Alma Mater Studiorum · Università di Bologna

Scuola di Scienze
Dipartimento di Fisica e Astronomia
Corso di Laurea in Fisica

Correlazione e causalità: le radici fisiche di una distinzione cruciale

Relatore:
Dott. Lorenzo Piroli

Correlatore:
Prof. Jorge Kurchan

Presentata da:
Francesco Contu

Anno Accademico 2023/2024

Abstract

La correlazione e la causalità sono due concetti distinti, ma strettamente legati. Nel secolo scorso, la presa di consapevolezza di tale differenza ha portato la comunità statistica ad abbandonare la nozione di causalità. Negli ultimi decenni si è però riacceso un forte interesse intorno a questo tema, dovendo prendere atto del fatto che l'eliminazione del concetto di causa dal lessico scientifico ci impedisce di comprendere e analizzare in modo efficace molti fenomeni naturali e sociali. Negli ultimi 40 anni sono stati introdotti dei metodi analitici fondati su una particolare formalizzazione della causalità che hanno avuto un successo tale da arrivare a parlare di *rivoluzione causale*. Il presente elaborato ripercorre i principali risultati inerenti tale dibattito e offre una prospettiva fisica sui metodi attuali dell'analisi causale.

Indice

Capitolo 1

Introduzione

Felix qui potuit rerum cognoscere causas.
Virgilio, *Georgiche*, 53 a.C.

La *causalità* è uno dei concetti più intriganti, complessi e dibattuti con cui il pensiero scientifico si sia mai confrontato: l'analisi della sua natura ha rappresentato una costante preoccupazione per la filosofia della natura e poi per la scienza fin dai loro albori. Per molti secoli questa è stata intesa come un attributo fondamentale della realtà. Lo scienziato era colui che studiava *le cause delle cose*, cioè il motivo per cui gli eventi si manifestano in un modo piuttosto che in un altro. Secondo Aristotele, "È evidente che noi siamo alla ricerca dei principi e delle cause, e perciò delle sostanze, essendo convinti che conosciamo ogni cosa solo quando ne conosciamo il perché, cioè la causa" (Aristotele (circa 350 a.C.)). Secondo Platone, "Ogni cambiamento è prodotto da qualche causa; niente accade senza una ragione" (Platone (circa 360 a.C.)).

Le cose iniziarono a cambiare con l'avvento della modernità, ed in particolare con la nascita della fisica newtoniana. Con il tempo, ci si accorse che la nuova fisica riduceva le nozioni di causa ed effetto a relazioni funzionali tra varie grandezze e che, piuttosto che di causalità, era opportuno parlare di *correlazione*. Ernst Mach (1838-1916), fisico e filosofo tedesco noto per il suo empirismo radicale, per primo dichiarò che la causalità fosse una nozione ormai desueta (Mach (1976), Mach (1986)). La scienza moderna, fondata su un ideale di oggettività e chiarezza, richiedeva basi ben più solide per le sue teorie. L'eco di tale posizione fu fortissima, e si declinò nell'eliminativismo (*eliminativism*) causale di Bertrand Russell (1872-1970), logico e filosofo inglese che diede contributi decisivi in moltissimi nodi del dibattito epistemologico dell'epoca. Il suo saggio *On the notion of cause* (Russell (1912)) fu, probabilmente, l'opera di maggiore influenza per le discussioni sulla causalità degli ultimi secoli. In tale sede, egli attaccò con veemenza tale nozione, sostenendo che:

> The law of causality, I believe, like much that passes muster among philosophers, is a relic of a bygone age, surviving, like the monarchy, only because it is erroneously supposed to do no harm. (Russell (1912))

L'influenza di questa prospettiva fu estremamente significativa per tutte le scienze dell'epoca e in particolare per la statistica, nata nella sua formulazione moderna sul finire del XIX secolo. Resosi conto che correlazione e causalità sono concetti differenti, Karl Pearson (1884-1934), uno dei padri fondatori della statistica, cercò di bandire in modo definitivo la nozione di causa dal

lessico scientifico (Pearson (1900)). La comunità statistica lo seguì quasi per intero, rifiutandosi per decenni di utilizzare qualsiasi principio di causalità, considerando la correlazione come l'unica categoria oggettiva da porre a fondamento del discorso scientifico.

Tuttavia, l'analisi di questa scuola di pensiero con il tempo divenne dogmatica, e, progressivamente, ci si rese conto che era insufficiente a cogliere alcune caratteristiche fondamentali della realtà. Poco alla volta si realizzò che la causalità non era semplicemente un modo erroneo di parlare di correlazione, ma che, piuttosto, ne era una declinazione, *quella più interessante*. La scienza *ha bisogno* di discutere in modo efficace tale distinzione, e rigettarla *in quanto tale* è una strada spesso controproducente. Ad esempio, sia fumare che respirare male sono eventi correlati con il cancro ai polmoni, ma è fondamentale distinguere il fatto che il fumo causi il cancro, mentre andare all'ospedale no (Ismael (2023)). Per fondare tale distinzione, ciononostante, era necessario essere in possesso di un adeguato sistema formale, e la statistica, da sola, non poteva fornirlo.

Il primo studioso a proporre una formalizzazione matematica che permettesse di studiare sistemi statistici in termini causali fu il genetista statunitense Sewall Wright (1899-1988), negli anni '20 del secolo scorso. La proposta di Wright, soprattutto in ragione dell'avversione epistemologica dell'epoca, non fu ben accolta, e rimase nell'oblio per molti anni. Negli ultimi decenni, tuttavia, c'è stato un rinnovamento dell'interesse nei confronti della nozione di causa, grazie ai lavori svolti in campi come l'intelligenza artificiale (Judea Pearl (2009b), Spirtes, Glymour e Scheines (2001)), la sociologia (Morgan e Winship (2007)), l'economia (Imbens e Rubin (2015)) e la filosofia (Beebee et al. (2009)).

I metodi di *analisi causale* moderni sono, in effetti, eredi del lavoro di Wright (J. Pearl e Mackenzie (2018)). Le nozioni causali sono oggi espresse attraverso il formalismo delle *reti bayesiane* (*bayesian networks*), erede dei diagrammi causali introdotti dal genetista, e il concetto di *intervento* su un sistema. Assieme, questi ci permettono di *esprimere* delle assunzioni causali, ad esempio il fatto che il fumo causi il cancro mentre il respiro affannato no, anche se entrambi questi eventi possono comparire in presenza di un cancro, e di *distinguere* le associazioni *causali* da quelle *spurie*, cioè quelle in cui è chiaro che l'associazione dipenda dal fatto che due variabili sono entrambe dipendenti da una terza, piuttosto che legate da una dipendenza dinamica. In questo modo, si riesce ad andare oltre il noto mantra "*Correlation is not Causation*" ("La correlazione non è causalità"), e comprendere *quali* associazioni hanno carattere causale. Questi metodi esaminano un insieme statistico di variabili correlate in termini di *relazioni causali* e permettono di interrogarsi su cosa accadrebbe *se* si intervenisse sul sistema modificando il valore di alcune di queste, cioè *simulando* gli esperimenti della di scienze quali la fisica e la chimica. Secondo una simile prospettiva, sostenere che fumare causi il cancro, mentre respirare male no, significa che, se vietassimo alla gente di fumare, la percentuale di popolazione che contrae il cancro diminuirebbe, ma che ciò non accadrebbe se impedissimo alle persone di respirare. Questo modo di intendere la causalità è illuminante, perché riflette quello in cui la concepiamo quotidianamente e permette di introdurre un formalismo matematico in grado di approcciare con successo la questione. Non a caso, le pubblicazioni sul tema sono aumentate a dismisura negli ultimi 40 anni in moltissimi campi di ricerca, tanto che si è iniziato a parlare di *rivoluzione causale* (J. Pearl e Mackenzie (2018)). I testi capitali per quanto riguarda la teoria delle reti bayesiane sono Judea Pearl (1988) e Neapolitan et al. (2004). Le monografiche più complete sul tema dell'analisi causale sono Spirtes, Glymour e Scheines (2001) e Judea Pearl (2009b). Il concetto di *intervento* è stato oggetto di un'analisi filosofica di grande rilevanza in Woodward (2005).

Questi metodi, tuttavia, sono spesso introdotti con fare dottrinario: le assunzioni dell'analisi causale sembrano frequentemente derivare da una scelta filosofica del suo utilizzatore. È inverosimile, tuttavia, che un successo empirico dipenda da una posizione aprioristica e non giustificata. In questo elaborato cercheremo di mostrare *come mai* questi metodi sono ragionevoli, facendo luce su ipotesi spesso proposte in modo assiomatico che, in realtà, sono profondamente radicate nella termodinamica, e, quindi, nella meccanica classica. Non estenderemo le nostre considerazioni al contesto della meccanica quantistica, ma segnaleremo, quando rilevante, che qualcuno sta tentando di farlo, e che il discorso, dopotutto, non cambia in modo radicale.

1.1 In questo elaborato

Nel capitolo 2 (*Correlazione e causalità: da Galton a Pearson a Wright*), mostreremo come la statistica standard non sia in grado di parlarci di rapporti di causalità. In primo luogo, vedremo che un'associazione statistica non può distinguere tra una causa e un effetto, per via della natura simmetrica dei sui indici. Per superare questo scoglio occorrerà introdurre l'asimmetria causale con metodi differenti. In secondo luogo, vedremo che la statistica non può neanche dirci se un'associazione è dovuta a una relazione che effettivamente sussiste tra due variabili o, piuttosto, al fatto che dipendano entrambe da una variabile *confondente* (*confunding*). Questo è il problema delle associazioni spurie il quale, per essere risolto, necessità dell'introduzione del formalismo dei grafi causali. Successivamente, introdurremo il profilo del genetista Sewall Wright (1889-1988), il primo studioso a tentare di reintrodurre la causalità nell'analisi statistica. Wright sviluppò un metodo per studiare le relazioni causali tra variabili statistiche noto come *analisi di percorso* (*path analysis*). Illustreremo tale metodo e ne mostreremo un'applicazione in biologia.

Nel capitolo 3 (*L'asimmetria causale*), discuteremo più approfonditamente la critica di Russell e la assumeremo come corretta, sebbene incompleta. Discuteremo che, nel limite termodinamico, la reversibilità della fisica fondamentale crolla, e che una descrizione *macroscopica* del reale permette di introdurre l'asimmetria necessaria per parlare in modo compiuto di causalità.

Nel capitolo 4 (*Le cause comuni di Reichenbach*) introdurremo il formalismo dei grafi causali, giustificando su base fisica le strutture che Wright disegnò seguendo il suo "buon senso". Una concettualizzazione utile a tale scopo è stata per la prima volta proposta dal filosofo tedesco Hans Reichenbach (1891-1953) nel suo libro *The Direction of Time*, pubblicato postumo nel 1956 (Reichenbach (1991)). In tale opera, il filosofo introdusse il *principio di causa comune* (*Common Cause Principle*), che permette di giustificare in termini di processi dinamici le proprietà di fattorizzazione che osserviamo tra due variabili statistiche causalmente indipenti. La generalizzazione al caso a molte variabili è immediata, e porta all'introduzione dei grafi causali, le reti bayesiane utilizzate dai metodi di analisi causale, e alla *condizione causale markoviana*, una proprietà delle distribuzioni di probabilità associate a un grafo causale di importanza cruciale. Sottolineeremo che, nonostante queste proprietà siano spesso presentate come *assunzioni filosofiche*, in realtà si radicano nell'interpretazione fisica della realtà, in particolare nei suoi aspetti termodinamici (Rovelli (2022a)).

Nel capitolo 5 (*L'inferenza causale*), analizzeremo una problematica emersa nel capitolo precedente, ossia il fatto che le strutture causali che interpretano in modo soddisfacente un sistema statistico sono molteplici. Mostreremo che le correlazioni che osserviamo tra i dati non sono sufficienti per *dedurre* una struttura causale a partire dalle indipendenze statistiche. Questo introduce il tema dell'inferenza causale, ossia il tentativo di identificare, a partire da correlazioni

statiche, la struttura causale che genera tali correlazioni. Inizieremo discutendo casi semplici in cui differenti grafi sono compatibili con distribuzioni di probabilità analoghe e tratteremo brevemente le condizioni necessarie per inferire quale tra questi sia quello corretto. Accenneremo al fatto che questo *modus operandi*, pur sembrando estraneo alla fisica, è in realtà utilizzato in certi ambiti di ricerca, come nello studio delle disuguaglianze di Bell, dove si indaga la possibilità di esistenza di uno schema causale deterministico in grado di spiegare le correlazioni osservate in alcuni fenomeni quantistici.

Il capitolo 6 (*Il do-calculus: dalle correlazioni alla causalità*) sarà l'approdo dell'elaborato, in cui discuteremo i metodi con i quali, in analisi causale, si distinguono le correlazioni spurie da quelle causali. Introdurremo il *do-operator*, l'operatore che *simula* l'effetto di un esperimento su un sistema statistico, e mostreremo come questo ci permetta di esprimere domande causali sul nostro sistema e, talvolta, di ottenere una risposta. Vedremo che, se tutte le variabili rilevanti sono osservate, possiamo sempre ottenere tale risposta dai dati. Successivamente, estenderemo il discorso al caso in cui alcune variabili non sono osservate, introducendo il *do-calculus*, un metodo analitico che ci permette di determinare quando una questione causale è risolvibile - entro certe determinate assunzioni - utilizzando esclusivamente dati osservazionali. Inoltre, evidenzieremo alcuni parallelismi recentemente osservati tra il *do-calculus* e certi metodi della fisica teorica.

In conclusione (capitolo 7) discuteremo l'esito dell'elaborato, sostenendo che l'essere riusciti a radicare l'analisi causale nella termodinamica dimostri come non sia necessario fare assunzioni aprioristiche per riportare *in auge* la nozione di causa.

Capitolo 2

Correlazione e causalità: da Galton e Pearson a Wright

The present paper is an attempt to present a method of measuring the direct influence along each separate path [...] and thus of finding the degree to which variation of a given effect is determined by each particular cause.
S. Wright, *Correlation and causation*, 1921

In questo capitolo mostreremo in che modo i metodi della statistica standard non permettono di approcciare nozioni causali. L'analisi si svolgerà nel quadro teorico della regressione lineare, ma avrà validità generale in quanto non si rivolge alle questioni tecniche dei metodi utilizzati, quanto ai loro presupposti metodologici. In primo luogo vedremo che un indice di correlazione lineare non dà informazioni sulla direzionalità di una relazione, ma si limita a descrivere associazioni dei dati. Successivamente, osserveremo che tale indice non è neanche in grado di differenziare casi in cui è evidente la presenza di un nesso causale e casi in cui l'associazione è chiaramente spuria. Infine, introdurremo il profilo del primo studioso che ha cercato di andare oltre a questo *impasse*, il genetista Sewall Wright.

