(n-1)!}{k!(n-1-k)!} + \frac{(n-1)!}{(k-1)!(n-k)!} \qquad (2.29)\end{aligned}$$

Per sommare le due frazioni, osserviamo che $k!(n-k)!$ è un multiplo comune dei denominatori. Infatti, tale numero si ottiene moltiplicando il primo denominatore per $n-k$ (che moltiplicato per $(n-1-k)!$ lo fa diventare $(n-k)!$), ma anche moltiplicando il secondo denominatore per k (che moltiplicato per $(k-1)!$ lo fa diventare $k!$). Usando allora $k!(n-k)!$ come multiplo comune, possiamo sommare le frazioni esattamente come nel caso della somma di due frazioni numeriche. Otteniamo che la (2.29) è uguale a

$$\frac{(n-k)\cdot(n-1)! + k\cdot(n-1)!}{k!(n-k)!},$$

ovvero, mettendo in evidenza $(n-1)!$ nei due addendi a numeratore,

$$\frac{[(n-k)+k]\cdot(n-1)!}{k!(n-k)!} = \frac{n\cdot(n-1)!}{k!(n-k)!} = \frac{n!}{k!(n-k)!} = \binom{n}{k}$$

proprio come volevamo dimostrare. □

Come abbiamo accennato, la (2.28) ci permette di calcolare velocemente i coefficienti binomiali al crescere di n. Incominciamo col disporre in riga, per $n = 0,1,2,\ldots,$ i coefficienti binomiali $\binom{n}{k}$ al variare di $k = 0,1,\ldots,n$. Ovvero scriviamo la tabella

	$(k=0)$	$(k=1)$	$(k=2)$	$(k=3)$	$(k=4)$	$(k=5)$
$(n=0)$	$\binom{0}{0}$					
$(n=1)$	$\binom{1}{0}$	$\binom{1}{1}$				
$(n=2)$	$\binom{2}{0}$	$\binom{2}{1}$	$\binom{2}{2}$			
$(n=3)$	$\binom{3}{0}$	$\binom{3}{1}$	$\binom{3}{2}$	$\binom{3}{3}$		
$(n=4)$	$\binom{4}{0}$	$\binom{4}{1}$	$\binom{4}{2}$	$\binom{4}{3}$	$\binom{4}{4}$	
$(n=5)$	$\binom{5}{0}$	$\binom{5}{1}$	$\binom{5}{2}$	$\binom{5}{3}$	$\binom{5}{4}$	$\binom{5}{5}$
			$\dots$			

La (2.28) ci dice che l'entrata $\binom{n}{k}$ della tabella (che si trova all'incrocio di riga n-esima e colonna k-esima) si ottiene sommando l'entrata sopra di lei e quella che si trova a sinistra di quest'ultima. Infatti l'entrata sopra si trova nella stessa colonna, la k-esima, ma una riga sopra, la $(n-1)$-esima (corrisponde quindi al coefficiente binomiale $\binom{n-1}{k}$). Allo stesso modo, l'entrata alla sua sinistra è sempre nella riga sopra, la $(n-1)$-esima, e nella colonna più a sinistra, la $(k-1)$-esima (cioè corrisponde al coefficiente $\binom{n-1}{k-1}$).

Con questa regola si ricostruiscono rapidamente tutte le entrate della tabella conoscendo le prime due righe (in quanto $\binom{0}{0} = \binom{1}{0} = \binom{1}{1} = 1$):

	$(k=0)$	$(k=1)$	$(k=2)$	$(k=3)$	$(k=4)$	$(k=5)$
$(n=0)$	1					
$(n=1)$	1	1				
$(n=2)$	1	2	1			
$(n=3)$	1	3	3	1		
$(n=4)$	1	**4**	**6**	4	1	
$(n=5)$	1	5	**10**	10	5	1
			$\dots$			

Abbiamo messo in evidenza in grassetto il fatto che il 10 dell'ultima riga è stato ottenuto sommando il 6 sopra di lui più il 4 a sinistra del 6.

La tabella appena scritta si dice anche *triangolo di Pascal-Tartaglia*.

I coefficienti ottenuti servono, tra le altre cose, per calcolare le potenze $(a+b)^n$

di un binomio. Infatti, vale la formula

$$(a+b)^n = \\ = \binom{n}{0}a^n + \binom{n}{1}a^{n-1}b + \binom{n}{2}a^{n-2}b^2 + \cdots + \binom{n}{n-1}ab^{n-1} + \binom{n}{n}b^n .$$

In altre parole vale il seguente risultato.

Proposizione 2.52. *Per ogni $n \in \mathbb{N}$ vale la seguente formula:*

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^{n-k}b^k . \tag{2.30}$$

Qui il simbolo $\sum$, simbolo di sommatoria, indica che dobbiamo sommare tutti gli addendi del tipo $\binom{n}{k}a^{n-k}b^k$ con k che varia da 0 a n.

Ad esempio, applicando questa formula e guardando i coefficienti della tabella si trova

$$\begin{aligned}(a+b)^2 &= a^2 + 2ab + b^2\\ (a+b)^3 &= a^3 + 3a^2b + 3ab^2 + b^3\\ (a+b)^4 &= a^4 + 4a^3b + 6a^2b^2 + 4ab^3 + b^4\\ &\vdots\end{aligned}$$

La (2.30) può essere dimostrata per induzione nel modo seguente.

Dimostrazione della Proposizione 2.52. Innanzitutto, tale formula è vera per $n = 0$, in quanto, da una parte $(a+b)^0 = 1$, e dall'altra, si ha $\sum_{k=0}^{0}\binom{0}{k}a^{0-k}b^k = \binom{0}{0}a^0b^0 = 1$.

Supponiamo ora che la formula valga per n (ipotesi induttiva), ovvero

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^{n-k}b^k \tag{2.31}$$

e dimostriamo che vale per $n+1$, ovvero

$$(a+b)^{n+1} = \sum_{k=0}^{n+1} \binom{n+1}{k} a^{n+1-k}b^k. \tag{2.32}$$

Poiché $(a+b)^{n+1} = (a+b)(a+b)^n$, possiamo usare l'ipotesi induttiva e scrivere

$$\begin{aligned}(a+b)^{n+1} =& (a+b)\sum_{k=0}^{n}\binom{n}{k}a^{n-k}b^k = \\ =& a\sum_{k=0}^{n}\binom{n}{k}a^{n-k}b^k + b\sum_{k=0}^{n}\binom{n}{k}a^{n-k}b^k. \end{aligned} \tag{2.33}$$

Ora, possiamo portare l'a e il b che moltiplicano rispettivamente la prima e la seconda sommatoria all'interno delle sommatorie stesse, in quanto in generale per le sommatorie vale[17] la proprietà $c\sum_{j=1}^{N} x_j = \sum_{j=1}^{N} cx_j$. Si ottiene quindi

$$(a+b)^{n+1} = \sum_{k=0}^{n}\binom{n}{k}a^{n+1-k}b^k + \sum_{k=0}^{n}\binom{n}{k}a^{n-k}b^{k+1}.$$

Per motivi che saranno chiari nei passaggi successivi, separiamo ora dalla prima sommatoria il termine corrispondente a $k=0$, cioè $\binom{n}{0}a^{n+1-0}b^0 = a^{n+1}$, e dalla seconda sommatoria il termine che corrisponde a $k=n$, cioè $\binom{n}{n}a^{n-n}b^{n+1} = b^{n+1}$:

$$(a+b)^{n+1} = a^{n+1} + \sum_{k=1}^{n}\binom{n}{k}a^{n+1-k}b^k + \sum_{k=0}^{n-1}\binom{n}{k}a^{n-k}b^{k+1} + b^{n+1}.$$

Ora utilizziamo un trucco spesso usato quando si ha a che fare con sommatorie: nella seconda sommatoria denotiamo $k+1$ con h (ovvero $k = h-1$) e usiamo h come nuovo indice di sommatoria. Quando $k=0$, si ha $h=1$ e quando $k=n-1$ si ha $h=n$. Cambiati in questo modo gli estremi della sommatoria e sostituito ovunque in tale sommatoria $k = h-1$, si ha

$$(a+b)^{n+1} = a^{n+1} + \sum_{k=1}^{n}\binom{n}{k}a^{n+1-k}b^k + \sum_{h=1}^{n}\binom{n}{h-1}a^{n-h+1}b^h + b^{n+1}.$$

A questo punto, dal momento che l'indice di una sommatoria può essere denotato in qualunque modo, ridenominiamo h come k:

$$(a+b)^{n+1} = a^{n+1} + \sum_{k=1}^{n}\binom{n}{k}a^{n+1-k}b^k + \sum_{k=1}^{n}\binom{n}{k-1}a^{n-k+1}b^k + b^{n+1}.$$

[17] Infatti, si ha $c\sum_{j=1}^{N} x_j = c(x_1 + x_2 + \cdots + x_N) = cx_1 + cx_2 + \cdots + cx_N = \sum_{j=1}^{N} cx_j$, dove abbiamo applicato la proprietà distributiva nella seconda uguaglianza.

Dopo queste trasformazioni, in entrambe le sommatorie l'indice k varia sempre da 1 a n e in entrambe compare $a^{n-k+1}b^k$, moltiplicato nella prima per $\binom{n}{k}$ e nella seconda per $\binom{n}{k-1}$. Quindi possiamo raggrupparle, usando la proprietà[18] generale delle sommatorie $\sum_{j=1}^{N} x_j + \sum_{j=1}^{N} y_j = \sum_{j=1}^{N}(x_j + y_j)$ e mettendo in evidenza $a^{n-k+1}b^k$:

$$(a+b)^{n+1} = a^{n+1} + \sum_{k=1}^{n}\left[\binom{n}{k} + \binom{n}{k-1}\right] a^{n+1-k}b^k + b^{n+1}.$$

A questo punto, all'interno della parentesi quadra possiamo usare la proprietà (2.28) del coefficiente binomiale, che ci dice che $\binom{n}{k} + \binom{n}{k-1} = \binom{n+1}{k}$, ottenendo

$$(a+b)^{n+1} = a^{n+1} + \sum_{k=1}^{n}\binom{n+1}{k} a^{n+1-k}b^k + b^{n+1}.$$

Ora, possiamo inglobare di nuovo gli addendi a^{n+1} e b^{n+1}, che nei passaggi precedenti avevamo separato, all'interno della sommatoria centrale. Notiamo che $a^{n+1} = \binom{n+1}{0}a^{n+1-0}b^0$ e $b^{n+1} = \binom{n+1}{n+1}a^{n+1-(n+1)}b^{n+1}$. Quindi perchè nella sommatoria siano compresi anche a^{n+1} e b^{n+1} basta aggiungere $k = 0$ e $k = n+1$, ottenendo finalmente

$$(a+b)^{n+1} = \sum_{k=0}^{n+1}\binom{n+1}{k} a^{n+1-k}b^k.$$

Abbiamo quindi dimostrato la (2.32) partendo dall'ipotesi che fosse vera la (2.31), quindi come afferma il principio d'induzione la formula è vera per ogni $n \in \mathbb{N}$. □

(3) Un altro problema tipico del calcolo combinatorio è il seguente: dato un insieme X con n elementi, e scelto un $k \leqslant n$, in quanti modi possiamo scegliere k distinti elementi da X tenendo conto dell'ordine? In altre parole vogliamo contare le k-uple composte da elementi distinti (diversamente da quanto visto in (2) per i sottoinsiemi di cardinalità k in cui l'ordine degli elementi non conta).

Ad esempio, se $X = \{a, b, c\}$ (cioè $n = 3$) e scegliamo $k = 2$, abbiamo le seguenti possibilità:

$$ab, ba, ac, ca, bc, cb.$$

[18] Infatti, usando le proprietà associativa e commutativa della somma, che ci permettono di permutare e accorpare gli addendi come vogliamo, si nota che $\sum_{j=1}^{N} x_j + \sum_{j=1}^{N} y_j = (x_1 + x_2 + \cdots + x_N) + (y_1 + y_2 + \cdots + y_N)$ è uguale a $(x_1 + y_1) + (x_2 + y_2) + \cdots + (x_N + y_N) = \sum_{j=1}^{N}(x_j + y_j)$.

In calcolo combinatorio si dice che vogliamo determinare il numero delle **disposizioni semplici di n elementi di classe k.**

La risposta a questa domanda è facile se si tiene conto di quanto già detto per risolvere il problema (2). In quel caso abbiamo visto che il numero di sottoinsiemi di k elementi presi da X è $n(n-1)(n-2)\cdots(n-k+1)$ diviso per $k!$ (il numero di permutazioni dei k elementi del sottoinsieme) perché ordinamenti diversi di questi elementi danno lo stesso sottoinsieme.

Dal momento che qui vogliamo invece tener conto dell'ordine, non dobbiamo dividere per $k!$ e il numero cercato è semplicemente

$$n(n-1)(n-2)\cdots(n-k+1).$$

Allo scopo di scrivere questo numero in una forma più conveniente, moltiplichiamolo e dividiamolo per $(n-k)(n-k-1)\cdots 2\cdot 1$ (ovvero $(n-k)!$). In questo modo otteniamo

$$\begin{aligned} &n(n-1)(n-2)\cdots(n-k+1) = \\ =&\frac{n(n-1)(n-2)\cdots(n-k+1)(n-k)(n-k-1)\cdots 2\cdot 1}{(n-k)!} = \frac{n!}{(n-k)!}. \end{aligned}$$

(4) Se nel problema (3) ammettiamo anche le ripetizioni otteniamo le cosiddette **disposizioni con ripetizione di n elementi di classe k.**

Ovvero, dato un insieme X con n elementi, e scelto un $k \leqslant n$, ci chiediamo in quanti modi possiamo scegliere k elementi da X, eventualmente con ripetizioni e tenendo conto dell'ordine.

Ad esempio, se $X = \{a,b,c\}$ (cioè $n = 3$) e $k = 2$, abbiamo le seguenti possibilità:

$$aa, bb, cc, ab, ba, ac, ca, bc, cb.$$

In questo caso, la risposta è particolarmente semplice: il primo elemento può essere scelto in n modi (può essere un qualunque elemento di X), il secondo anche (non dobbiamo escludere il primo già scelto perché sono ammesse ripetizioni) e così via per tutti gli altri: quindi abbiamo $n\cdot n\cdot\cdots\cdot n = n^k$ possibilità, tenendo conto anche dell'ordine.

Si osservi che nello scegliere k elementi da X ammettendo ripetizioni e tenendo conto dell'ordine, stiamo equivalentemente considerando tutte le k-uple ordinate di elementi di X, ovvero l'insieme prodotto cartesiano X^k di X per se stesso k volte. Ad esempio per l'insieme $X = \{a,b,c\}$ considerato sopra, si ha

$$X^2 = \{(a,a),(b,b),(c,c),(a,b),(b,a),(a,c),(c,a),(b,c),(c,b)\}$$

ed è chiaro che dare le coppie ordinate o le disposizioni con ripetizione è sostanzialmente la stessa cosa.