2.1 La simmetria delle correlazioni

2.1.1 La regressione lineare

Supponiamo di essere in possesso di due popolazioni statistiche di dati $x = \{x_1, x_2, ..., x_n\}$ e $y = \{y_1, y_2, ..., y_n\}$ e di voler comprendere la relazione *quantitativa* che li associa. Ad esempio, potremmo essere interessati a sapere quale è la migliore previsione che possiamo fare per una variabile una volta nota l'altra. Supponiamo di aver collezionato dati sulle ore di studio e sui risultati degli studenti di un'università negli esami di una sessione. Ci aspettiamo che maggiore sia il numero di ore di studio, maggiore sarà il risultato negli esami, ma vorremmo essere in grado di darne la *migliore* previsione dai dati che abbiamo. Non ci aspettiamo certamente di trovare una relazione *perfetta* tra le due variabili: la prestazione di uno studente ad un esame è determinata da una moltitudine di altri fattori, quali la concentrazione, la tranquillità, la predisposizione alla materia di studio e così via. Tuttavia, abbiamo ragione di sospettare che,

tendenzialmente, più ore si studi maggiore sia il profitto. La nostra domanda quindi è: sapendo che lo studente ha studiato in media x ore al giorno in questo semestre, quale è il voto medio y più probabile che otterrà negli esami della sessione?

Se supponiamo che le due variabili siano legate da una relazione lineare, la risposta alla nostra domande è fornita dalla teoria della *regressione lineare*, formalizzata da Karl Pearson e Udny Yule (1871-1951) tra la fine del XIX e l'inizio del XX secolo. Una retta di regressione ci fornisce la migliore previsione *probabilistica* per il valore di y una volta noto quello di x sotto l'ipotesi di dipendenza lineare, spesso indicata come $\mathbb{E}[y|x]$.

Supponiamo di conoscere con precisione assoluta il valore della variabile x. (È semplice generalizzare il calcolo al caso in cui ci siano incertezze anche lungo l'asse x.) Dato che abbiamo una vastissima popolazione di studenti, probabilmente per ogni valore medio di ore di studio ci saranno studenti che hanno ottenuto risultati diversi. Ad esempio, potremmo avere 100 studenti che hanno studiato 6 ore al giorno in media, e che hanno ottenuto risultati differenti all'esame. Possiamo anche supporre che tale popolazione - cioè quella degli studenti che hanno studiato in media 8 ore e 15 minuti - sia una distribuita in modo gaussiano.[1] Di conseguenza, per massimizzare la probabilità $P(y|x)$, dobbiamo massimizzare l'esponente di ognuna di queste gaussiane. Esprimendo $y = Ax + B$, tale massimizzazione ci conduce alle migliori stime per i coefficienti A e B. Si trova che:

$$A = \frac{\sum x^2 \sum y - \sum x \sum xy}{\Delta} \qquad B = \frac{N \sum xy - \sum x \sum y}{\Delta} \tag{2.1}$$

dove le sommatorie si estendono all'intero insieme di misure, e $\Delta = N \sum x^2 - (\sum x)^2$.[2] A e B sono i migliori parametri che descrivono la relazione lineare tra le due popolazioni statistiche, cioè quelli che minimizzano lo scarto tra la previsione del modello e la misura reale. Un esempio di regressione lineare, con dati simulati dall'esempio delle ore di studio e del profitto degli studenti, è riportato in figura 2.1a.

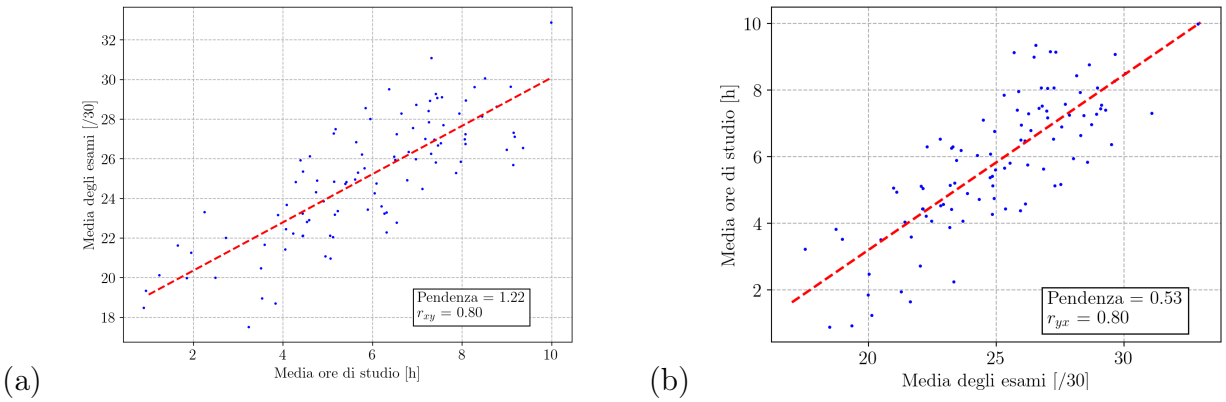
Supponiamo di voler quantificare quanto la nostra ipotesi di relazione lineare si conformi ai dati raccolti. Tale informazione può essere ottenuta dal coefficiente di correlazione lineare r, definito come:

$$r_{xy} = \frac{\mathrm{cov}(x,y)}{\sigma_x \sigma_y} = \frac{1}{N} \frac{\sum_{i=1}^{N} (x_i - \overline{x})(y_i - \overline{y})}{\sigma_x \sigma_y} \tag{2.2}$$

dove $\overline{x} = \sum_{i=1}^{N} x_i / N$ è la media aritmetica di x, e $\sigma_x = \sqrt{\sum_{i=1}^{N} (x_i - \overline{x})^2 / N}$ è la deviazione standard di x, utilizzata come misura della dispersione della variabile (analogo con y). Il termine $\mathrm{cov}(x,y)$, detto *covarianza* di x e y, è quello che ci fornisce indicazioni sull'interdipendenza tra x e y. Se le due variabili fossero indipendenti, entrambe oscillerebbero attorno al valore medio in modo aleatorio, e, sommando su un numero abbastanza grande di campioni, la covarianza dovrebbe essere nulla. Se questa è diversa da zero, significa che c'è una relazione tra x e y tale

[1] Questa è un'ipotesi del tutto naturale in statistica. Poiché le interazioni in questo genere di sistemi sono trattate in modo aleatorio, ogni popolazione deriva da una somma di variabili aleatorie. Quando sommiamo un numero abbastanza grande di tali variabili, in virtù del teorema del limite centrale, dimostrato per la prima volta nel 1810 da Pierre Simon Laplace (1749-1827), sappiamo che la distribuzione della somma sarà una gaussiana, la cui media e varianza dipendono dal numero di variabili che sommiamo e dalle loro caratteristiche. (In realtà, il teorema così come lo si formula oggi fu provato solo nel 1922 da Jarl Waldemar Lindeberg (1876-1932), un matematico e statistico finlandese.)

[2] Una dimostrazione delle precedenti può essere trovata al capitolo 10 di Fornasini et al. (2008).



(a) (b)

Figura 2.1: Relazioni lineari tra il profitto in un esame (espresso in trentesimi) e le ore medie di studio per giorno. Sono mostrate: a) Retta di regressione tra il profitto in un esame (asse y) e le ore di studio medie di uno studente (asse x); b) retta di regressione tra le ore di studio medie di uno studente (asse x) e il profitto in un esame (asse y). Sono riportate anche la pendenza della retta (il parametro A in equazione (2.1)) e il coefficiente di regressione lineare r, definito in equazione (2.2). I dati sono simulati tramite un programma scritto in Python.

per cui, quando una oscilla *sopra* la media, l'altra faccia mediamente lo stesso, e viceversa.[3] Di conseguenza, quando la covarianza non è nulla, possiamo supporre una relazione tra le due variabili. Le caratteristiche di tale relazione sono quantificate dal coefficiente r. In particolare, si dimostra che $r \in [-1, 1]$, e vale 1 per correlazione perfetta positiva (ad un incremento di σ_X corrisponde un incremento di σ_Y), -1 per correlazione perfetta negativa (ad un incremento di σ_X corrisponde un incremento di $-\sigma_Y$) e 0 quando non c'è correlazione (le due variabili sono indipendenti).

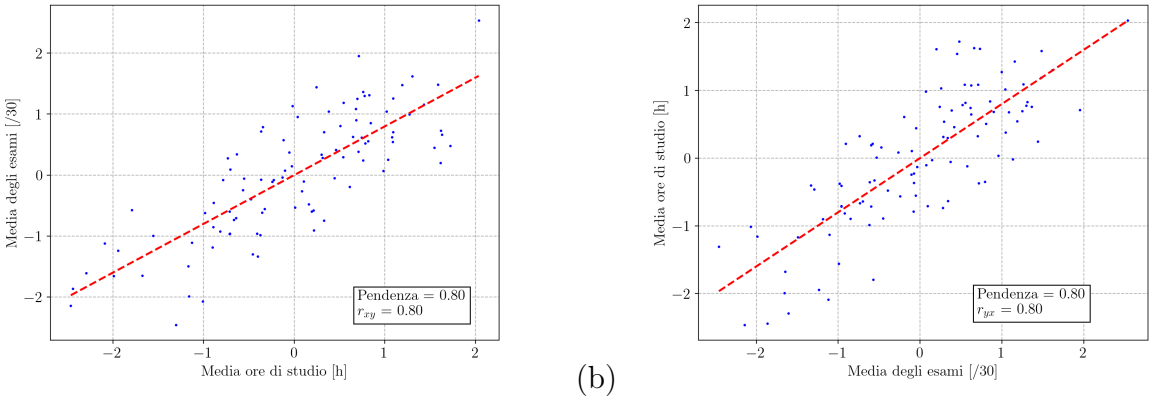
2.1.2 Adirezionalità e variabili normalizzate

Torniamo al nostro studio sul profitto negli esami e la media di ore di studio. Avevamo deciso di studiare il valore di aspettazione del profitto di un esame in funzione delle ore di studio medie durante il semestre perché *sospettiamo* ci sia un nesso di causalità tra le due variabili. Tuttavia, tale informazione è soggettiva: come vedremo in questa sezione, non c'è nulla nel coefficiente di correlazione che introduce un simile nesso causale. Ad esempio, proviamo ad esprimere il numero di ore di studio medie in termini di profitto all'esame, come in figura 2.1b. Anche in questo caso, la relazione è lineare, e il coefficiente di correlazione è lo stesso. Quindi la regressione lineare caratterizza la forza della relazione tra le due variabili, ma non la direzione di influenza causale. Se riscriviamo l'equazione (2.2) in funzione del coefficiente di correlazione, otteniamo:

$$A = \frac{\mathrm{cov}(x, y)}{\sigma_x^2} = r_{xy} \frac{\sigma_y}{\sigma_x} \tag{2.3}$$

Da tale espressione notiamo che la pendenza della retta dipende dalle deviazioni standard delle due variabili. Sarebbe più interessante studiare le popolazioni statistiche utilizzando delle unità di misura adimensionali, tramite le quali l'unità di variazione è proprio la deviazione standard. Inoltre, vorremmo esprimere le nostre relazioni prescindendo dal valore medio che tali variabili

[3] In caso di una relazione lineare a pendenza negativa accade il contrario: quando una variabile oscilla sopra la media, in generale l'alta oscilla *sotto*. Per evitare tali problematiche, spesso ci si interessa al valore di r^2, piuttosto che al semplice r.



(a) (b)

Figura 2.2: Dati di figura 2.1 espressi in variabili normalizzate. Sono mostrate: a) Retta di regressione tra il profitto in un esame (asse y) e le ore di studio medie di uno studente (asse x); b) retta di regressione tra le ore di studio medie di uno studente (asse x) e il profitto in un esame (asse y). Le pendenze delle rette sono le stesse in entrambe le regressioni, e coincidono con il coefficiente di correlazione.

assumono. Delle variabili che ci permettono di avere le proprietà discusse sono le *variabili normalizzate*, definite per una popolazione $x = \{x_i, i = 1, ..., N\}$ come:

$$X_i = \frac{x_i - \overline{x}}{\sigma_x} \tag{2.4}$$

(analogo con y). Queste variabili hanno media nulla e deviazione standard unitaria.[4] È semplice verificare che $r_{xy} = r_{XY}$, quindi utilizzeremo sempre la notazione r_{xy}. Inserendo le variabili normalizzate in equazione (2.1) si trova che la *regressione* di Y su X è data dal coefficiente di correlazione lineare r_{xy}, ossia, sempre assumendo un modello lineare $Y = \beta X$,[5] si dimostra che $\beta = r_{xy}$. Utilizzando tali variabili, dunque, si ottiene la stessa pendenza sia quando si proietta Y in funzione di X sia quando si proietta X in funzione di Y, come riportato in figura 2.2. Formalmente:

$$\mathbb{E}[Y|X] = \mathbb{E}[X|Y] \tag{2.5}$$

Ciò significa che, in un modello lineare $Y = \beta X$ dove β è la regressione di Y su X, è anche vero che $X = \beta Y$.[6] Utilizzando queste variabili, anche l'apparente asimmetria manifestata dalla pendenza delle rette di regressione in figura figura 2.1 scompare. In figura 2.2 riportiamo i grafi di figura 2.1 espressi in variabili normalizzate. In figura 2.2, notiamo che, come dimostrato precedentemente, il coefficiente di correlazione è lo stesso che in figura 2.1, e la pendenza delle rette di regressione non cambia quando invertiamo le coordinate da figura 2.2a a figura 2.2b.

[4] La dimostrazione di tali proprietà è banale: 1) $\overline{X} = \overline{(X - \overline{X})/\sigma_X} = (\overline{X} - \overline{X})/\sigma_X = 0$ e 2) $\sigma_X^2 = \sum_i ((X - \overline{X}) - 0)^2/N\sigma_X^2 = \sigma_X^2/\sigma_X^2 = 1 \Rightarrow \sigma_X = 1$ (analogo con Y).

[5] È semplice mostrare che, utilizzando variabili normalizzate, non c'è alcun termine di offset. Infatti, se si assume che le due variabili siano legate da una relazione lineare $Y = \beta x + \gamma$, poiché $\overline{X} = 0$ (si veda la nota a piè di pagina 4), mediando entrambi i termini della relazione lineare si ottiene $0 = 0 + \gamma$, e quindi $\gamma = 0$.

[6] L'apparente contraddizione sorge dal fatto che, in realtà, con $Y = \beta X$ si intende $\mathbb{E}[Y|X]$.