(5) Infine, consideriamo le cosiddette **combinazioni con ripetizione di n elementi di classe k**: rispetto alle combinazioni semplici, ovvero i sottoinsiemi di k elementi, studiate in (2), ammettiamo anche le ripetizioni. Più precisamente, dato un insieme X con n elementi, e scelto un $k \leqslant n$, ci chiediamo in quanti modi possiamo scegliere k elementi da X, eventualmente con ripetizioni, ma sempre senza tener conto dell'ordine come nelle combinazioni semplici.

Ad esempio, se $X = \{a,b,c\}$ (cioè $n = 3$) e $k = 2$, abbiamo le seguenti possibilità:

$$aa, bb, cc, ab, ac, bc.$$

Notiamo che dare una di queste combinazioni significa semplicemente dire il numero di volte che si ripete a, il numero di volte che si ripete b e il numero di volte che si ripete c, con la condizione che la somma di questi tre numeri deve essere 2, perché in tutto dobbiamo avere solo 2 elementi. Questi numeri possono essere anche 0, ad esempio in aa, a si ripete 2 volte, b zero volte e c zero volte; in bc, a si ripete 0 volte, b una volta e c una volta.

Possiamo rappresentare allora ognuna di queste combinazioni come uno schema del tipo

$$* * \cdots * \circ * * \cdots * \circ * * \cdots *$$

dove il numero di asterischi della prima serie ci dice quante volte si ripete a, il numero di asterischi della seconda serie (dopo il primo cerchietto) ci dice quante volte si ripete b, e il numero di asterischi della terza serie (quelli dopo il secondo cerchietto) ci dice quante volte si ripete c.

Ad esempio, ab è rappresentato da $*\circ *\circ$ (il primo asterisco ci dice che a si ripete una volta, il secondo che b si ripete una volta, e l'assenza di asterischi dopo il secondo cerchietto significa che c si ripete 0 volte, ovvero non compare nella combinazione). O ancora, cc è rappresentato da $\circ\circ **$ (nessun asterisco prima del primo cerchietto significa che a si ripete 0 volte, nessun asterisco tra il primo e il secondo cerchietto significa che b si ripete 0 volte, 2 asterischi dopo il secondo cerchietto significano che c si ripete 2 volte).

In generale, se l'insieme X ha n elementi, avremo n serie di asterischi, separate da $n-1$ cerchietti, e il numero totale di asterischi deve essere k.

La nostra domanda equivale allora a chiedere: quanti possibili schemi di questo tipo con k asterischi e $n-1$ cerchietti possiamo formare?

Tali schemi saranno tanti quanti i modi diversi di permutare gli asterischi e i cerchietti tra loro a meno di permutazioni che lasciano lo schema invariato. Avendo in tutto $k+n-1$ elementi tra asterischi e cerchietti, le loro permutazioni sono $(k+n-1)!$. Dobbiamo però dividere per il numero di permutazioni che lasciano invariato uno schema dato, ovvero quelle che permutano i cerchietti tra loro e gli asterischi tra loro. Essendoci k asterischi, le permutazioni che li scambiano sono $k!$; mentre essendoci $n-1$ cerchietti, le permutazioni che li scambiano sono $(n-1)!$. Quindi dobbiamo dividere il numero totale di permutazioni per $k!(n-1)!$, ottenendo

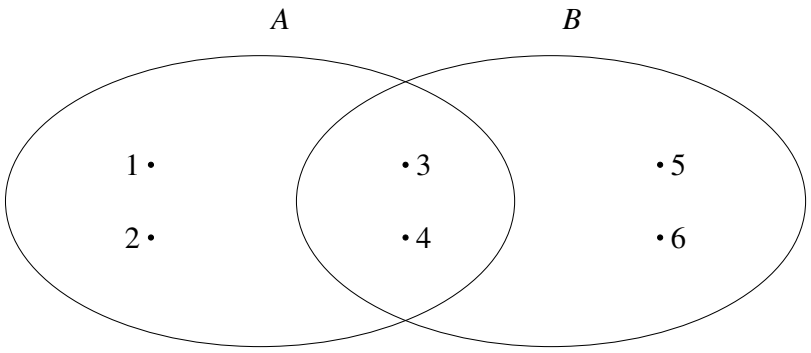
$$\frac{(k+n-1)!}{k!(n-1)!} = \binom{k+n-1}{k}.$$

Le formule appena viste possono essere riassunte nel seguente schema

	con ripetizioni	senza ripetizioni
ordine	n^k	$\frac{n!}{(n-k)!}$
no ordine	$\binom{n+k-1}{k}$	$\binom{n}{k}$

Concludiamo questa parte dedicata alla combinatoria con un altro problema di calcolo del numero di elementi di un certo insieme. Con le formule precedenti abbiamo imparato a contare il numero di elementi dell'insieme potenza $P(X)$ di un insieme finito dato e il numero di elementi del prodotto cartesiano X^n di un insieme finito dato per se stesso (n volte). Concludiamo ora il capitolo con una formula utile per il calcolo del numero di elementi di unioni e intersezioni di insiemi: il cosiddetto **principio di inclusione/esclusione**.

Per iniziare a illustrare tale principio, supponiamo di avere due insiemi finiti A e B e di voler calcolare il numero $|A \cup B|$ di elementi dell'unione $A \cup B$ in funzione del numero $|A|$ di elementi di A e del numero $|B|$ di elementi di B. Come si vede nel disegno seguente, se $A = \{1,2,3,4\}$ e $B = \{3,4,5,6\}$



non vale l'uguaglianza $|A \cup B| = |A| + |B|$, in quanto sommando $|A|$ e $|B|$ stiamo contando due volte gli elementi 3 e 4 dell'intersezione. Per ottenere un'uguaglianza dobbiamo quindi sottrarre il numero di elementi che stiamo contando più volte. La formula corretta diventa

$$|A \cup B| = |A| + |B| - |A \cap B|. \tag{2.34}$$

Usando questa semplice formula e uguaglianze elementari di teoria degli insiemi, possiamo vedere anche senza l'aiuto di disegni cosa succede se consideriamo l'unione $A \cup B \cup C$ di tre insiemi A, B, C. Infatti, per associatività dell'unione possiamo intendere l'unione $A \cup B \cup C$ come unione $(A \cup B) \cup C$ dei due insiemi $A \cup B$ e C. Per la formula (2.34) abbiamo allora

$$|A \cup B \cup C| = |A \cup B| + |C| - |(A \cup B) \cap C|. \tag{2.35}$$

Il primo addendo al secondo membro può essere riscritto usando di nuovo la formula (2.34), da cui

$$|A \cup B \cup C| = |A| + |B| - |A \cap B| + |C| - |(A \cup B) \cap C|. \tag{2.36}$$

Per quello che riguarda l'ultimo addendo, in base alla commutatività dell'intersezione a alla distributività dell'intersezione rispetto all'unione si ha $(A \cup B) \cap C = C \cap (A \cup B) = (C \cap A) \cup (C \cap B)$ e quindi possiamo applicare anche all'ultimo addendo della (2.36) la formula (2.35):