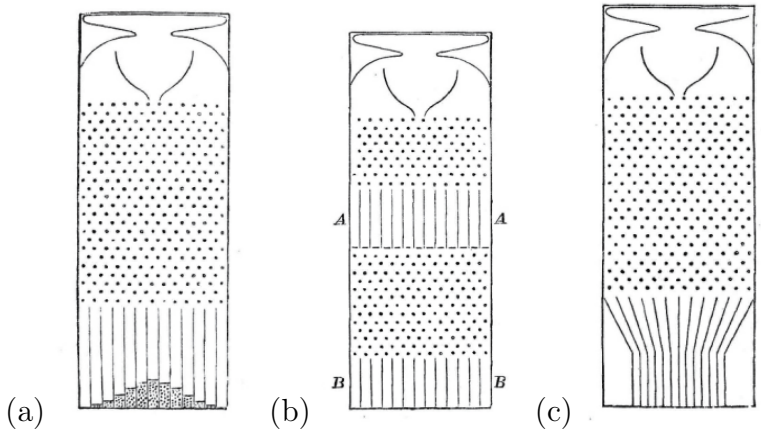


Figura 2.3: Scatola di Galton, utilizzata per mostrare visivamente il processo ereditario tra un padre e suo figlio. Ogni pallina nella sezione in alto rappresenta un padre, ogni pallina nella sezione in basso un figlio. a) Quando le palline vengono fatte cadere dall'alto ed incontrano degli ostacoli che le fanno scostare in modo aleatorio, la distribuzione delle paline al termine del processo è una gaussiana, in virtù del teorema del limite centrale. b) Simulazione del processo di eredità per due generazioni. c) Scivoli utilizzati per simulare la *regressione verso la media*, utilizzata da Galton per spiegare la stabilità della varianza della gaussiana limite.

2.2 Galton e la regressione verso la media

Il primo ad accorgersi che l'analisi statistica non forniva informazioni di carattere causale fu Francis Galton, uno dei padri della neo-nata scienza. È interessante analizzare in che modo lo studioso inglese si rese conto di tale problematica, infatti, in principio, egli era alla ricerca di leggi *causali* che spiegassero i processi di eredità, ma finì con lo scoprire che quello che pensava fosse un processo biologico era in realtà un fenomeno unicamente statistico. In questa sezione discuteremo l'esperienza dello studioso inglese.

A fine '800, una delle più grandi sfide che la scienza stava affrontando era la comprensione dei fenomeni ereditari. Il lavoro di Gregor Mendel (1822-1884), che pose le basi della genetica e della biologia evolutiva, era andato perduto, e sarebbe stato riscoperto solamente circa 35 anni dopo la sua morte, in modo indipendente, da Hugo de Vries (1848-1935), Carl Correns (1864-1935) e Erich von Tschermak-Seysenegg (1871-1962). Ciò che si era compreso, in ogni caso, era che i processi ereditari dipendevano, almeno in parte, da leggi aleatorie.

Nel 1810 Pierre-Simon Laplace (1749-1827) aveva per la prima volta provato il teorema del limite centrale, mostrando che, dato un insieme di N variabili aleatorie indipendenti e identicamente distribuite, la loro somma tende ad una distribuzione gaussiana nel limite $N \to \infty$.[7] Negli anni '80 dello stesso secolo, Galton stava collezionando una grande quantità di misurazioni di alcune caratteristiche antropometriche, come la statura, la lunghezza delle ossa e la larghezza del cranio, e stava cercando di capire tramite quale processo queste venivano tramandate. Lo studioso si accorse che queste erano sempre distribuite in modo gaussiano nella popolazione. Se era vero che l'eredità dipendeva da processi aleatori, allora tale curva doveva emergere dalla somma di tali variabili.

Per mostrare il tutto graficamente, Galton utilizzò uno strumento di sua ideazione, oggi chiamato in suo onore *scatola di Galton* (*Galton board*), riportata in figura 2.3. Nella scatola di

[7] Si veda anche la nota a piè di pagina 1.

Galton, ogni pallina nella sezione superiore rappresenta un padre. Durante il processo ereditario, la pallina viene fatta cadere, e deve attraversare una serie di ostacoli che la faranno spostare a destra o a sinistra, in modo aleatorio (in fisica si direbbe che ogni pallina effettua una *marcia aleatoria*). Quando arriva allo strato inferiore, ogni pallina rappresenta l'espressione della caratteristica ereditaria nel figlio. Poiché la posizione di ogni pallina è determinata dalla somma di molte variabili aleatorie, al termine del processo è una variabile gaussiana. Quando eseguiamo il processo per moltissime palline, osserviamo dunque una curva a campana. Secondo Galton, ciò equivaleva a simulare un processo ereditario da una generazione all'altra. Tuttavia, la scatola non funzionava esattamente come lo studioso inglese si attendeva.

Secondo il teorema del limite centrale, la varianza della gaussiana finale è proporzionale al numero delle variabili, ossia al numero di salti che la pallina compie discendendo nella scatola di Galton. Di conseguenza, se ripetiamo il processo ancora un'altra volta, poiché ogni pallina è soggetta al doppio dei salti, la varianza della seconda generazione dovrebbe essere maggiore. Tuttavia, in natura, si riscontra una forte stabilità sulla distribuzione delle popolazioni, per lo meno tra una generazione e l'altra. Di conseguenza, a questo stadio, la scatola di Galton non rappresenta correttamente il processo evolutivo. Per conservare l'ampiezza della distribuzione della popolazione, Galton introdusse degli *scivoli* (si veda figura 2.3c) i quali, dopo ogni processo ereditario, avvicinano le palline al centro, in modo da riportare la varianza a quella originaria, e li interpretò come il processo *causale* che la natura utilizza per stabilizzare le popolazioni.

L'introduzione degli scivoli era volta motivata da un'osservazione statistica. Infatti, Galton si accorse che, quando un padre è molto più alto della media, tendenzialmente il figlio sarà comunque più alto della media, ma non quanto il padre, e che questo valeva per qualsiasi caratteristica antropometrica che misurava. Galton chiamò tale processo *regressione verso la media*, e oggi sappiamo che è un fenomeno esibito da qualsiasi *trend* statistico, piuttosto che una legge di natura. Infatti, in sezione 2.1.2 abbiamo dimostrato che il coefficiente di correlazione r_{xy} è compreso tra 0 e 1, e che $E(Y|X) = r_{xy}X$, quindi $\mathbb{E}(Y|X) = r_{xy}X < X$, ossia le popolazioni statistiche regrediscono verso la media. Tale concetto, dunque, non è un fenomeno di carattere causale.

Galton si convinse del carattere puramente statistico di tale fenomeno notando che la regressione verso la media si verifica anche quando si cerca di predire l'altezza di un padre conoscendo quella del figlio. Come abbiamo visto, infatti, se $\mathbb{E}[X|Y] = \beta X$, allora anche $\mathbb{E}[Y|X] = \beta Y < Y$. Ovviamente, in alcun modo l'altezza di un figlio può essere causalmente determinante per quella del padre, quindi la regressione verso la media non è un processo causale.

Accortosi di tale problematica, Galton abbandonò la ricerca di un'interpretazione causale, iniziò a dire che le variabili erano "correlate" e si disinteressò dello studio della causalità. Tuttavia, egli non cercò di espellere il concetto di causa dal lessico scientifico. Piuttosto, resosi conto che la correlazione non dava informazioni causali, sospese il suo interesse nei confronti della causalità e si dedicò allo sviluppo dei metodi statistici, che divennero con il tempo sempre più importanti per qualsiasi scienza, sia essa sociale o naturale. La causalità, semplicemente, era qualcosa di differente, a cui la statistica non aveva accesso.

2.3 Pearson e le correlazioni spurie

Nella sezione precedente abbiamo visto che i coefficienti di correlazione non forniscono indicazioni di carattere causale. Piuttosto, questi riflettono dei *pattern* nei dati e ci dicono quanto due variabili sono *interdipendenti*. Il fatto che un nesso causale non è descritto da una cor-

relazione statistica deriva dalla costatazione che i coefficienti di correlazione sono simmetrici all'inversione di causa e effetto. In questa sezione estendiamo tale problematica, introducendo il concetto di correlazione *spuria*.

Nonostante la consapevolezza che la causalità fosse estranea ai metodi statistici, Galton si accorse che, talvolta, tale nozione è necessaria per interpretare delle associazioni tra variabili che sarebbero altrimenti inspiegabili. In più passaggi egli sostenne che, in alcuni casi, la presenza di correlazione tra due variabili è dovuta ad una *causa comune* che le influenza entrambe:

> It is easy to see that correlation [between the sizes of two organs] must be the consequence of the variations of the two organs being partly due to common causes. (Galton (1889))

Quello che sta sostenendo lo studioso inglese è che due variabili potrebbero essere correlate non solo perché una è causa dell'altra, ma anche perché entrambe potrebbero essere dipendenti da una terza.[8] Il fatto che *alcune* associazioni non implichino un nesso causale tra le due variabili ha introdotto nel lessico scientifico il mantra:

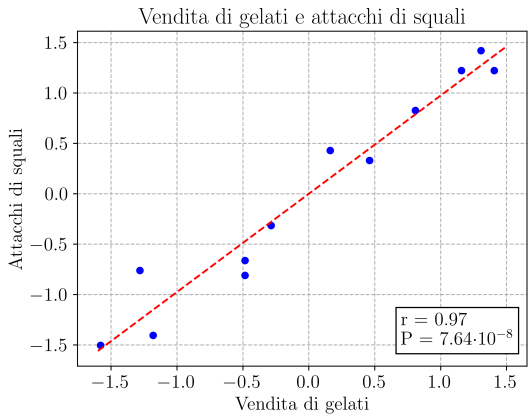
> *Correlation is not causation.* (La correlazione non è causalità.)

Un esempio classico della letteratura in cui un'associazione statistica non è evidentemente dovuta ad un legame causale è quello relativo alla forte associazione che si osserva tra il consumo pro capite di cioccolato e il numero di premi nobel vinti da una nazione. Chiaramente, non sussiste un nesso di causalità tra le due variabili ma, piuttosto, sono entrambe effetto di uno o più fattori comuni, quali, ad esempio, la ricchezza della nazione e la sua distribuzione nella popolazione. Quando restringiamo l'insieme statistico alle nazioni che hanno vinto molti premi nobel, stiamo considerando le nazioni più ricche, quindi sarà molto più semplice trovare un alto tasso di consumo di cioccolato. Tramite una terza variabile, che chiameremo *confondente* (*confounding*), stiamo modificando il tasso di variazione dell'altra, rendendo correlate le due originali, nonostante queste non siano causalmente legate, introducendo un'associazione spuria.

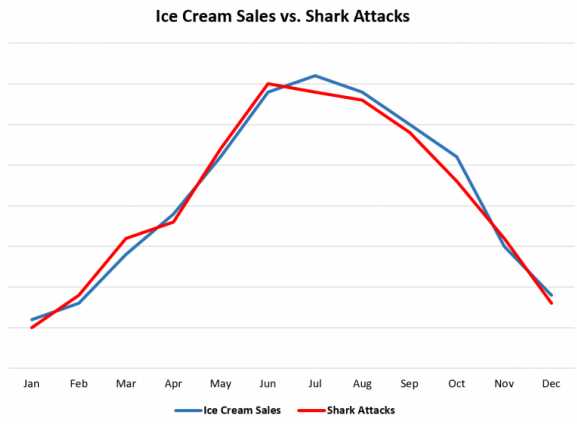
Un altro esempio si può trovare nell'alta associazione che si osserva tra il consumo di gelato e il numero di attacchi di squali. Ancora una volta, non sussiste un nesso di causalità tra le due variabili ma, piuttosto, sono entrambe effetto di uno o più fattori comuni, in questo caso il mese dell'anno. (Banalmente in estate, essendo più caldo, si consumano più gelati e, allo stesso tempo, più persone fanno il bagno, rendendo più probabile l'attacco di uno squalo.) La forte correlazione tra le due variabili è mostrata in figura 2.4. In figura 2.4a si mostra che l'andamento delle due variabili è con buonissima approssimazione lineare ($r \approx 0.97$), cioè che le due sono fortemente associate. Dal grafico di figura 2.4b notiamo che una possibile variabile confondente potrebbe essere, come già discusso, il mese dell'anno, infatti i due insiemi di dati variano assieme al variare dei mesi, ed entrambi raggiungono picchi in quelli più caldi.

Come possiamo capire se l'associazione tra la vendita dei gelati e il numero di attacchi di squali è dovuta ad una relazione causale o, come sospettiamo, semplicemente dal fatto che sono entrambe effetto di una stessa causa? Il modo più semplice è utilizzare le probabilità condizionate. Se ci fosse un nesso causale tra le due variabili, questo dovrebbe valere in ogni mese dell'anno. Quindi, se guardassimo ai dati di un mese isolato, dovremmo avere la stessa

[8] Ovviamente, potrebbero anche essere entrambe le cose contemporaneamente, o potrebbe esserci un complesso sistema di relazioni causali che danno luogo ad una dipendenza statistica. Torneremo con precisione su questo punto nel prossimo capitolo.



(a)



(b)

Figura 2.4: Correlazione tra attacchi di squali e consumo di gelati. I dati sono presentati in variabili normalizzate. Sono mostrati: a) correlazione lineare tra le due variabili (i dati sono stati estratti graficamente dalla fig. 2.4b); b) il periodo dell'anno (in mesi) sull'asse delle ascisse e il numero di attacchi di squali (in rosso) e il numero di gelati venduti (in blu) sull'asse delle ordinate (immagine tratta da https://www.statology.org/correlation-does-not-imply-causation-examples/).

gennaio	febbraio	marzo	aprile	maggio	giugno
$r \approx 0.03$	$r \approx -0.08$	$r \approx 0.09$	$r \approx 0.05$	$r \approx 0.05$	$r \approx -0,02$
luglio	agosto	settembre	ottobre	novembre	dicembre
$r \approx 0.04$	$r \approx 0.14$	$r \approx 0.05$	$r \approx 0.01$	$r \approx 0.01$	$r \approx 0.02$

Tabella 2.1: Indici di correlazione calcolati per i dati di figura 2.5a in funzione del mese dell'anno.

relazione lineare. Se ciò non accade, e si trova che le variabili sono indipendenti sotto tale condizionamento, l'associazione era spuria.

Per visualizzarlo, supponiamo che lo studio sia stato fatto su 1000 spiagge diverse. Quindi, per ogni mese, abbiamo 1000 dati di vendita di gelato e di attacchi di squali. Simuliamo questi dati nel grafico di figura 2.5a. Nelle figure 2.5b, 2.5c e 2.5d mostriamo i dati condizionando sui mesi, rispettivamente, di febbraio, giugno e ottobre. Notiamo che, in questo caso, i coefficienti di correlazione sono nulli con buonissima approssimazione, suggerendo che la nostra analisi sia corretta, e che il mese dell'anno sia una variabile confondente, che introduce associazione spuria tra la vendita dei gelati e l'attacco degli squali. In tabella 2.1 mostriamo gli indici di correlazione calcolati per ogni mese dell'anno, e notiamo che tutti quanti sono prossimi a 0.