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

Tenendo conto che $(C \cap A) \cap (C \cap B) = A \cap B \cap C$ scriviamo

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

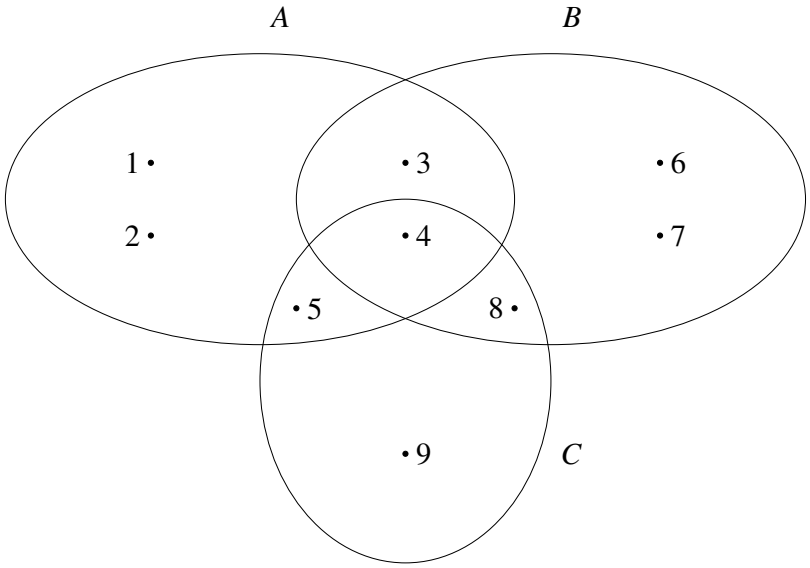
Sostituendo questa espressione nella (2.36) otteniamo allora

$$|A \cup B \cup C| = |A| + |B| - |A \cap B| + |C| - [|C \cap A| + |C \cap B| - |A \cap B \cap C|]$$

ovvero, riordinando gli addendi,

$$|A \cup B \cup C| = |A| + |B| + |C| - |A \cap B| - |A \cap C| - |B \cap C| + |A \cap B \cap C|. \quad (2.37)$$

Questa formula mostra che per contare il numero di elementi dell'unione di tre insiemi dobbiamo sommare le cardinalità dei tre insiemi, escludere tutte le possibili intersezioni di questi insiemi a due a due (che sono state contate due volte nella somma $|A|+|B|+|C|$ e poi reincludere però l'intersezione di tutti e tre gli insiemi, che è stata esclusa una volta di troppo quando abbiamo sottratto $|A \cap B|$, $|A \cap C|$ e $|B \cap C|$. Il seguente disegno, in cui $A = \{1,2,3,4,5\}$, $B = \{3,4,6,7,8\}$, $C = \{4,5,8,9\}$, può essere utile a illustrare quello che sta succedendo.



A questo punto ci si potrebbe chiedere se è possibile ricavare una formula per la cardinalità dell'unione $A_1 \cup A_2 \cup \cdots \cup A_n$ di n insiemi, con n qualunque. La risposta è affermativa e la formula è la seguente:

$$|A_1 \cup A_2 \cup \cdots \cup A_n| = \sum_{k=1}^{n} (-1)^{k-1} \sum_{1 \leqslant i_1 < \cdots < i_k \leqslant n} |A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k}|\,. \quad (2.38)$$

La formula esprime il fatto che per calcolare la cardinalità dell'unione bisogna sommare le cardinalità degli insiemi dati, sottrarre le cardinalità delle intersezioni di due

insiemi (tutte quelle possibili), sommare le cardinalità delle intersezioni di tre insiemi (tutte quelle possibili), sottrarre le cardinalità delle intersezioni di quattro insiemi (tutte quelle possibili), e così via, a segni alterni, includendo ed escludendo (da cui il nome di principio di inclusione/esclusione). La formula (2.38) può essere dimostrata per induzione.

Esempio 2.53. Vediamo un esempio di applicazione delle formule appena viste. Supponiamo che i calzini contenuti in un cassetto possano essere grossi, bianchi e lunghi. Sapendo che il cassetto contiene 7 calzini grossi, che i calzini bianchi e grossi sono 2, quelli lunghi e grossi 3, quelli contemporaneamente grossi, bianchi e lunghi 2, e che ci sono 5 calzini che sono bianchi o lunghi, quanti sono i calzini in totale?
Se denotiamo B l'insieme dei calzini bianchi, L l'insieme dei calzini lunghi e G l'insieme dei calzini grossi, il numero totale dei calzini è dato da $|G \cup B \cup L|$, che come sappiamo dal principio di inclusione/esclusione è

$$|G \cup B \cup L| = |G| + |B| + |L| - |B \cap L| - |B \cap G| - |G \cap L| + |B \cap G \cap L|.$$

In base ai dati del problema, abbiamo $|G| = 7$, $|B \cap G| = 2$, $|G \cap L| = 3$, $|G \cap B \cap L| = 2$ e $|B \cup L| = 5$, quindi

$$|G \cup B \cup L| = 7 + |B| + |L| - |B \cap L| - 2 - 3 + 2 = |B| + |L| - |B \cap L| + 4.$$

I dati del problema non ci danno le quantità $|B|$, $|L|$, $|B \cap L|$, ma ci danno $|B \cup L| = 5$, che in base al principio di inclusione/esclusione (caso di due insiemi) è uguale esattamente alla somma $|B| + |L| - |B \cap L|$ che compare nell'espressione di $|G \cup B \cup L|$. Sostituendo si ha allora

$$|G \cup B \cup L| = |B| + |L| - |B \cap L| + 4 = |B \cup L| + 4 = 5 + 4 = 9.$$

2.10 L'insieme delle permutazioni come gruppo

Definizione 2.54. Per ogni numero naturale n, le funzioni biiettive dall'insieme $\{1, 2, \ldots, n\}$ in se stesso si dicono **permutazioni di n elementi**.

Questo nome è dovuto al fatto che una tale funzione deve assegnare a ogni numero tra 1 e n un numero tra 1 e n evitando ripetizioni (altrimenti non sarebbe iniettiva) e in modo che tutti compaiano come immagini (altrimenti non sarebbe suriettiva): quindi, essa non fa altro che "cambiare l'ordine" degli elementi $1, 2, \ldots, n$. Ad esempio, se $n = 3$, un esempio di funzione biiettiva da $\{1, 2, 3\}$ in se stesso si ottiene ponendo

$$f(1) = 3$$
$$f(2) = 1$$
$$f(3) = 2$$

oppure

$$f(1) = 3$$
$$f(2) = 2$$
$$f(3) = 1$$

che scambia 3 e 1 tra loro lasciando fisso 2.

Si noti che anche la funzione identica su $\{1, 2, \ldots, n\}$, essendo chiaramente una funzione biiettiva, è una permutazione in base alla definizione data.

Definizione 2.55. In generale, dato un insieme $A = \{a_1, a_2, \ldots a_n\}$ con n elementi, chiameremo **permutazione di** A (o **permutazione di** $a_1, a_2, \ldots, a_n$) una funzione biiettiva $\{1, 2, \ldots, n\} \longrightarrow A$.

Ad esempio se $A = \{a, b, c\}$, la funzione $f\colon \{1, 2, 3\} \longrightarrow \{a, b, c\}$ definita da

$$f(1) = b$$
$$f(2) = a$$
$$f(3) = c$$

è una permutazione di a, b, c.

Quante sono le possibili permutazioni di n elementi $a_1, a_2, \ldots, a_n$? Ricordiamo che abbiamo già calcolato questo numero all'inizio della Sezione 2.9 rispondendo alla domanda a pagina 63. Aveveamo quindi trovato che le scelte possibili come immagini di $1, 2, \ldots, n$ che diano una funzione biiettiva sono $n!$.
In questa sezione vogliamo considerare l'insieme delle permutazioni di $\{1, 2, \ldots, n\}$, che da ora denoteremo con S_n, assieme all'operazione di composizione di funzioni, e studiarne le proprietà; non diversamente da quello che si fa quando si considera ad esempio l'insieme dei numeri naturali $\mathbb{N}$, assieme a un'operazione (ad esempio la somma o il prodotto) e se ne studiano le proprietà.