In questo caso, la variabile confondente assumeva valori discreti, ma il discorso è del tutto analogo nel caso continuo. Se possediamo i dati su tutte e 3 le variabili, la teoria della regressione lineare ci permette di calcolare il coefficiente di regressione di una variabile su un'altra quando la terza è costante. Se x, y e z sono tre popolazioni statistiche, si dimostra che il coefficiente di regressione di x su y quando z è costante vale:

$$ {}_z r_{xy} = \frac{r_{xy} - r_{xz} r_{yz}}{\sqrt{\left(1 - r_{xz}^2\right)\left(1 - r_{yz}^2\right)}} \tag{2.6} $$

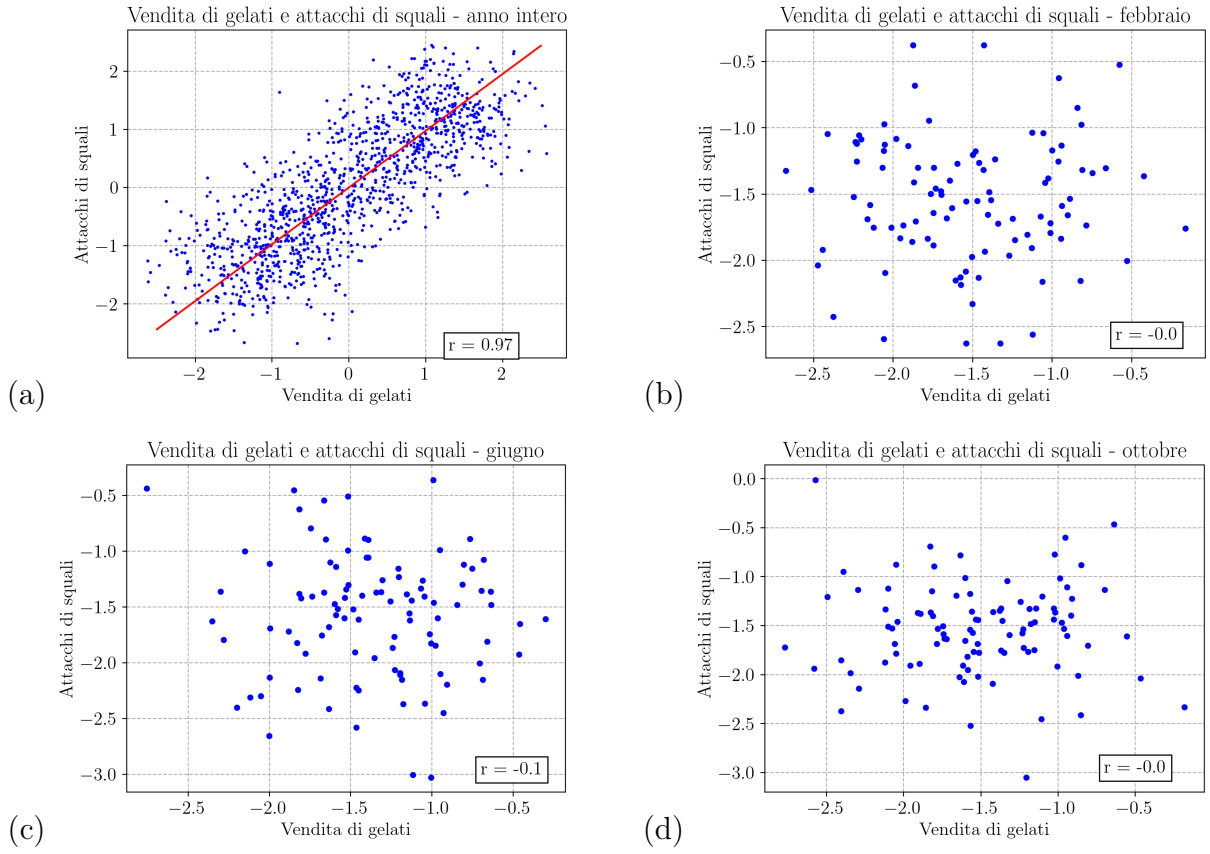


Figura 2.5: Correlazione tra attacchi di squali e consumo di gelati, con dati (simulati) provenienti da 1000 spiagge diverse per ogni mese dell'anno. I dati sono presentati in variabili normalizzate. Sono mostrati: a) correlazione lineare tra le due variabili analizzando tutti i mesi dell'anno; b) correlazione lineare tra le due variabili nel mese di febbraio; c) correlazione lineare tra le due variabili nel mese di giugno; d) correlazione lineare tra le due variabili nel mese di ottobre.

dove con ${}_zr_{xy}$ si intende il coefficiente di correlazione tra x e y quando z è costante.

Quando il sistema di variabili ne contiene una confondente, dunque, per ottenere la correlazione tra le due di interesse occorre condizionare sulla terza. Tuttavia, definire cosa sia una variabile confondente non fu affatto semplice per la comunità statistica. Pearson e il suo allievo Yule cercarono per primi di interpretare tale concetto in termini esclusivamente statistici, e tale tentativo continuò fino alla fine del secolo scorso, ma cadde ogni volta in definizioni insoddisfacenti, o, meglio ancora, sbagliate (J. Pearl e Mackenzie (2018)). Ad esempio, potremmo cercare di definire una variabile confondente come una variabile correlata con entrambe le altre. In effetti, nel caso degli attacchi degli squali e della vendita dei gelati, il mese dell'anno è correlato con entrambe. Tuttavia, ciò non è sufficiente. Ad esempio, torniamo al caso analizzato precedentemente sul numero di ore di studio degli studenti e sul loro profitto negli esami. Supponiamo che l'unico modo in cui studiare di più renda migliore il voto all'esame è il numero di esercizi che si svolgono. Dunque, se introduciamo la variabile "numero di esercizi svolti", questa sarà correlata sia con il numero di ore di studio che con il risultato all'esame, rientrando nella definizione statistica di variabile confondente. In questo caso, condizionare su tale variabile significherebbe guardare la correlazione tra ore di studio medie e risultato all'esame analizzando solo studenti che hanno svolto lo stesso numero di esercizi. Di conseguenza, probabilmente

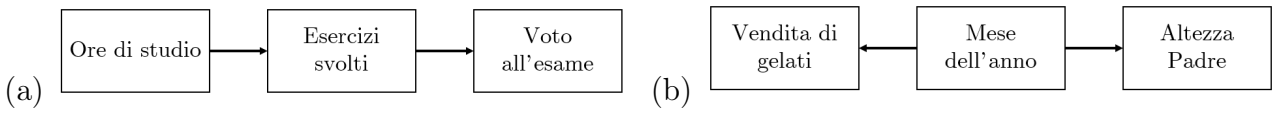


Figura 2.6: Grafi causali rappresentanti i sistemi descritti in questa sezione: a) il numero di ore di studio medio causa il numero di esercizi svolti che, a sua volta, causa il voto ottenuto all'esame; b) il mese dell'anno causa sia il numero di gelati venduti che il numero di attacchi di squali. Nel caso di figura 2.6a la variabile centrale non è confondente, mentre nel caso di figura 2.6a sì.

troveremmo una correlazione molto nulla tra le ore di studio e il risultato agli esami perché, se guardiamo a studenti che hanno svolto lo stesso numero di esercizi, stiamo restringendo la popolazione statistica a quelli che hanno studiato approssimativamente lo stesso numero di ore, e che quindi hanno avuto all'incirca lo stesso voto finale. Possiamo davvero concludere, come nel caso precedente, che l'associazione tra le due variabili era *spuria*? Ovviamente, ciò sarebbe assurdo.

Il punto è che, per parlare di variabile confondente, occorre avere in mente uno schema causale. Una volta introdotta la nozione di causalità, si può far luce su *pattern* statistici che sarebbero altrimenti inspiegabili sulla base dei semplici dati, come nel caso della correlazione tra attacchi di squali e consumo di gelati. Tale intuizione, tuttavia, aveva bisogno di essere formalizzata per divenire utile, e, all'inizio del '900, la comunità scientifica non possedeva una semantica adatta.[9]

Un modo funzionale per notare la differenza tra le situazioni descritte precedente è quello di utilizzare dei *gradi causali*, strutture che rappresentano le relazioni causali tra le variabili del nostro sistema statistico. I grafi associati alle situazioni descritte negli esempi precedenti sono presentati in figura 2.6. Nel grafo di figura 2.6b la variabile "mese dell'anno" è *causa* sia del numero di gelati venduti sia del numero di attacchi di squali, ma non c'è un nesso causale tra queste due. Di conseguenza, guardando ai dati provenienti dallo stesso mese dell'anno, le due variabili diventano indipendenti. Al contrario, nel grafo di figura 2.6a, il numero di ore di studio medio causa il numero di esercizi svolti che, a sua volta, causa il voto ottenuto all'esame. Quindi, se guardiamo solo agli studenti che hanno svolto lo stesso numero di esercizi, impediamo alla variabile "ore di studio medie" di avere effetto causale sull'esito finale dell'esame, e le due variabili, anche se dipendenti causalmente, sembrano indipendenti.

Il primo ad introdurre rigorosamente grafi di questo tipo fu Sewall Wright, un genetista statunitense, noto soprattutto per i suoi lavori in teoria dell'evoluzione e per l'ideazione dell'*analisi di percorso* (*path analysis*), su cui ci soffermeremo nella prossima sezione.

2.4 Wright e i diagrammi causali

In questa sezione discuteremo l'*analisi di percorso*, ideata da Sewall Wright negli anni '20 del secolo scorso. Tale metodo ha fatto da apripista per tutte le linee di ricerca sull'analisi causale.

[9] In realtà, il termine *causa* non era stato del tutto eliminato dal lessico statistico. L'unico caso in cui anche gli eliminativisti più convinti accettavano di parlare di causalità era quello in cui veniva effettuato uno studio randomizzato controllato, detto RCT (*randomized controlled trial*). Possiamo pensare ad un RCT come un modo per eliminare i *bias* confondenti in un'analisi statistica. Poiché, una volta introdotti i metodi della nuova analisi causale, gli RCT non sono che un caso particolare, non li discuteremo in questa sede.

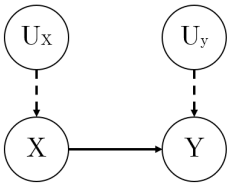


Figura 2.7: Schema causale che rappresenta l'influenza di una variabile X su un'altra variabile Y.

Il merito di Wright è stato quello di fornire alla scienza un lessico - seppure embrionale - per *esprimere* le assunzioni causali. Infatti, come discusso nelle sezioni precedenti, la statistica da sola non è in grado di fornirci strumenti per introdurre delle relazioni di causalità.

I metodi di Wright sono stati ripresi in moltissime altre discipline, rielaborati e potenziati dalla comunità scientifica negli ultimi 50 anni. Uno degli studiosi più importanti e influenti che si è occupato di analisi causale è Judea Pearl, ricercatore nel campo dell'intelligenza artificiale noto per lo sviluppo delle reti bayesiane, l'introduzione dell'approccio probabilistico all'intelligenza artificiale e, soprattutto, per aver sviluppato una teoria dell'analisi causale. Il problema da cui sorge la nostra discussione è il seguente:

> How can one express mathematically the common understanding that symptoms do not cause diseases? (Judea Pearl (2009a))

Se ci limitiamo al caso lineare, potremmo pensare di scrivere un'equazione quale:

$$Y = \beta X + U_Y \tag{2.7}$$

dove Y rappresenta il sintomo, X la malattia, β l'influenza causale del sintomo sulla malattia e U_Y tutte le variabili non osservate che modificano il valore di Y quando X è tenuto costante. (Si sono utilizzate le lettere maiuscole per indicare l'utilizzo delle variabili normalizzate.) Chiameremo *rumore* le variabili come U_Y. Nel capitolo 4 vedremo che queste emergono in modo naturale quando il discorso si cala in fisica. Per adesso, supponiamo che U_Y sia una somma di effetti casuali, e che quindi sia approssimativamente gaussiana, con media nulla. Ciò è ragionevole in virtù del teorema del limite centrale. Chiameremo *variabili osservate* le variabili come X e Y, delle quali possediamo un insieme statistico di dati, e quindi una distribuzione di probabilità.

In questo caso, poiché $\overline{U_Y} = 0$, $\mathbb{E}(Y|X) = r_{xy}X$. Ma è anche vero che $\mathbb{E}(X|Y) = r_{xy}Y$, quindi la semplice equazione non è sufficiente ai nostri scopi, perché non ci permette di distinguere tra causa e effetto. Per far fronte a questa problematica, Wright accostò all'equazione un diagramma, come quello di figura 2.7, dove si è aggiunta anche una variabile esogena U_X relativa a X. Un tale schema causale si associa alla serie di relazioni:

$$X = U_X \qquad Y = \beta X + U_Y \tag{2.8}$$

Wright utilizzò questo formalismo per quantificare l'effetto di una causa. I diagrammi sono necessari perché, come vedremo, le correlazioni osservate si esprimono in termini di percorsi che congiungono due variabili in un grafico. Essi esprimono la considerazione *qualitativa* che il sintomo non causi la malattia. L'assunzione di causalità è codificata graficamente con l'assenza di una freccia da Y a X. Algebricamente, risiede nel fatto che Y sia funzione di X, ma non viceversa. Come abbiamo visto, tale informazione non può dedursi dai dati, perché le distribuzioni di probabilità sono invarianti per inversione delle variabili. Nel capitolo 5 discuteremo

più approfonditamente la problematica dell'introduzione di un grafo causale per un sistema di variabili correlate. Per il momento limitiamoci a dire che derivano da considerazioni qualitative dello scienziato sul suo sistema.[10] Il parametro β, in questo caso, è detto *coefficiente di percorso* (*path coefficient*), e quantifica l'effetto causale di X su Y: una variazione di X di una deviazione standard causerà una variazione di β deviazioni standard di Y, a prescindere dai valori assunti dalle altre variabili del modello e dalla sorgente della variazione di X (cioè, non fa differenza se il cambiamento in X sia imposto dall'esterno o dipenda da una fluttuazione di U_X). Ovviamente, non vale il contrario: imporre una variazione su Y non cambierà il valore di X. Definiremo in modo rigoroso i coefficienti di percorso più avanti in questa sezione, e vedremo che tale asimmetria dipende dalla loro definizione. Assumeremo scorrelate le variabili esogene U_X e U_Y. Nel capitolo 4 motiveremo formalmente tale assunzione. Nel nostro caso, calcolare β è estremamente semplice: basta studiare la correlazione tra X e Y. Tuttavia, gli schemi causali possono essere molto complessi, e non sempre è così semplice calcolare l'influenza diretta di una variabile sull'altra. I metodi di Wright permettono di generalizzare a schemi causali molto più complessi.

Un profilo biografico, storico ed epistemologico dell'opera di Wright e dell'influenza che ebbe nella ricerca dell'epoca può essere trovato in Provine (1989) - il capitolo 5 è dedicato esplicitamente all'analisi di percorso -, una contestualizzazione nel quadro dei metodi contemporanei di analisi causale si trova in J. Pearl e Mackenzie (2018) (capitoli 2 e 4).

2.4.1 Il metodo dei coefficienti di percorso

In questa sottosezione introduciamo in modo formale il metodo dei coefficienti di percorso. Per comprendere gli obbiettivi di Wright, è opportuno utilizzare le sue parole:

> The ideal method of science is the study of the direct influence of one condition on another in experiments in which all other possible causes of variation are eliminated. Unfortunately, causes of variation often seem to be beyond control. In the biological sciences, especially, one often has to deal with a group of characteristics or conditions which are correlated because of a complex of interacting, uncontrollable, and often obscure causes. The degree of correlation between two variables can be calculated by well-known methods, but when it is found it gives merely the resultant of all connecting paths of influence. The present paper is an attempt to present a method of measuring the direct influence along each separate path in such a system and thus of finding the degree to which variation of a given effect is determined by each particular cause. The method depends on the combination of knowledge of the degrees of correlation among the variables in a system with such knowledge as may be possessed of the causal relations. (Wright (1921))

Quindi, Wright vuole poter quantificare l'effetto causale di una variabile su un'altra in modo *isolato*. Non vuole semplicemente guardare a ciò che risulta dal complesso sistema di interazioni, ma intende quantificare quale è il peso di ognuna di queste variabili nel processo di generazione dei dati.