Prima di fare ciò, dobbiamo però essere sicuri che quando componiamo due permutazioni otteniamo ancora una permutazione. Non avrebbe infatti senso studiare le proprietà di questa operazione su S_n se succedesse che quando la applichiamo rischiamo di ottenere qualcosa che non sta più in S_n, allo stesso modo in cui non ha senso per esempio

studiare le proprietà della sottrazione su $\mathbb{N}$ quando non sempre questa operazione si può fare in questo insieme (pur essendo 2 e 5 numeri naturali, 2-5 non è un numero naturale).

Dal momento che una permutazione per definizione è una funzione biiettiva da $X = \{1, 2, \ldots, n\}$ in sè, basta ricordare che per il Lemma 2.44 la composizione di due funzioni biiettive è ancora biiettiva. Ora che siamo sicuri che la composizione di permutazioni sia ancora una permutazione, e quindi che l'operazione di composizione sia ben definita sull'insieme S_n delle permutazioni, possiamo studiarne le proprietà.
Abbiamo già visto che la composizione di funzioni è associativa, si veda la (2.11). Sappiamo inoltre che in generale la composizione non è commutativa, e non è difficile mostrare un esempio di permutazioni che non commutano, ad esempio sull'insieme $\{1, 2, 3\}$ se definiamo f e g come segue

$$\begin{array}{ccc} 1 & \mapsto & 2 \\ 2 & \mapsto & 1 \\ 3 & \mapsto & 3 \end{array}$$
$$f$$

$$\begin{array}{ccc} 1 & \mapsto & 3 \\ 2 & \mapsto & 2 \\ 3 & \mapsto & 1 \end{array}$$
$$g$$

allora si ha

$$\begin{array}{ccc} 1 & \mapsto & 2 \\ 2 & \mapsto & 3 \\ 3 & \mapsto & 1 \end{array}$$
$$g \circ f$$

mentre

$$\begin{array}{ccc} 1 & \mapsto & 3 \\ 2 & \mapsto & 1 \\ 3 & \mapsto & 2 \end{array}$$
$$f \circ g$$

Un'altra proprietà della composizione sull'insieme S_n è che in S_n esiste un elemento neutro per tale operazione. Come abbiamo visto a pagina 50, l'elemento neutro per la composizione è la funzione identica id_X, che appartiene all'insieme S_n delle permutazioni di $X = \{1, 2, \ldots, n\}$ in quanto biiettiva. Per semplicità di notazione, da questo momento in poi denoteremo la funzione identica semplicemente con id.

Infine, essendo ogni permutazione una funzione $f\colon X \longrightarrow X$ biiettiva, essa, come abbiamo visto nella Proposizione 2.41, è invertibile, ovvero esiste una funzione $g\colon X \longrightarrow X$

tale che $g \circ f = f \circ g = id$.
Chiaramente, anche la funzione g inversa di f è biiettiva, perché le uguaglianze $g \circ f = f \circ g = id$ ci dicono che anche lei è invertibile (ha f come sua inversa). Quindi anche g, l'inversa di f, che da questo momento denoteremo con f^{-1}, è una permutazione di X.
In altre parole S_n contiene l'inversa di ogni sua permutazione.

Le proprietà che stiamo mettendo in evidenza rientrano in quelle elencate nella seguente, importantissima

Definizione 2.56. Un **gruppo** è un insieme G dotato di un'operazione (che denotiamo con il simbolo $\cdot$) per cui valgono le tre seguenti proprietà.

1. (**Associatività.**) Per ogni $g_1, g_2, g_3 \in G$ si ha $(g_1 \cdot g_2) \cdot g_3 = g_1 \cdot (g_2 \cdot g_3)$.
2. (**Elemento neutro.**) Esiste un elemento neutro, ovvero esiste un $u \in G$ tale che $g \cdot u = u \cdot g = g$ per ogni $g \in G$.
3. (**Inverso.**) Ogni $g \in G$ ha un inverso in G, ovvero per ogni $g \in G$ esiste $g^{-1} \in G$ tale che $g \cdot g^{-1} = g^{-1} \cdot g = u$

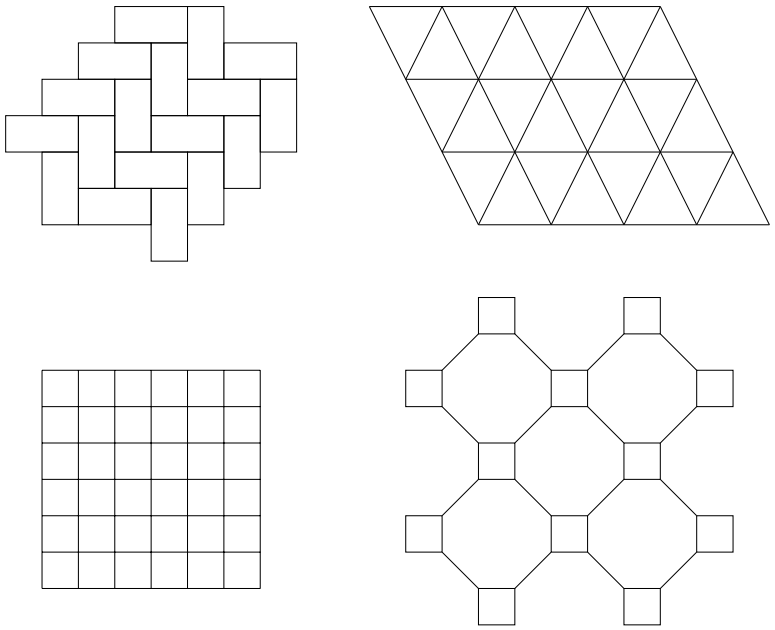
Si usa anche scrivere che $(G, \cdot)$ è un gruppo.

Definizione 2.57. Un gruppo $(G, \cdot)$ si dice **abeliano** o **commutativo** se $\cdot$ gode della proprietà commutativa, cioè se per due qualunque elementi $g_1, g_2 \in G$ vale $g_1 \cdot g_2 = g_2 \cdot g_1$.

Quindi $(S_n, \circ)$ (l'insieme delle permutazioni dotato dell'operazione di composizione) è un gruppo (non abeliano).

Esempio 2.58. Un altro esempio di gruppo, stavolta abeliano, è $(\mathbb{Z}, +)$, ovvero l'insieme degli interi dotato dell'operazione di somma: l'elemento neutro è il numero 0, e l'inverso di ogni numero intero n è semplicemente $-n$.
Invece, $(\mathbb{Z}, \cdot)$, cioè sempre l'insieme degli interi ma dotato della moltiplicazione, non è un gruppo. Pur essendo valide la proprietà associativa e l'esistenza dell'elemento neutro (in questo caso il numero 1), non è vero che ogni elemento di $\mathbb{Z}$ ha un inverso rispetto alla moltiplicazione (ad esempio, l'inverso di 2 è $\frac{1}{2}$, che è un razionale non intero).

Esempi di gruppi si trovano numerosi in ogni campo della matematica e nelle sue applicazioni. Ad esempio, in geometria, data una figura nel piano l'insieme delle trasformazioni che lasciano la figura invariata mandandola in se stessa (ad esempio, dato un quadrato la rotazione di 90 gradi attorno al centro del quadrato manda il quadrato in sè) forma un gruppo, detto *gruppo di simmetria della figura*. Questa definizione si estende anche a figure non limitate, ad esempio i motivi geometrici illlimitati, spesso usati nelle decorazioni o in architettura, che ricoprono in maniera regolare il piano per ripetizione di un motivo. Si consideri ad esempio il seguente disegno



Queste sono figure infinite con un loro gruppo di simmetria, e grazie alla teoria dei gruppi si può dimostrare che esistono solo 17 possibili gruppi di simmetria di tali motivi. Infine, la nozione di gruppo si rivela fondamentale in tutta la fisica contemporanea: grazie alla teoria dei gruppi si è riusciti a prevedere l'esistenza di particelle elementari che poi sono state effettivamente osservate.

Mettiamo ora in evidenza un'altra importante differenza tra il gruppo delle permutazioni $(S_n, \circ)$ e quello degli interi $(\mathbb{Z}, +)$. In $\mathbb{Z}$ ogni elemento può essere scritto usando solo il numero 1 o il suo inverso -1 e l'operazione $+$:

$$2 = 1 + 1, 3 = 1 + 1 + 1, 4 = 1 + 1 + 1 + 1, \ldots$$

$$-2 = (-1) + (-1), -3 = (-1) + (-1) + (-1), \ldots$$

$$0 = 1 + (-1)$$

In questo senso, quindi, è sufficiente un elemento a *generare* tutto $\mathbb{Z}$. Questo è un caso particolare della seguente

Definizione 2.59. Dato un gruppo $(G, \cdot)$, si dice che $g_1, g_2, \ldots, g_s$ **generano** G (o **sono generatori di** G) se ogni elemento di G può essere scritto come prodotto finito dei g_i e dei loro inversi.

Il gruppo $(\mathbb{Z}, +)$ ha quindi la caratteristica di essere generato da un solo elemento, 1. I gruppi che godono di questa proprietà son detti **ciclici**.

Si ha invece che il gruppo S_n non è ciclico, a meno che $n = 2$ (in questo caso, infatti, S_2 contiene solo l'identità *id* e la permutazione f che scambia 1 e 2, che genera *id* in quanto $f \circ f = id$).

Troveremo ora un insieme di generatori del gruppo S_n. Più precisamente, chiamiamo **trasposizioni** le permutazioni di S_n che scambiano tra loro due elementi e lasciano fissi tutti gli altri. Ad esempio, in S_4, la permutazione

$$1 \mapsto 1$$

$$2 \mapsto 3$$

$$3 \mapsto 2$$

$$4 \mapsto 4$$

è una trasposizione (scambia tra loro 2 e 3 lasciando fissi 1 e 4).

Mostreremo ora che qualunque permutazione può essere scritta come composizione di trasposizioni. Da questo si deduce che le trasposizioni generano il gruppo S_n.

Per far ciò, procederemo in modo costruttivo. Iniziamo col mostrare su un esempio come effettivamente, data una permutazione, si può trovare la sua decomposizione in trasposizioni. Consideriamo la seguente permutazione σ in S_6:

$$1 \mapsto 4$$

$$2 \mapsto 1$$

$$3 \mapsto 6$$

$$4 \mapsto 2$$

$$5 \mapsto 5$$

$$6 \mapsto 3$$

Procediamo come segue: la permutazione manda 1 in 4, 4 in 2, e 2 in 1. Questi tre elementi vengono quindi permutati tra loro in quello che si chiama un **ciclo** di lunghezza 3.

Usiamo la notazione $(1\ 4\ 2)$ per indicare tale ciclo (ogni numero che appare nel ciclo viene mandato nel successivo, e l'ultimo viene rimandato nel primo).

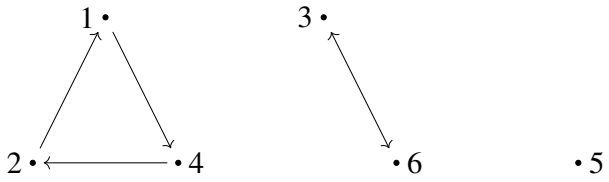
Concentriamoci ora sul primo elemento rimasto fuori da questo ciclo, cioè 3. Questo viene mandato in 6, e 6 viene mandato in 3. Quindi questi due elementi vengono permutati tra loro in un ciclo di lunghezza 2, che seguendo la notazione di sopra denotiamo con $(3\ 6)$ (si noti che un ciclo di lunghezza 2 è una trasposizione).

Infine, l'ultimo elemento che rimane fuori dai cicli trovati è 5, che viene fissato dalla

permutazione, quindi non appartiene a nessun ciclo[19].
La nostra permutazione è quindi scrivibile come *composizione di cicli disgiunti*:

$$\sigma = (3\ 6)(1\ 4\ 2) \tag{2.39}$$

(da questo momento per facilità di notazione ometteremo spesso il simbolo di composizione) e agisce sugli elementi $1, 2, 3, 4, 5, 6$ come nel disegno seguente.



Quanto visto sopra può essere generalizzato ad una qualunque permutazione per ottenere il seguente risultato.

Proposizione 2.60. *Ogni permutazione è la composizione di cicli disgiunti.*

Idea della dimostrazione. Sia $\sigma \in S_n$ una permutazione. Consideriamo l'immagine dell'elemento 1, cioè l'elemento $\sigma(1)$. Questo viene mandato da σ in un altro elemnto di $\{1, \dots, n\}$ potenzialmente diverso. Stiamo quindi considerando l'elemento $\sigma \circ \sigma(1)$. Iterando questo procedimento otterremo nuovamente l'elemento 1 dopo un certo numero k_1 di passaggi. In altre parole avremo che $\big(1,\ \sigma(1),\ \sigma(\sigma(1)),\ \dots, \sigma^{k_1}(1)\big)$ è un ciclo. Ora se $k_1 = n$ abbiamo dimostrato la tesi. In caso contrario, cioè se $k_1 < n$, consideriamo il primo elemento m in $\{1, \dots, n\}$ che non appartiene al ciclo appena trovato. Con lo stesso ragionamento otteniamo un ciclo $\big(m, \sigma(m), \dots, \sigma^{k_m}(m)\big)$. Dato che n è un numero finito, dopo al più n passaggi otterremo una decomposizione di σ in cicli. Per dimostrare che i cicli così ottenuti sono effettivamente disgiunti dobbiamo usare il fatto che una permutazione è una funzione biiettiva. □

Ora, il nostro obiettivo è decomporre ogni permutazione in trasposizioni. Se riusciamo a mostrare che ogni ciclo può essere decomposto in trasposizioni avremo raggiunto lo scopo. In effetti, dato un ciclo $(a_1\ a_2\ \cdots\ a_k)$ di lunghezza k, si può dimostrare che si ha sempre

$$(a_1\ a_2\ \cdots\ a_{k-1}\ a_k) = (a_1\ a_k)(a_1\ a_{k-1}) \cdots (a_1\ a_2). \tag{2.40}$$

Non diamo una dimostrazione generale di tale fatto: limitiamoci a illustrarlo con un esempio. Il ciclo $(1\ 4\ 2)$ in base alla (2.40) si dovrebbe decomporre come $(1\ 2)(1\ 4)$. Verifichiamolo esplicitamente: ricordando che le permutazioni si applicano da destra a sinistra, si ha che $(1\ 2)(1\ 4)$ agisce su 1,2,4 come segue:

[19] O, volendo, possiamo interpretarlo come un ciclo di lunghezza 1 e denotarlo (5), ma in generale lo si omette.

$$\begin{array}{ccccc} 1 & \mapsto & 4 & \mapsto & 4 \\ 2 & \mapsto & 2 & \mapsto & 1 \\ 4 & \mapsto & 1 & \mapsto & 2 \\ & (1\,4) & & (1\,2) & \end{array}$$

cioè esattamente come il ciclo $(1\ 4\ 2)$ che manda $1 \mapsto 4$, $4 \mapsto 2$, $2 \mapsto 1$.
In conclusione, la (2.39) può essere riscritta come

$$\sigma = (3\ 6)(1\ 2)(1\ 4)$$

ovvero come prodotto di trasposizioni.