[10] Tali considerazioni sono spesso quasi evidenti. Nel caso che analizzeremo nella prossima sezione, ad esempio, Wright studia il peso alla nascita dei porcellini d'india, e - in modo del tutto naturale - assume che questo sia causato dal tasso di crescita prenatale e dal periodo di gestazione.

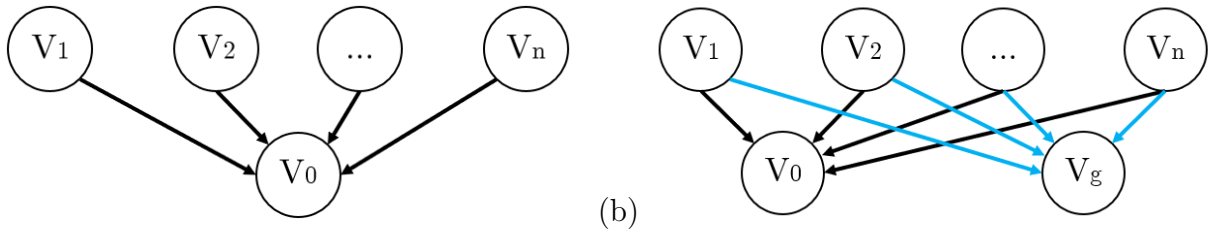


Figura 2.8: Dipendenza causale di a) una variabile V_0 e b) due variabili V_0 e V_g da un insieme di altre variabili $\{V_i, i = 1, ..., n\}$.

Consideriamo un insieme di variabili casuali (estratte da popolazioni statistiche). Assumeremo che tutte le popolazioni hanno uno stesso numero di campioni.[11] Vogliamo disporre queste variabili in uno *schema causale* e studiare l'influenza di una variabile su un'altra quando tutte le altre possibili cause sono fissate. Ogni deviazione di una variabile dal suo valore medio sarà dovuta a variazioni delle variabili da cui dipende direttamente. Se tali relazioni sono lineari, scriveremo che, per la variazione di V_0 in figura 2.8a:

$$(V_0 - \overline{V_0}) = c_1(V_0 - \overline{V_0}) + ... + c_n(V_n - \overline{V_n}) \tag{2.9}$$

dove c_n sono dei coefficienti che rappresentano la forza del legame causale. Utilizzando le variabili normalizzate $X_i = (V_i - \overline{V_i})/\sigma_i$, dove σ_i è la deviazione standard della variabile V_i, si ottiene:

$$\frac{(V_0 - \overline{V_0})}{\sigma_0} = \sum_{i=1}^{n} \left(c_i \frac{\sigma_i}{\sigma_0} \right) \frac{(V_i - \overline{V_i})}{\sigma_i} \tag{2.10}$$

Introducendo i *coefficienti di percorso* (*path coefficients*) P_i:

$$P_i = c_i \frac{\sigma_i}{\sigma_0} \tag{2.11}$$

otteniamo che:

$$X_0 = P_1 X_1 + ... + P_n X_n \tag{2.12}$$

Ognuno di questi coefficienti rappresenta la frazione di deviazione standard di X_0 per la quale la variabile di riferimento è responsabile. Usando la notazione di Wright (1934), cerchiamo di esprimere le correlazioni tra le variabili X_0 e X_g in un grafo come quello di figura 2.8b. Le loro deviazioni dalla media sono sviluppabili sulle loro cause come in equazione (2.12):

$$r_{0g} = \frac{1}{N} \sum_{i=1}^{N} X_g(P_1 X_1 + ... + P_n X_n) = P_1 r_{g1} + ... + P_n r_{gn} = \sum_{i=1}^{n} P_i r_{gi} \tag{2.13}$$

I termini di correlazione residui r_{gi} possono essere analizzati utilizzando lo stesso metodo. Di conseguenza, generalizzando, si ottiene che il coefficiente di correlazione tra due variabili è dato dalla somma di tutti i possibili coefficienti di percorso che le legano nel diagramma causale.[12]

[11] Ovviamente, l'assunzione non ha ripercussioni sulla validità del metodo poiché, generalmente, considereremo N molto grandi, ossia le distribuzioni limite delle variabili. In questo caso è molto semplice eliminare le complicazioni dovute ad un numero finito di campioni.

[12] Nei suoi scritti, in realtà, Wright tiene aperta la possibilità che ci siano dei termini di correlazione tra variabili che non possono essere spiegati in termini di diagrammi di percorso. Nella sua analisi vengono trattati anche questi casi, introducendo delle frecce bidirezionali. Estendere il metodo anche a questa casistica non è complicato, ma lo faremo direttamente nella formulazione moderna dell'analisi causale, nel capitolo 6.

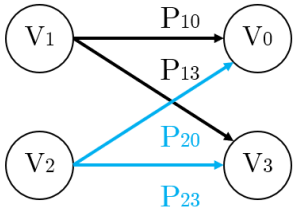


Figura 2.9: Diagramma causale per l'applicazione del metodo dei coefficienti di percorso.

Quindi, in generale:

$$r_{xy} = \sum_{\substack{\text{possibili percorsi che} \\ \text{congiungono } X \text{ e } Y}} \left(\prod_{\substack{\text{coefficienti di percorso} \\ \text{lungo tale cammino}}} \right) \tag{2.14}$$

Vediamo un esempio concreto per comprendere meglio si applica il metodo dei coefficienti di percorso. Consideriamo il diagramma causale di figura 2.9. In questo caso, avremo:

$$\begin{aligned} r_{03} = \frac{1}{N} \sum_{i=1}^{N} \left(P_{10}X_1 + P_{20}X_2\right)\left(P_{20}X_2 + P_{23}X_1\right) = \\ = \frac{1}{N} \sum_{i=1}^{N} \left(P_{10}P_{13}X_1^2 + P_{20}P_{23}X_2^2\right) = \\ = P_{10}P_{13} + P_{20}P_{23} \end{aligned} \tag{2.15}$$

dove nel primo passaggio si è utilizzato il fatto che le variabili X_1 e X_2 sono indipendenti (tale assunzione risiede nell'assenza di frecce tra le due variabili nel diagramma) e nel secondo che le variabili normalizzate abbiamo varianza unitaria.

Si possono ottenere delle ulteriori relazioni utilizzando un caso particolare della precedente. Consideriamo la correlazione di una variabile con se stessa, che sarà ovviamente pari a 1. Nel caso di figura 2.9, possiamo scrivere per X_0:

$$\begin{aligned} r_{00} = \frac{1}{N} \sum_{i=1}^{N} \left(P_{10}X_1 + P_{20}X_2\right)\left(P_{10}X_1 + P_{20}X_2\right) = \\ P_{10}^2 + P_{20}^2 = 1 \end{aligned} \tag{2.16}$$

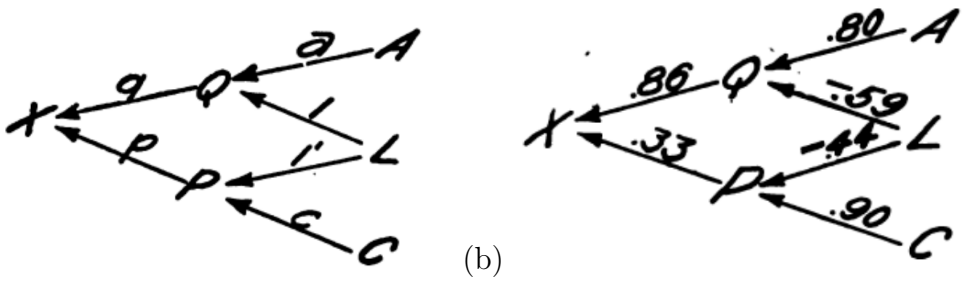
Generalizzando:

$$\sum (\text{coefficienti che puntano a X})^2 = 1 \tag{2.17}$$

Da tali considerazioni possiamo ottenere delle relazioni le quali, noti i coefficienti di correlazione, ci permettono di ottenere il valore dei coefficienti di percorso. In questo modo, utilizzando delle considerazioni qualitative per strutturare le relazioni causali tra le variabili, saremo riusciti a dedurre la forza di un legame causale riferendoci semplicemente a dati osservativi. Nella prossima sottosezione presentiamo un'applicazione immediata di tale procedura.

2.4.2 Il peso alla nascita dei porcellini d'india

L'applicazione più immediata ma non triviale del metodo dei coefficienti di percorso è quella per lo studio di un effetto di due cause correlate, per le quali conosciamo la variabile confondente. Uno schema causale che rappresenta tale situazione è quello riportato in figura 2.10a. Per ana-



(a) (b)

Figura 2.10: Sono riportati: a) diagramma che rappresenta il sistema di relazioni causali tra le variabili rilevanti per determinare il peso dei porcellini d'india (le lettere maiuscole sono le variabili introdotte nel testo, le lettere minuscole i coefficienti di percorso); b) schema di figura 2.10a integrato con i valori dei coefficienti di percorso ottenuti nel testo. Immagini tratte da Wright (1921).

lizzare tale sistema, introduciamo un caso di studio reale. Consideriamo dei porcellini d'india, e studiamo il peso che hanno alla nascita, che chiameremo B. Questo sarà determinato da due fattori: il periodo di gestazione (il tempo trascorso nel ventre della madre), che indicheremo con P, e il tasso di crescita prenatale, che chiameremo Q.[13] Notiamo che non possiamo misurare il tasso di crescita, perché non possiamo pesare i porcellini nel ventre della madre. Potremmo quindi pensare di stimarlo tramite la regressione del peso sul periodo di gestazione. Infatti, $B = QP$, quindi il coefficiente di regressione r_{BP} ci fornisce Q, in variabili normalizzate.

Tuttavia, seguendo Minot (1891), Wright osservò che c'è una variabile confondente tra il tasso di crescita e il periodo di gestazione, ossia la dimensione della cucciolata, che chiameremo L. Infatti, se un cucciolo ha meno fratelli, crescerà più velocemente, e rimarrà di più nel ventre della madre, perché l'ambiente è più favorevole. Di conseguenza, quando guardiamo periodi di gestazione più lunghi, sarà più facile che il tasso di crescita sia maggiore, e viceversa. Se siamo interessati all'influenza di Q su B quando tutte le altre variabili sono costanti il valore ottenuto dal semplice coefficiente di regressione potrebbe essere una sovrastima. Per isolare l'effetto di Q, Wright utilizzò con successo l'analisi di percorso. Iniziò scrivendo uno schema causale, riportato in figura 2.10a. Le variabili sono:

— X è il peso alla nascita del porcellino d'india;

— Q è il tasso di crescita prenatale;

— L è la dimensione della cucciolata;

— P è il periodo di gestazione;

— A sono tutti i fattori ambientali ed ereditari che determinano Q, quindi A è una variabile esogena;

— C sono tutti i fattori che determinano il periodo di gestazione a parte la dimensione della cucciolata, quindi C è una variabile esogena.

[13] Notiamo che in questo caso non ci aspettiamo di avere relazioni lineari: il peso alla nascita sarà il prodotto di questi due fattori. Tuttavia, in Wright (1921) si mostra che, in questo caso, i termini non lineari sono trascurabili. Nei prossimi capitoli discuteremo i diagrammi di percorso in un contesto più generale, dove le relazioni funzionali possono assumere qualsiasi forma, e mostreremo come la maggior parte delle proprietà di questi sistemi risiede nella loro struttura causale, quindi non ci preoccupiamo di queste problematiche.

Variabile	Deviazione standard
X	18.60 g
P	1.91 giorni
L	1.26

Tabella 2.2: Dati dei porcellini d'India e relative deviazioni standard, tratti da Wright (1934).

Il diagramma si interpreta nel seguente modo: il peso alla nascita X di un porcellino d'india è determinato in modo univoco dal tempo che ha passato nel ventre della madre P e dal tasso di crescita Q nel ventre. Entrambi dipendono da una causa comune, la dimensione della cucciolata L, ed hanno rispettive variabili esogene, A e C. Wright raccolse dati sulle variabili X, P e L, riportati in tabella 2.2 e ne calcolò i coefficienti di correlazione:

$$r_{XP} = +0.56 \qquad r_{XL} = -0.66 \qquad r_{PL} = -0.48 \qquad \text{(Wright (1934))} \tag{2.18}$$

Utilizzando l'equazione (2.14) e l'equazione (2.17) sul diagramma di figura 2.10a si ottiene:

$$r_{XP} = p + qll' = +0.5547 \quad r_{XL} = ql + pl' = -0.6578 \quad r_{PL} = l' = -0.4444$$
$$q^2 + p^2 + 2qpll' = 1 \quad a^2 + l^2 = 1 \quad (l')^2 + c^2 = 1$$

che è un sistema di 6 equazioni in 6 incognite.[14] Il sistema ammette le seguenti soluzioni:

$$a = 0.80 \quad l = -0.59 \quad l' = -0.44$$
$$c = 0.90 \quad p = 0.33 \quad q = 0.86$$

In questo modo possiamo ottenere l'influenza del tasso di crescita sul peso dei porcellini quando tutte le altre variabili sono costanti (sotto l'assunzione che lo schema causale sia corretto):

$$r_{X \cdot P} = p_{X \cdot P} \frac{\sigma_X}{\sigma_P} = 0.33 \cdot \frac{18.6}{1.91} = 3.2\text{g/giorno} \tag{2.19}$$

Si ottiene, dunque, un risultato di 3.2 grammi al giorno, molto diverso da quello ottenuto dalle semplici correlazioni, che era pari a 5.66 grammi al giorno.

In conclusione, abbiamo visto che i metodi di Wright permettono di quantificare relazioni sulle quali la statistica sarebbe dovuta rimanere qualitativa.[15] Con il metodo dell'analisi di percorso, dunque, siamo in grado di dedurre la forza di un'influenza causale guardando semplicemente ai coefficienti di correlazione tra le variabili, sotto l'assunzione che il diagramma causale che abbiamo strutturato rappresenti il sistema in modo adeguato.

Tuttavia, questo metodo può sembrare strano agli occhi di un fisico. Ciò è dovuto in parte al contesto assolutamente inusuale per la fisica (i sistemi trattati sono estremamente approssimativi, e la nostra conoscenza delle variabili è imprecisa e non esaustiva), ma soprattutto per il

[14] Il sistema non è lineare, e spesso ciò può complicare le procedure di risoluzione. In questo caso, tuttavia, i coefficienti si ottengono agevolmente.

[15] Per un'evidenza lampante di tale contrasto ci si può riferire a Wright (1918), dove è evidente come Wright riesca ad aggiungere informazioni quantitative alle considerazioni approssimative del prof. con il quale si era formato, William Castle. Per approfondire la questione, rimandiamo ancora una volta al capitolo 5 di Provine (1989).

fatto che il concetto di *causa* non viene spesso utilizzato in fisica con lo stesso significato che assume nelle altre scienze. Per comprendere se (e in che senso) questi metodi possono dare qualcosa alla fisica, oppure se la fisica può essere utile per uno sviluppo ulteriore di questo genere di analisi, occorre cercare di formalizzare in modo migliore alcuni nodi cruciali del metodo di Wright. Dovremo specificando cosa si intende con il termine *causa* e capire meglio in cosa consistono le variabili che stiamo trattando, e motivare l'emergere dei diagrammi causali, discutendo le proprietà di cui godono.

Capitolo 3

L'asimmetria causale

$S = k \ln \Omega$
Epigrafe sulla tomba di Ludwig Boltzmann, Vienna

In questo capitolo dimostreremo come una descrizione *macroscopica* dei fenomeni fisici introduca una direzionalità nella maggior parte dei processi naturali, e discuteremo come ciò possa fornire radici fisiche alla nozione di causalità.

3.1 La critica di Russell

La nozione di causa è un concetto asimmetrico: l'eruzione del Vesuvio ha causato la distruzione di Pompei, non il contrario. Inoltre, essa è orientata temporalmente: una causa accade *prima* del suo effetto. Tuttavia, le equazioni elementari che descrivono la natura non possiedono tale peculiarità. Le leggi della fisica fondamentale sono tutte invarianti per inversione temporale: se un processo può avvenire in una direzione temporale, allora può avvenire anche in quella opposta. Tali leggi, infatti, non sono formulate in termini di cause e effetti, ma si limitano a descrivere le regolarità che osserviamo tra i fenomeni. Di conseguenza, apparentemente, da tali leggi non può emergere alcuna nozione asimmetrica di causa. Tale questione fu sottolineata per la prima volta da Ernst Mach, e ribadita con forza da Bertrand Russell nel suo noto saggio *On the notion of cause* (Russell (1912)). In tale opera, il logico e filosofo inglese sostenne che la causalità era un concetto estraneo al lessico scientifico, e che doveva dunque essere espulsa da tale dominio.

L'analisi di Russell si svolge nella prospettiva del meccanicismo classico, prototipo di una *teoria del tutto*: qualsiasi fenomeno naturale, in ultima istanza, poteva essere ricondotto alle leggi di Newton. È utile seguirlo su tale assunzione perché tale approccio offre un contesto ottimale per discutere i metodi di analisi causale (Judea Pearl (2009b), pagina 26). L'espressione più estrema di tale impostazione filosofica si trova negli scritti di Pierre-Simon Laplace (1749-1827), filosofo, matematico e fisico francese vissuto tra il XVIII e il XIX secolo:

> Nous devons donc envisager l'état présent de l'Univers comme l'effet de son état antérieur et comme la cause de celui qui va suivre. Une intelligence qui, pour un instant donné, connaîtrait toutes les forces dont la nature est animée, et la situation respective des êtres qui la composent, si d'ailleurs elle était assez vaste

> pour soumettre ces données à l'Analyse, embrasserait dans la même formule les mouvements des plus grands corps de l'univers et ceux du plus léger atome : rien ne serait incertain pour elle et l'avenir, comme le passé serait présent à ses yeux. (Laplace (1840))

Le discussioni che faremo, tuttavia, non dipenderanno da tale assunzione: anche le leggi della meccanica quantistica sono reversibili temporalmente, ma il discorso si complica leggermente in quanto dobbiamo rinunciare all'assunzione di determinismo. Nei capitoli 4 e 5 accenneremo alle differenze di cui occorre dar ragione per calare la discussione nel quadro teorico della meccanica quantistica.

Seguendo Russell (1912), diremo che un sistema è *deterministico* se, dato un insieme di dati $e_1, e_2, ..., e_n$ ai tempi $t_1, t_2, ..., t_n$, se $E(t)$ è lo stato del sistema ad un istante di tempo qualsiasi t, questo è determinato in modo univoco dai dati, cioè se:

$$E(t) = f(e_1, t_1, e_2, t_2, ..., e_n, t_n) \tag{3.1}$$

Ad esempio, la fisica classica è deterministica. Un possibile insieme di dati che determina E ad ogni possibile istante di tempo t sono la posizione e il momento delle particelle del sistema ad un istante di tempo qualsiasi. Una volta nota tale informazione, cioè le *condizioni iniziali del sistema*, la sua evoluzione è completamente determinata dalle equazioni di Newton, *in entrambe le direzioni temporali*. Infatti, tali variabili soddisfano un sistema di equazioni differenziali la cui soluzione, per funzioni abbastanza regolari, esiste ed è unica. Ciò è garantito dal teorema di esistenza e unicità di Cauchy-Lipschitz.

Supponiamo, per semplicità, che le forze siano solo funzione della posizione. La seconda legge della dinamica ci dice che:

$$F(x) = m\ddot{x} \tag{3.2}$$

Quindi, per un sistema di n variabili (ad esempio, un sistema di n particelle) avremo un insieme di n equazioni differenziali al secondo ordine che, almeno in linea di principio, ha sempre una soluzione. In formulazione hamiltoniana, gli assiomi di Newton si traducono nelle equazioni di Hamilton:

$$\dot{q_j} = \frac{\partial \mathcal{H}}{\partial p_j} \qquad \dot{p_j} = -\frac{\partial \mathcal{H}}{\partial q_j} \tag{3.3}$$

dove $\mathcal{H}$ è l'hamiltoniana del sistema (la sua energia), q_j sono le coordinate generalizzate e p_j i momenti generalizzati associati. L'assunzione che le forze dipendano solo dalla posizione si traduce nel fatto che $\mathcal{H}$ sia una funzione pari dei momenti.

Introduciamo lo spazio delle fasi Γ, ossia lo spazio vettoriale in cui ogni coordinata rappresenta la posizione o il momento di una particella del sistema. Di conseguenza, se il sistema è composto da n particelle, lo spazio delle fasi sarà lo spazio $6n$-dimensionale:

$$\Gamma = \underbrace{\mathbb{R}^3}_{\substack{\text{posizione della} \\ \text{particella 1}}} \otimes \underbrace{\mathbb{R}^3}_{\substack{\text{momento della} \\ \text{particella 1}}} \otimes \ ... \ \otimes \underbrace{\mathbb{R}^3}_{\substack{\text{momento della} \\ \text{particella } n}} = \mathbb{R}^{\otimes 6n} \tag{3.4}$$

In questo modo, il sistema è rappresentato da un punto X nello spazio delle fasi:

$$X = (\mathbf{r}_1, \mathbf{p}_1, ..., \mathbf{p}_n) \tag{3.5}$$

dove $\mathbf{r}_i = (x^{(i)}, y^{(i)}, z^{(i)})$ e $\mathbf{p}_i = (p_x^{(i)}, p_y^{(i)}, p_z^{(i)})$ sono la posizione e il momento della i-esima particella. Quando il sistema è isolato, cioè l'hamiltoniana non dipende dal tempo, la sua

evoluzione è descritta dalle equazioni di Hamilton. Per quanto discusso, dato un microstato X al tempo t_0, che indicheremo con $X(t_0)$, tali equazioni determinano lo stato del sistema a qualsiasi tempo t. Possiamo esprimere tale concetto tramite la relazione:

$$X(t) = T_{t_0,t} X_0 \tag{3.6}$$

dove $T_{t_0,t}$ è una funzione di evoluzione dinamica generalmente molto complessa. In questo modo, l'evoluzione del sistema può essere descritta come una funzione nello spazio delle fasi.

Consideriamo due microstati del sistema a due tempi differenti $X(t_0)$ e $X(t_0 + \tau)$ con $\tau > 0$. Se invertiamo tutte le velocità del microstato $X(t_0 + \tau)$, otteniamo un nuovo microstato, con tutte le velocità invertite. Tale operazione corrisponde ad invertire la direzione dello scorrere del tempo. Dato che l'hamiltoniana del sistema è una funzione pari dei momenti, le equazioni di Hamilton sono le stesse e, di conseguenza, dopo un tempo τ il sistema tornerà allo stato iniziale, ma con tutte le velocità invertite. Questo significa che se un processo è possibile in una direzione temporale, allora è possibile anche in quella opposta (Lebowitz (2007)).

Tale reversibilità sembrerebbe dunque impedirci di introdurre processi temporalmente asimmetrici nella nostra descrizione della realtà. Tuttavia, in natura, percepiamo un'inevitabile asimmetria temporale. Un gas inizialmente confinato in metà di una stanza e poi lasciato libero di espandersi, occuperà (quasi) sempre la stanza per intero, piuttosto che rimanere confinato in una sola metà. Questo dipende dalla rappresentazione macroscopica del reale che adottiamo in tale contesto, studiata dalla meccanica statistica. Nella prossima sezione vedremo in che modo, con tale approccio, emerga un'asimmetria in quasi tutti i processi fisici dove facciamo questa operazione di astrazione.

3.2 Asimmetrie macroscopiche e freccia causale

Le critiche di Russell sono fondate e convincenti, ma, contrariamente a come aveva pensato Pearson, non concludono il dibattito sulla causalità. Infatti, come abbiamo visto nel capitolo precedente, la nozione di causa è spesso indispensabile per comprendere la realtà. Questa constatazione genera un contrasto teorico? Il fatto che la causalità, nonostante la sua centralità nella vita di tutti i giorni e nelle scienze deboli, non affondi le sue radici nella fisica fondamentale significa semplicemente che non può essere discussa su base fisica? Dobbiamo introdurre un'ontologia causale per sopperire a tale contraddizione?

In realtà, il punto è che l'analisi di Russell è incompleta: è vero che la fisica fondamentale non contiene l'asimmetria tipica della nozione di causa, ma ciò non significa che tale nozione sia priva di senso. Anche il termine *gatto* non compare nelle leggi della fisica fondamentale, ma ciò non ne sminuisce l'importanza per la nostra comprensione del mondo (Rovelli (2023)). Così come il concetto di gatto può essere ridotto all'insieme dei suoi componenti biologici, poi alle sue strutture chimiche e, infine, alla fisica quantistica e subnucleare, anche l'asimmetria causale può essere radicata nella fisica, ma occorre adottare un approccio diverso. In particolare, si deve mostrare come questa, in ultima istanza, si radichi nella termodinamica. Infatti, poiché la nozione di causa è orientata temporalmente, essa non può che fondarsi nell'unica legge della fisica che gode di tale peculiarità, cioè la seconda legge della termodinamica.

La meccanica statistica, sviluppata sul finire del XIX secolo grazie ai contributi decisivi di James Clerk Maxwell (1831-1879), Ludwig Boltzmann (1844-1906) e Josiah Willard Gibbs (1839-1903), permette di approcciare in modo efficace queste problematiche, mostrando come

i processi *irreversibili* con cui abbiamo a che fare tutti i giorni *non* sono in contraddizione con la simmetria delle leggi della fisica fondamentale.

3.2.1 I macrostati e il loro volume nello spazio delle fasi

In primo luogo, occorre definire cosa sia un macrostato. Supponiamo di voler descrivere lo stato di un sistema di N particelle. Ciò significa avere dati sulla posizione e i momenti di tutte le particelle, quindi $6N$ numeri reali. Per calcolarne l'evoluzione dinamica, dovremmo riuscire a risolvere $2N$ equazioni differenziali in modo esatto. Quando le particelle sono molte, ciò diventa pressoché impossibile. Di conseguenza, vorremo isolare delle variabili macroscopiche che ci permettano di analizzare il comportamento *collettivo* del sistema. Ad esempio, potremmo specificare un macrostato M fornendo l'energia del sistema e il numero di particelle in ogni metà della stanza, trascurando le differenze tra le particelle. Ovviamente, un microstato determina il relativo macrostato, ma non vale il contrario: ci possono essere molti microstati compatibili con lo stesso macrostato. Ad esempio, il macrostato "il gas ha temperatura T" è realizzato da moltissime combinazioni delle energie cinetiche delle N particelle: quello che conta è che la loro media sia $3/2 \times k_B T$, dove $k_B = 1.380649 \times 10^{-23}$ è la costante di Boltzmann. Ci aspettiamo anche che esistano macrostati che abbiano associati molti più microstati di altri. Ad esempio, il macrostato "il gas occupa tutta la stanza" è intuitivamente compatibile con molti più microstati di "il gas occupa tutta della stanza". Poiché ogni microstato è associato in modo univoco ad un macrostato, una volta che specifichiamo il tipo di macrostati che vogliamo osservare, lo spazio delle fasi viene diviso in insiemi ad intersezione nulla corrispondenti ognuno ad un macrostato differente.[1] Ad esempio, se il macrostato di riferimento è "metà del gas in un lato della stanza", dovremo dividere lo spazio delle configurazioni in due settori, senza interessarci al momento delle particelle.[2] Di conseguenza, ogni macrostato M occuperà una regione dello spazio delle fasi Γ_M:

$$\Gamma_M = \frac{1}{N! h^{3N}} \int_{\Gamma_M} \prod_{i=1}^{N} \mathrm{d}\mathbf{r}_i \, \mathrm{d}\mathbf{p}_i \tag{3.7}$$

dove il fattore costante $N!h$ serve a normalizzare le probabilità.[3] Intuitivamente, ci si può aspettare che, se il sistema è all'equilibrio ed evolve nel tempo in modo stocastico (e dunque il corrispondente punto nello spazio delle fasi vaga in modo casuale al suo interno), questo si troverà in un determinato macrostato con una probabilità proporzionale al volume che tale macrostato occupa nello spazio delle fasi. Di conseguenza, se prepariamo un sistema in un macrostato con bassa probabilità, cioè se lo isoliamo in una regione dello spazio delle fasi molto ristretta, a seguito del suo vagare caotico molto probabilmente si troverà, dopo un certo intervallo di tempo, in una regione che corrisponde ad un macrostato con volume maggiore. In questo modo, si è rotta la simmetria della fisica microscopica: è vero che un sistema può improvvisamente invertire tutte le sue velocità e tornare indietro nel tempo nel macrostato con

[1] In realtà, tale ripartizione dipende dall'accuratezza con la quale definiamo i macrostati. Se non siamo abbastanza precisi nella descrizione dei macrostati, può darsi che questi si intersechino. Negli esempi che porteremo in questo capitolo, tuttavia, non dovremo affrontare tale problematica.

[2] Una descrizione più precisa potrebbe ad esempio dividere la stanza in K celle, dove K è grande ma comunque tale per cui $K \ll N$. Questo genere di descrizioni conducono alla dinamica dei mezzi continui e all'equazione di Navier-Stockes, alla base dell'idrodinamica.

[3] Il fattore h deriva dalla quantizzazione elementare dello spazio delle fasi che si dà in meccanica quantistica in virtù del principio di indeterminazione di Heisenberg. Il fattore $N!$ è introdotto perché, in generale, le particelle vengono trattate come indistinguibili. Tale assunzione è naturale in meccanica quantistica, ma può sembrare artificiosa in meccanica classica. In realtà, anche in tale contesto teorico è pienamente giustificata, come discusso in Frenkel (2014).