Quello che abbiamo mostrato è un modo per determinare la decomposizione di una permutazione in trasposizioni, tuttavia si può vedere che tale decomposizione non è in generale unica. Ad esempio, è facile verificare che la permutazione

$$\begin{array}{ccc} 1 & \mapsto & 3 \\ 2 & \mapsto & 4 \\ 3 & \mapsto & 1 \\ 4 & \mapsto & 2 \end{array}$$

si decompone sia come $(1\ 3)(2\ 4)$ che come $(1\ 4)(1\ 2)(4\ 3)(1\ 4)$.
Tuttavia, vale il seguente risultato.

Teorema 2.61. *Il numero di trasposizioni in cui si decompone una permutazione data è o sempre pari o sempre dispari.*

Alla luce di ciò, possiamo dare la seguente

Definizione 2.62. Una permutazione si dice **pari** (risp. **dispari**) se si decompone in un numero pari (risp. dispari) di trasposizioni.

L'identità id è una permutazione pari in quanto può essere realizzata come prodotto di una qualunque trasposizione τ per se stessa. Infatti $\tau \circ \tau = id$ perché τ scambia due elementi tra loro lasciando gli altri fissi, e riapplicando τ gli unici due elementi invertiti vengono di nuovo scambiati tornando alla situazione iniziale.
Osserviamo che se componiamo due permutazioni σ e σ' entrambe pari, il risultato sarà ancora una permutazione pari. Questo perché se la prima si decompone in $2k$ trasposizioni

$$\sigma = \tau_1 \tau_2 \cdots \tau_{2k}$$

e la seconda in $2l$ trasposizioni

$$\sigma' = \tau'_1 \tau'_2 \cdots \tau'_{2l},$$

allora si ha chiaramente

$$\sigma\sigma' = \tau_1\tau_2\cdots\tau_{2k}\tau'_1\tau'_2\cdots\tau'_{2l}.$$

Cioè $\sigma\sigma'$ si decompone come prodotto di $2k+2l = 2(k+l)$ trasposizioni, ed è quindi anche lei una permutazione pari.

Inoltre, l'inversa σ^{-1} di una permutazione pari σ, sarà una permutazione pari. Infatti, se $\sigma = \tau_1\tau_2\cdots\tau_{2k-1}\tau_{2k}$, è facile vedere che la sua inversa è

$$\sigma^{-1} = \tau_{2k}\tau_{2k-1}\cdots\tau_2\tau_1, \tag{2.41}$$

cioè il prodotto delle stesse trasposizioni ma nell'ordine inverso (quindi sempre un numero pari).

Questo è un caso particolare del seguente risultato, che ci dice come calcolare l'inversa di un prodotto in un gruppo qualunque.

Lemma 2.63. *Sia G un gruppo e siano $g_1, g_2, \ldots, g_k \in G$. Allora*

$$(g_1g_2\cdots g_k)^{-1} = g_k^{-1}\cdots g_2^{-1}g_1^{-1} \tag{2.42}$$

Prima di dimostrare il lemma, osserviamo che da esso segue subito la (2.41). Infatti, in base al lemma si ha che l'inversa di $\tau_1\tau_2\cdots\tau_{2k-1}\tau_{2k}$ è $\tau_{2k}^{-1}\tau_{2k-1}^{-1}\cdots\tau_2^{-1}\tau_1^{-1}$. Ma come abbiamo già osservato sopra, per qualunque trasposizione τ si ha $\tau\circ\tau = id$, che significa che τ ha come inversa se stessa, ovvero $\tau^{-1} = \tau$. Quindi $\tau_{2k}^{-1}\tau_{2k-1}^{-1}\cdots\tau_2^{-1}\tau_1^{-1} = \tau_{2k}\tau_{2k-1}\cdots\tau_2\tau_1$, che dimostra la (2.41).

Dimostrazione del Lemma 2.63. In base alla definizione di inversa basta vedere se moltiplicando $g_1g_2\cdots g_{k-1}g_k$ con $g_k^{-1}g_{k-1}^{-1}\cdots g_2^{-1}g_1^{-1}$ si ottiene l'elemento neutro:

$$(g_1g_2\cdots g_{k-1}g_k)(g_k^{-1}g_{k-1}^{-1}\cdots g_2^{-1}g_1^{-1}) =$$

(per associatività dell'operazione in un gruppo, possiamo disporre le parentesi come vogliamo)

$$= (g_1g_2\cdots g_{k-1})(g_kg_k^{-1})(g_{k-1}^{-1}\cdots g_2^{-1}g_1^{-1}) = \tag{2.43}$$

Ma la parentesi centrale $g_kg_k^{-1}$ è uguale all'elemento neutro u, quindi la (2.43) si riduce a

$$= (g_1g_2\cdots g_{k-1})u(g_{k-1}^{-1}\cdots g_2^{-1}g_1^{-1}) = \tag{2.44}$$

ovvero, tenendo conto della definizione di elemento neutro,

$$= (g_1g_2\cdots g_{k-1})(g_{k-1}^{-1}\cdots g_2^{-1}g_1^{-1}). \tag{2.45}$$

A questo punto possiamo nello stesso modo eliminare g_{k-1}: infatti, sempre per associatività la (2.45) si scrive

$$= (g_1 g_2 \cdots)(g_{k-1} g_{k-1}^{-1})(\cdots g_2^{-1} g_1^{-1}) = \tag{2.46}$$

e come prima la parentesi centrale si elimina in quanto $g_{k-1} g_{k-1}^{-1} = u$. Continuando così, si eliminano via via tutti i fattori del prodotto, fino a che non rimane solo $g_1 g_1^{-1} = u$. La dimostrazione è conclusa[20]. □

Quanto detto mostra in effetti che l'insieme delle permutazioni pari in S_n, che si denota A_n, può essere considerato un gruppo a sè, in quanto quando compongo due elementi di A_n rimango in A_n (cioè l'operazione di composizione è ben definita dentro A_n), l'elemento neutro sta dentro A_n e l'inversa di ogni elemento di A_n sta ancora dentro A_n.

Definizione 2.64. Un gruppo G' contenuto in un gruppo G si dice **sottogruppo di** G (si intende che devono essere gruppi rispetto alla stessa operazione).

Quindi A_n è un sottogruppo di S_n. Notiamo che invece il sottoinsieme di S_n costituito dalle permutazioni dispari non forma un sottogruppo: basterebbe già il fatto che l'elemento neutro *id* (che come abbiamo osservato sopra è una permutazione pari) non sta in tale sottoinsieme. Inoltre possiamo notare che il prodotto di due permutazioni dispari non è più dispari in quanto se $\sigma = \tau_1 \tau_2 \cdots \tau_{2k+1}$ e $\sigma' = \tau'_1 \tau'_2 \cdots \tau'_{2l+1}$, allora $\sigma\sigma' = \tau_1 \tau_2 \cdots \tau_{2k+1} \tau'_1 \tau'_2 \cdots \tau'_{2l+1}$ risulta essere prodotto di $(2k+1) + (2l+1) = 2(k+l+1)$ trasposizioni, cioè un numero pari.