(a)

A	AB
B	0

(b)

A	AB	BD	ACD
B	AC	CD	BCD
C	AD	ABC	ABCD
D	BC	ABD	0

Figura 3.1: Configurazioni microscopiche di: a) 2 particelle distinguibili; b) 4 particelle distinguibili. I macrostati, indicati con colori differenti, sono identificati dal numero di molecole sul lato sinistro della stanza.

volume inferiore, ma ciò non accade perché, nel limite termodinamico, la probabilità di una tale evoluzione è praticamente nulla.[4] Cerchiamo adesso di rendere quantitativa tale analisi, riferendoci all'esempio del gas nella stanza.

Consideriamo un sistema di 2 particelle A e B, e supponiamo che queste si muovano sempre alla stessa velocità. Di conseguenza, possiamo disinteressarci al valore dei loro momenti, cioè limiteremo lo spazio delle fasi a quello delle configurazioni, dove ogni punto rappresenta una possibile disposizione del sistema. Discretizziamo lo spazio delle configurazioni in due parti: la metà a sinistra e quella a destra della stanza. I microstati possibili sono riportati in figura 3.1a. Alcuni di questi corrispondono allo stesso macrostato. In particolare, abbiamo 4 microstati e 3 macrostati possibili. I macrostati sono: "solo una molecola è a sinistra"; "nessuna molecola è a sinistra" e "entrambe le molecole sono a sinistra". Il macrostato in cui solo una molecola è a sinistra è realizzato da più microstati degli altri, quindi occupa un volume maggiore nello spazio delle fasi. Tuttavia, ne occupa solo la metà, e gli altri due macrostati un quarto: la probabilità che il sistema si sposti da un macrostato all'altro è ancora molto alta.

Vediamo cosa accade quando abbiamo 4 particelle. Questa volta abbiamo $2^4 = 16$ microstati, riportati in figura 3.1b, di cui 6 corrispondono al macrostato "metà delle molecole sono a sinistra". Notiamo che, anche questa volta, questo macrostato è sempre il più probabile, ma, ancora, le probabilità degli altri non sono trascurabili. La situazione, tuttavia, cambia radicalmente continuando ad aggiungere molecole. In generale, la probabilità $P(n_S)$ di trovare n_S particelle nella metà sinistra è:

$$P(n_S) = \binom{N}{n_S} p^{n_S} (1-p)^{N-n_S} \tag{3.8}$$

dove N è il numero di molecole totali e p la probabilità che una particella si trovi a sinistra. Cosa accade a tale distribuzione continuando ad aggiungere molecole? Dato che lo spazio delle configurazioni diventa molto più grande, possiamo approssimare a continua la distribuzione di probabilità. È un risultato noto dell'analisi che il limite di una distribuzione binomiale è una distribuzione gaussiana:

$$P(n_S) = \frac{1}{\sqrt{2\pi Npq}} e^{-\frac{1}{2Npq}(n-n^*)^2} \tag{3.9}$$

dove n^* è la media della gaussiana, cioè il punto di massimo della distribuzione di probabilità. Sappiamo che la maggior parte dell'integrale gaussiano è concentrato nell'intervallo $[n^* - \sigma, n^* + \sigma]$, dove σ è la deviazione standard della distribuzione. In tale limite, le

[4] In realtà, il *teorema del ritorno* di Henri Poincaré (1854-1912), dimostra che, se il sistema è isolato, prima o poi questo tornerà nel punto da cui è partito. Tuttavia, i tempi caratteristici di tutti i sistemi con cui abbiamo a che fare sono tali da rendere irrilevante tale problematica.

fluttuazioni dal valore medio sono trascurabili, infatti si dimostra che:

$$\frac{\Delta n}{n^*} \sim \frac{1}{\sqrt{N}} \sim 10^{-13} \tag{3.10}$$

quando $N = 10^{23}$, cioè quando il gas è composto da una mole di molecole. Questo significa che i macrostati molto vicini al macrostato "metà del gas è nella metà a sinistra" occuperanno pressoché tutto lo spazio delle fasi, cioè:

$$\frac{\Gamma_M}{\Gamma} \xrightarrow{N\to\infty} 1 \tag{3.11}$$

3.2.2 L'asimmetria termodinamica e la nozione di causa

Supponiamo di preparare il sistema in modo che tutte le molecole siano in una sola metà della stanza. Il macrostato avrà quindi un volume molto piccolo nello spazio delle fasi, e tale volume, anche se deformato, sarà sempre conservato, in virtù del teorema di Liouville. Una volta lasciato libero di evolvere, il sistema avrà a disposizione uno spazio enorme per farlo. Ad esempio, se consideriamo una mole di gas, il rapporto tra lo spazio totale che rimane da esplorare al sistema e lo spazio occupato inizialmente è molto più grande del rapporto tra la dimensione dell'universo conosciuto e il volume di un protone (Lebowitz (2007)). Di conseguenza, il sistema entrerà progressivamente nella zona dello spazio delle fasi associata al macrostato più probabile e ci rimarrà per moltissimo tempo.

Un modo conveniente per comprendere tale irreversibilità è quello di utilizzare l'entropia di Boltzmann:

$$S = k_B \ln \Gamma_M \tag{3.12}$$

Nel nostro caso, quando confiniamo il sistema in una regione dello spazio delle fasi dove tutte le molecole sono sulla sinistra, il volume associato al macrostato è Γ_i. Quando viene lasciato libero di evolvere, si sposterà in un microstato di volume Γ_f, con $\Gamma_f > \Gamma_i$. Per monotonia del logaritmo:

$$S_f > S_i \tag{3.13}$$

Ossia il sistema evolve verso stati a entropia maggiore. Per questa ragione si dice che l'entropia è associata ad una nozione di disordine: un sistema è tanto più disordinato quanto maggiore è il numero di microstati associati al macrostato che si sta osservando. Nei processi macroscopici, dunque, l'entropia aumenta sempre (o meglio, la probabilità che non lo faccia è trascurabile), perché il sistema evolverà verso macrostati la cui regione dello spazio delle fasi ha un volume maggiore.

Tuttavia, rimane ancora un problema: consideriamo il macrostato del nostro sistema all'istante t_0, con associata entropia S_0. Poiché, generalmente, $S_0 \ll S_{\max}$, il sistema evolverà sicuramente verso uno stato in cui la sua entropia aumenta, per avvicinarsi allo stato di equilibrio. Consideriamo ora il tempo $t < t_0$ e chiediamoci quali sono i valori dell'entropia dei macrostati che, con maggiore probabilità, hanno prodotto tale microstato. In modo analogo, *anche nel passato*, $S(t) > S(t_0)$. Di conseguenza, la legge di massimizzazione dell'entropia di Boltzmann vale in entrambe le direzioni temporali: un sistema tende a massimizzare la sua entropia sia nel futuro che nel passato! Come è possibile, allora, che vediamo tale processo solo nella direzione in cui il tempo aumenta? Per dar ragione di tale fenomeno occorre aggiungere un'ipotesi, detta *ipotesi del passato* (*past hypothesis*) (Feynman (1967)): il nostro universo è nato in una condizione di bassa entropia, e sta evolvendo nel tentativo di massimizzarla. Questa ipotesi ci permette così

di identificare la freccia termodinamica con la freccia cosmologica. In modo analogo, la freccia psicologica, che ci permette di distinguere il passato dal futuro, di ricordare il passato e così via, si può radicare nella termodinamica e, dunque, nella cosmologia (Rovelli (2022b)).

Il discorso è analogo per quanto riguarda la causalità: questa non è altro che l'asimmetria caratteristica dei processi dinamici che osserviamo nella nostra descrizione macroscopica del mondo. Tale asimmetria emerge non appena introduciamo una descrizione approssimativa della realtà. La nostra osservazione quotidiana del reale è *sempre* macroscopica, così come è macroscopica la descrizione della realtà fornita da scienze quali l'economia, la sociologia e così via. Per questa ragione, in tali contesti, parlare di causalità è *utile*: semplicemente perché è un tipo di processo che coinvolge i sistemi che tali scienze analizzano. Al contrario, quando in fisica si isolano processi dal resto dell'ambiente, e se ne studia la dinamica in termini microscopici, l'asimmetria termodinamica scompare, non ha senso parlare di entropia, e la nozione di causa si dissolve.

Capitolo 4

Le cause comuni di Reichenbach

The discovery of a common cause behind seemingly unrelated events reveals the hidden threads that bind the fabric of reality.
H. Reichenbach, *The direction of time*, 1956

Nel capitolo precedente abbiamo esaminato l'origine fisica dell'asimmetria che caratterizza la nostra concezione di causa, permettendoci di relazionare variabili macroscopiche in senso causale. In questo capitolo, utilizzeremo tale nozione per descrivere un sistema statistico in termini dinamici, dove le variabili interagenti sono governate da leggi deterministiche. Per far ciò, sarà cruciale poter replicare le associazioni e le indipendenze osservate nei dati usando variabili macroscopiche e un rumore di fondo non modellizzato. La possibilità di eseguire questa modellizzazione è definita come *compatibilità classica*. Questo approccio fu inizialmente proposto dal filosofo tedesco Hans Reichenbach per dar ragione fisica delle associazioni che si trovano tra variabili indipendenti, e discuteremo in che modo questo si può utilizzare per discutere la fattorizzazione delle probabilità nei grafi causali di alcuni sistemi statistici. Successivamente, generalizzeremo il concetto di compatibilità a grafi arbitrari, e vedremo che questi devono soddisfare ad una proprietà detta *condizione markoviana* per essere compatibili con una descrizione dinamica. Evidenzieremo che tale condizione non è sufficiente a dedurre uno schema causale dalle semplici correlazioni, infatti grafi differenti possono includere le stesse relazioni di indipendenza statistica. Questo aprirà le porte al capitolo successivo, dedicato all'importante tema dell'*inferenza causale*.

4.1 Le dipendenze causali

Consideriamo un sistema statistico composto da due variabili X e Y. Vogliamo esaminare questo sistema nella prospettiva deterministica introdotta nel capitolo precedente, supponendo che queste due variabili contengano tutta l'informazione necessaria per analizzare dinamicamente il sistema, includendo l'irreversibilità tipica dei processi macroscopici.

Consideriamo due variabili X e Y, e supponiamo che siano indipendenti. In una prospettiva dinamica, queste sono rappresentate da macrostati nello spazio delle fasi. Di conseguenza, la probabilità di osservare sia X che Y è semplicemente il prodotto delle due probabilità:

$$P(XY) = P(X)P(Y) \tag{4.1}$$

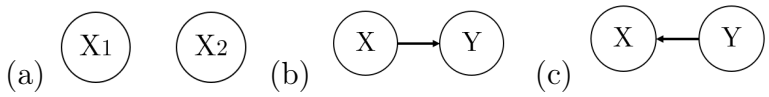


Figura 4.1: Grafi causali associati ad un sistema statistico a due variabili X e Y nel caso in cui: a) sono indipendenti; b) X causa Y; c) Y causa X.

infatti, per ogni istanza della variabile X, dobbiamo considerare tutte le possibili istanze di Y. In generale, dunque:

Due variabili causalmente indipendenti sono anche statisticamente indipendenti.

Quando due variabili sono causalmente indipendenti, possiamo rappresentarle graficamente tramite un grafo composto da due nodi non connessi da alcun arco, come illustrato in figura 4.1a Nella prossima sezione introdurremo formalmente tutta la notazione necessaria per discutere i grafi causali. Per ora, ci limitiamo a utilizzare nozioni che assumiamo note al lettore (nodo, arco, ...).

Supponiamo ora che X sia causa di Y. Tale sistema è rappresentato dal grafo in figura 4.1b. Il nesso causale va inteso nel senso asimmetrico introdotto nel capitolo precedente. In una prospettiva deterministica, una variabile determina univocamente l'altra. Per riprodurre le distribuzioni di probabilità del sistema statistico introduciamo delle variabili non osservate λ che rappresentano il rumore del nostro sistema. In questo modo, la distribuzione di probabilità di Y deve essere totalmente determinata da X assieme a λ, cioè:

$$P(Y) = f(X, \lambda) \tag{4.2}$$

dove f è la funzione di evoluzione del sistema. Questa funzione potrebbe rappresentare qualsiasi legge che lega due fenomeni macroscopici, come l'espansione di un gas discussa nel capitolo precedente, un processo biologico, un fenomeno economico, e così via. Se il nesso causale fosse inverso, ossia se Y causasse X, la discussione sarebbe analoga: il grafo associato sarebbe quello di figura 4.1c e $P(X)$ sarebbe univocamente determinata da Y e da λ:

$$P(X) = f(Y, \lambda) \tag{4.3}$$

Ovviamente, però, il sistema è causalmente differente, a causa della nozione asimmetrica di causa che stiamo adottando.

Supponiamo ora di voler calcolare la probabilità condizionata di Y su X, cioè $P(Y|X)$. In tal caso, poiché X evolve in Y, dobbiamo considerare solo le sue evoluzioni che generano Y. Matematicamente, ciò significa introdurre una δ di Kronecker. Quindi:

$$P(Y|X) = \sum_{\lambda} \delta\left(Y, f(X, \lambda)\right) P(\lambda) \tag{4.4}$$

dove: $\delta(Y, f(X, \lambda)) = 1$ se $Y = f(X, \lambda)$ e $\delta(Y, f(X, \lambda)) = 0$ se $Y \neq f(X, \lambda)$. Tale equazione viene spesso indicata come *dilatazione classica* (Allen et al. (2017)). Generalizzando al caso continuo:

$$P(Y|X) = \int_{\lambda} \mathrm{d}\lambda \delta\left(Y, f(X, \lambda)\right) P(\lambda) \tag{4.5}$$

Per ragioni di chiarezza espositiva, utilizzeremo in seguito sempre il caso discreto. Quando necessario, discuteremo come varia la trattazione nel caso di variabili continue.

Nella prossima sezione vedremo come tali nozioni ci permettono di comprendere fenomeni di correlazione tra variabili macroscopiche. Introducendo il *principio di causa comune* di Reichenbach, saremo in grado di definire in senso causale il concetto di variabile confondente, superando l'*impasse* epistemologico discusso nel capitolo 6.1.

4.2 Il principio di causa comune di Reichenbach

Il *principio di causa comune* è stato introdotto per la prima volta da Reichenbach nel libro *The Direction of Time* (Reichenbach (1991)), pubblicato postumo nel 1956. Questo principio è oggetto di ampie discussioni in numerose branche della filosofia della scienza; in questa sede ci focalizzeremo sugli aspetti rilevanti per i nostri obiettivi.[1] Inoltre, questo viene spesso formulato in modi diversi, il che può causare una certa confusione. In questo contesto, utilizzeremo le definizioni comunemente adottate nella ricerca in fisica (Allen et al. (2017), Cavalcanti e Lal (2014), Pienaar e Brukner (2015)).