Osservazione 2.65. Un'importante applicazione delle proprietà dei gruppi A_n e S_n in matematica è stata la dimostrazione del fatto che *non esiste una formula risolutiva generale per risolvere le equazioni di grado superiore al quarto*. In altre parole, non esiste una formula generale che permetta di trovare le radici di un polinomio di grado $d \geqslant 5$. La dimostrazione, che fa parte della cosiddetta teoria di Galois, fa uso del fatto che a ogni polinomio di grado n si può associare un gruppo, opportunamente definito, che permuta le sue radici, quindi si può pensare come un sottogruppo del gruppo delle permutazioni S_n. La risolubilità dell'equazione determinata da tale polinomio corrisponde a una particolare proprietà di questi gruppi di permutazioni. Si dimostra che S_n e A_n non soddisfano questa proprietà se $n \geqslant 5$.

Le decomposizione in cicli disgiunti di una permutazione σ è importante anche per ottenere un'altra informazione: l'*ordine di* σ.
Il fatto che S_n sia finito ha come conseguenza il fatto che per ogni permutazione $\sigma \in S_n$ esista un numero naturale k per cui $\sigma^k = id$, dove con σ^k intendiamo la composizione

[20] Dovremmo verificare anche che $(g_k^{-1} g_{k-1}^{-1} \cdots g_2^{-1} g_1^{-1})(g_1 g_2 \cdots g_{k-1} g_k) = u$, ma i calcoli sono analoghi.

di σ con se stessa k volte. Ad esempio, consideriamo in S_3 il ciclo $(1\ 2\ 3)$, ovvero la permutazione

$$\begin{array}{ccc} 1 & \mapsto & 2 \\ 2 & \mapsto & 3 \\ 3 & \mapsto & 1 \end{array}$$

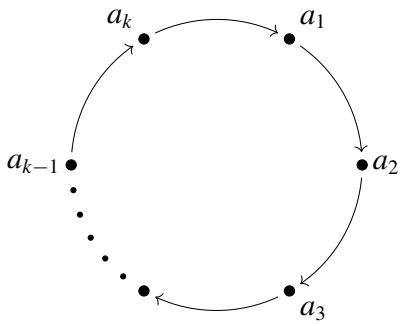
Si vede allora che $\sigma^3 = id$, ovvero che componendo σ tre volte con se stessa si ottiene l'identità:

$$\begin{array}{ccccccc} 1 & \mapsto & 2 & \mapsto & 3 & \mapsto & 1 \\ 2 & \mapsto & 3 & \mapsto & 1 & \mapsto & 2 \\ 3 & \mapsto & 1 & \mapsto & 2 & \mapsto & 3 \\ & \sigma & & \sigma & & \sigma & \end{array}$$

Chiaramente, se continuare ad applicare σ per la quarta volta sarebbe come comporre id con σ, ovvero $\sigma^4 = \sigma$, componendo ancora una volta otterremo $\sigma^5 = \sigma^2$ e componendo per la sesta volta avremmo $\sigma^6 = \sigma^3$, ovvero $\sigma^6 = id$. Iterando questo argomento, è facile vedere che in realtà vale $\sigma^k = id$ per tutti i multipli di 3, quindi ci sono in realtà infiniti numeri naturali per cui $\sigma^k = id$. Chiamiamo ordine il più piccolo numero naturale positivo per cui vale questa uguaglianza: nel nostro caso esso è chiaramente 3. In generale si può dare la seguente

Definizione 2.66. Sia G un gruppo con elemento neutro u e sia g un suo elemento. Si dice **ordine** di g il più piccolo intero positivo k tale che $g^k = u$.

Ora impareremo a calcolare l'ordine di qualunque permutazione σ senza dover applicare σ con se stessa fino a ottenere l'identità.
Per prima cosa, osserviamo che un ciclo $\sigma = (i_1\ i_2\ \ldots\ i_k)$ di lunghezza k ha ordine esattamente k.
Di ciò ci si può facilmente convincere guardando la rappresentazione grafica del ciclo:



Come si vede, ogni applicazione del ciclo manda un qualunque elemento nell'elemento successivo del ciclo (nel verso indicato dalla freccia), quindi applicando k volte il ciclo ogni elemento viene rimandato in se stesso.

A questo punto, per calcolare l'ordine di una generica permutazione σ, basta usare la sua decomposizione come prodotto di cicli disgiunti. Supponiamo che σ si decomponga in m cicli $C_1, C_2, \ldots, C_m$ rispettivamente di lunghezze $L_1, L_2, \ldots, L_m$. Essendo i cicli disgiunti, quando componiamo σ con se stessa ogni ciclo è composto con se stesso. Il primo ciclo C_1 ci darà l'identità se iteriamo la permutazione un numero di volte pari a un qualunque multiplo della sua lunghezza L_1; analogamente, il secondo ciclo C_2 ci darà l'identità se iteriamo la permutazione un numero di volte pari a un qualunque multiplo della sua lunghezza L_2, e così via per tutti i cicli. Quindi, avremo che tutti i cicli ci danno contemporaneamente l'identità se e solo se iteriamo la permutazione un numero di volte k che sia un multiplo comune di tutte le lunghezze $L_1, L_2, \ldots, L_m$. Dalla definizione otteniamo che *l'ordine della permutazione è il più piccolo di questi multipli comuni, ovvero il minimo comune multiplo di* $L_1, L_2, \ldots, L_m$.

Esempio 2.67. Consideriamo la permutazione σ di S_5 data da

$$\sigma(1) = 3,\ \sigma(2) = 5,\ \sigma(3) = 4,\ \sigma(4) = 1,\ \sigma(5) = 2.$$

Si vede subito che la permutazione si decompone in prodotto di cicli disgiunti come $\sigma = (1\ 3\ 4)(2\ 5)$. Poiché abbiamo due cicli di lunghezze due e tre, l'ordine della permutazione sarà il minimo comune multiplo di 2 e 3 ovvero 6. In altre parole, il più piccolo intero positivo k per cui $\sigma^k = id$ è $k = 6$. Per verificarlo direttamente, calcoliamo le potenze successive di σ.

		σ		σ^2		σ^3		σ^4		σ^5		σ^6
1	$\mapsto$	3	$\mapsto$	4	$\mapsto$	1	$\mapsto$	3	$\mapsto$	4	$\mapsto$	1
2	$\mapsto$	5	$\mapsto$	2	$\mapsto$	5	$\mapsto$	2	$\mapsto$	5	$\mapsto$	2
3	$\mapsto$	4	$\mapsto$	1	$\mapsto$	3	$\mapsto$	4	$\mapsto$	1	$\mapsto$	3
4	$\mapsto$	1	$\mapsto$	3	$\mapsto$	4	$\mapsto$	1	$\mapsto$	3	$\mapsto$	4
5	$\mapsto$	2	$\mapsto$	5	$\mapsto$	2	$\mapsto$	5	$\mapsto$	2	$\mapsto$	5

Come si vede, 1, 3 e 4 vengono rimandati in sè già da σ^3. Ma per quella potenza 2 e 5 non vengono mandati in se stessi perché questo accade solo per le potenze pari (cioè multiple di 2). Quindi vengono rimandati in sè contemporaneamente 1, 3, 4 e anche 2, 5 dopo il primo multiplo comune di 2 e 3, ovvero 6.

Come secondo esempio, consideriamo la seguente permutazione di S_{10}

$$\sigma(1) = 5,\ \sigma(2) = 7,\ \sigma(3) = 6,\ \sigma(4) = 10,\ \sigma(5) = 2,$$
$$\sigma(6) = 4,\ \sigma(7) = 1,\ \sigma(8) = 9,\ \sigma(9) = 3,\ \sigma(10) = 8$$

La sua decomposizione in prodotto di cicli disgiunti è $\sigma = (1\ 5\ 2\ 7)(3\ 6\ 4\ 10\ 8\ 9)$. Avendo due cicli di lunghezze rispettivamente 4 e 6, l'ordine della permutazione è il loro minimo comune multiplo, ovvero 12.