Il principio formalizza l'intuizione secondo cui una dipendenza statistica deve avere una spiegazione in termini causali, e può essere scomposto in due asserzioni distinte (Cavalcanti e Lal (2014)). La prima afferma che due variabili, non legate causalmente ma comunque correlate, devono avere una causa comune. La seconda stabilisce che le distribuzioni di probabilità delle due variabili devono fattorizzare quando si condiziona sulla causa comune. La seconda asserzione presenta delle problematiche quando viene applicata al contesto della meccanica quantistica (Pienaar e Brukner (2015), Fritz (2016)). Al contrario, nel quadro della meccanica classica, le due sono equivalenti (Allen et al. (2017)). In questo elaborato ci concentreremo sul caso classico, introducendo le due parti del principio in modo separato e dimostrando la loro identità logica.

Supponiamo che il nostro sistema sia composto da due variabili Y e Z e che esse siano correlate, cioè che $P(YZ) > P(Y)P(Z)$. Per dar ragione di tale correlazione, secondo Reichenbach, deve valere una delle seguenti condizioni:

1. Y è causa di Z (figura 4.2a);
2. Z è causa di Y (figura 4.2b);
3. non ci sono legami causali tra Y e Z, ma entrambe sono effetto di una causa comune X (figura 4.2c);
4. Y è una causa di Z ed entrambe sono effetto di una causa comune X (figura 4.2d);
5. Z è una causa di Y ed entrambe sono effetto di una causa comune X (figura 4.2e);

Tale osservazione costituisce la parte *qualitativa* del principio di Reichenbach.

Tale osservazione si applica, ad esempio, al caso della correlazione tra numero di attacchi di squali e di gelati venduti, discusso nel capitolo 2. In questo caso Y rappresenta la variabile "attacchi degli squali", e Z come la variabile "consumo di gelati". In tal caso, abbiamo assunto che non ci fosse un nesso causale tra le due variabili. Poiché queste sono comunque correlate, il principio di Reichenbach ci dice che ci deve essere una causa comune, cioè che il grafo causale associato a tale sistema statistico è quello di figura 4.2c. Avevamo discusso che una possibile causa comune era il mese dell'anno. Potremmo quindi identificare X con tale variabile. Ovviamente, non è detta che X sia l'*unica* causa comune: potrebbero esserci altri fattori che determinano

[1] Una panoramica storica ed epistemologica del principio è disponibile in Hitchcock e Rédei (2020).

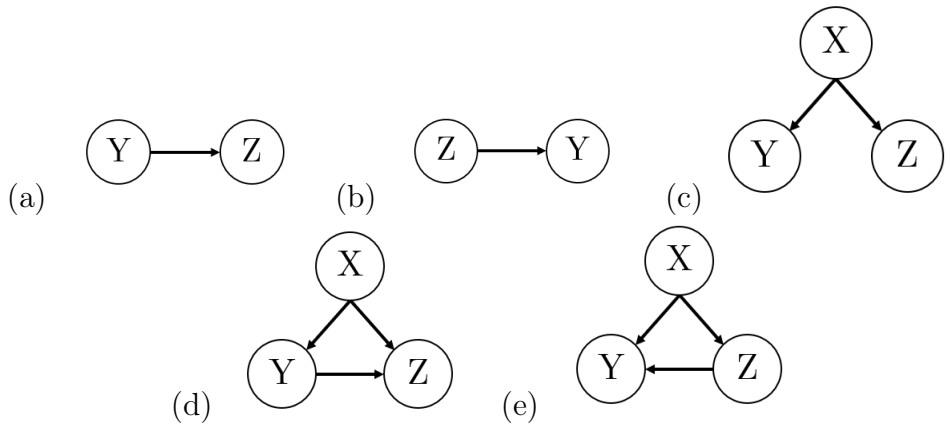


Figura 4.2: Possibili relazioni causali postulate dal principio di Reichenbach quando Y e Z sono due variabili statisticamente correlate.

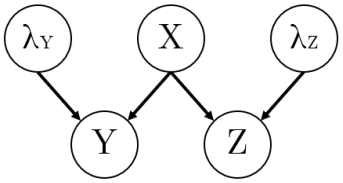


Figura 4.3: Struttura causale di figura 4.2c, aggiungendo le variabili di rumore λ_Y e λ_Z.

sia un aumento di attacchi di squali che di vendite di gelati. Supponiamo però che tali variabili non siano determinanti per il nostro sistema. In effetti, abbiamo visto che condizionare sul mese dell'anno rende le altre due variabili indipendenti con buonissima approssimazione. Ciò significa che questa contiene, nello spazio delle fasi del sistema, *tutte* le cause comuni di Y e Z, o comunque che ne esclude un insieme di misura nulla nello spazio delle fasi. Se così non fosse, infatti, anche condizionando su X, Y e Z rimarrebbero associate. In questi casi diremo che X è *causa comune completa*. La proprietà di fattorizzazione sulla causa comune completa costituisce la parte *quantitativa* del principio di Reichenbach. Formalmente, questa stabilisce che le probabilità di due variabili Y e Z devono fattorizzare quando condizioniamo su X, se questa è causa comune completa:

$$P(Y, Z|X) = P(Y|X)P(Z|X) \tag{4.6}$$

ossia che i due eventi sono indipendenti quando condizioniamo su X. Di conseguenza, applicando l'equazione (4.4):

$$P(YZ|X) = \sum_{\lambda_Y, \lambda_Z} \delta\left(Y, f_Y(\lambda_Y, X)\right) \delta\left(Z, f_Z(\lambda_Z, X)\right) P\left(\lambda_Y, \lambda_Z\right) \tag{4.7}$$

Poiché λ_Y e λ_Z sono indipendenti, $P(\lambda_Y, \lambda_Z) = P(\lambda_Y)P(\lambda_Z)$, quindi:

$$P(YZ|X) = \sum_{\lambda_Y \lambda_Z} \delta\left(Y, f_Y(\lambda_Y, X)\right) \delta\left(Z, f_Z(\lambda_Z, X)\right) P\left(\lambda_Y\right) P\left(\lambda_Z\right) \tag{4.8}$$

Chiameremo *compatibilità classica* la possibilità di tale sviluppo, dove con *compatibilità* si fa riferimento al fatto che le correlazioni osservate nei dati sono spiegabili in termini deterministici, a meno dell'aggiunta del rumore, che serve semplicemente a giustificare il fatto che stiamo trattando distribuzioni di probabilità, piuttosto che variabili fisse. Formalmente:

Definizione 4.2.1 (*Compatibilità classica - caso in 2 variabili*)**.** $P(YZ|X)$ è detta essere *compatibile* con X come causa comune di Y e Z se possiamo trovare:

1. delle variabili λ_Y e λ_Z;
2. delle distribuzioni $P(\lambda_Y)$ e $P(\lambda_Z)$;
3. delle funzioni $f_Y : (\lambda_Y, X) \to Y$ e $f_Z : (\lambda_Z, X) \to Z$

tali per cui queste costituiscono una dilatazione classica di $P(YZ|X)$, ossia:

$$P(YZ|X) = \sum_{\lambda_Y, \lambda_Z} \delta\left(Y, f_Y(\lambda_Y, X)\right)\left(Z, f_Z(\lambda_Z, X)\right) P\left(\lambda_Y\right) P\left(\lambda_Z\right) \tag{4.9}$$

Il grafo di figura 4.3 è la controparte grafica di tale proprietà. La correlazione tra due variabili indipendenti è dunque compatibile con una descrizione classica se può essere spiegata in termini di una causa comune e di un rumore di fondo non modellizzato. In generale, dunque:

$$\text{I termini di rumore sono } \textit{non} \text{ correlati per definizione.} \tag{4.10}$$

Se fattorizziamo i termini dell'equazione (4.8) otteniamo:

$$\begin{aligned} P(YZ|X) = \sum_{\lambda_Y} \delta\left(Y, f_Y(\lambda_Y, X)\right) P\left(\lambda_Y\right) \cdot \sum_{\lambda_Z} \delta\left(Z, f_Z(\lambda_Z, X)\right) P\left(\lambda_Z\right) \\ \Rightarrow P(YZ|X) = P(Y|X)P(Z|X) \end{aligned} \tag{4.11}$$

dove nel seconda passaggio si è utilizzata la dilatazione classica (equazione 4.4). Abbiamo quindi ottenuto la parte quantitativa del principio di Reichenbach assumendo quella qualitativa. Più in generale, abbiamo mostrato che, se X è compatibile con l'essere causa comune di Y e Z (equazione (4.8)), allora le probabilità di Y e Z fattorizzano quando condizioniamo su X. Inoltre, vale anche il contrario: se $P(YZ|X) = P(Y|X)P(Z|X)$, allora X deve essere l'unica causa comune (altrimenti il condizionamento non schermerebbe parte della correlazione), di conseguenza possiamo sviluppare la probabilità congiunta come in equazione (4.8), e quindi il sistema di variabili è *classicamente compatibile*. Possiamo riassumere quanto discusso nel seguente teorema:

Teorema 4.2.1. Data una distribuzione di probabilità $P(YZ|X)$, le seguenti proposizioni sono equivalenti:

1. $P(YZ|X)$ è compatibile con X come causa comune di Y e Z (definizione 4.8);
2. $P(YZ|X) = P(Y|X)P(Z|X)$.

Il teorema giustifica il modo in cui facciamo emergere una struttura causale a partire da una correlazione tra due variabili. in particolare:

— l'implicazione $1 \to 2$ significa che, se X è una causa comune di Y e Z, allora condizionare su X rende Y e Z indipendenti;

— L'implicazione $2 \to 1$ ci permette di dedurre una possibile *spiegazione causale* per una distribuzione osservata partendo da una caratteristica di tale distribuzione. Tuttavia, ciò suggerisce solo una *possibile* spiegazione causale, non che ce ne sia un'*unica* compatibile con la distribuzione. Infatti, come discuteremo nel prossimo capitolo, ci sono anche altre strutture causali che riproducono la stessa proprietà di fattorizzazione e sono compatibili con una descrizione dinamica.

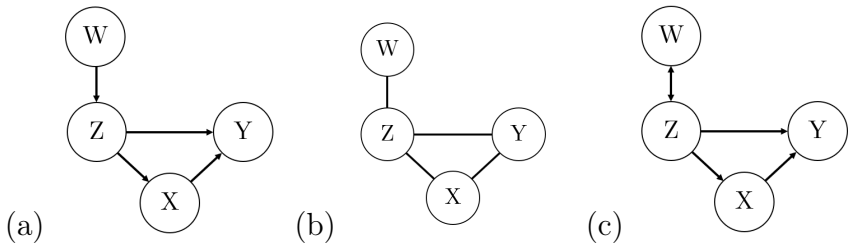


Figura 4.4: Esempi di grafi che rappresentano sistemi con stesse variabili ma archi diversi.

4.3 I grafi causali e la condizione causale markoviana

In questa sezione generalizzeremo le osservazioni della precedente a schemi causali arbitrariamente complessi. Dopo aver introdotto la notazione necessaria, generalizzeremo il concetto di compatibilità classica, e mostreremo come questo conduce alla condizione causale markoviana.

4.3.1 I grafi causali

Un *grafo* consiste in un insieme di *nodi* e un insieme di *archi* che connettono alcune coppie di nodi. I nodi del grafo rappresentano le *variabili* del sistema, mentre gli archi le *relazioni* tra le variabili. Due nodi legati da un arco sono detti *adiacenti*. Se l'arco punta da un nodo ad un altro è rappresentato da una freccia con una sola direzione, e si dice *diretto*. Un esempio di arco diretto è quello tra W e Z in figura 4.4a. Se l'arco è un segmento, si dice *adirezionale*. Un esempio di arco adirezionale è quello tra W e Z in figura 4.4b. Se rimuoviamo tutte le direzioni delle frecce e lasciamo indicati solo i segmenti (cioè se rendiamo tutti gli archi *adirezionali*), otteniamo lo *scheletro* del grafico. Ad esempio, il grafo di figura 4.4b è lo scheletro di quello di figura 4.4a. Se l'arco punta in entrambe le direzioni, è detto *bidirezionale*. Un arco bidirezionale rappresenta la presenza di una variabile che è causa comune di altre due ma che non è specificata nel modello. Un esempio di arco bidirezionale è quello che connette W e Z in figura 4.4c. Se tutti gli archi sono diretti, il grafo si dice *diretto*. Un esempio di grafo diretto è quello di figura 4.4a.

Un *percorso* è una sequenza di archi tali per cui ognuno inizia con il nodo con cui era terminato il precedente e termina con il nodo con cui inizia il successivo. Ogni percorso si può indicare specificando l'insieme delle coppie dei nodi che connette. Ad esempio, due percorsi che legano le variabili W e X in figura 4.4a sono $\{(W,Z),(Z,X)\}$ e $\{(W,Z),(Z,Y),(Y,X)\}$. Un percorso può passare per ogni arco sia nella direzione della freccia che in quella opposta. Se il percorso passa attraversa archi esclusivamente seguendo la loro direzione causale, il percorso si dice *causale*. Ad esempio, nel grafo di figura 4.4a, il percorso $\{(W,Z),(Z,Y)\}$ è causale, ma $\{(W,Z),(Z,Y),(Y,X)\}$ non lo è. Un percorso non causale si dice *anticausale*.

Se due nodi possono essere congiunti con un percorso (qualsiasi) si dicono *connessi*, in caso contrario si dicono *sconnessi*. Un percorso causale che parte da una variabile e finisce con la stessa si dice *loop*. Poiché una variabile non può essere causa di sé stessa, assumeremo che in un grafo causale non ci siano *loop* (Pienaar e Brukner (2015)). Il tal caso, questo si dice *aciclico*. Se il grafo è sia diretto che aciclico, si dice *grafo diretto aciclico* (*directed acyclic graph*). Tali grafi sono noti nella letteratura con l'acronimo DAG. Nel presente elaborato li indicheremo semplicemente come *grafi causali*. Un esempio di grafo causale è quello di figura 4.4a.

Utilizzeremo lettere maiuscole per indicare variabili o insiemi di variabili, e lettere minuscole per indicare le loro istanze. Ad esempio, scriveremo $V = V_1, ..., V_n$ per indicare l'insieme delle n variabili $V_i, i = 1, ..., n$, e $v = v_1, ..., v_n$ per indicare una loro istanziazione. Data una variabile V_i, chiameremo insieme dei *genitori causali*, e lo indicheremo con PA_i (dall'inglese *parents*) l'insieme delle variabili del grafico che hanno un arco che punta verso V_i. Ossia, i genitori di una variabile sono le variabili che, nel grafico, hanno influenza diretta su V_i. Chiameremo *antenati* di V_i tutti i nodi A_i per i quali esiste un percorso diretto che parte da A_i ed arriva a V_i. In modo analogo, indicheremo con CH_i (dall'inglese *children*) tutte le variabili C_i per le quali c'è un arco diretto che punta da V_i a C_i e *discendenti* di V_i tutte le variabili D_i per le quali esiste un percorso causale che parte da V_i ed arriva a D_i. pa_i sarà un'istanza delle variabili PA_i, ch_i sarà un'istanza delle variabili CH_i, e così via.

Le definizioni di questa sezione, che utilizzeremo nel resto dell'elaborato, sono riassunte in tabella 4.1.

4.3.2 La condizione causale markoviana

Consideriamo un sistema di un numero arbitrario di variabili statistiche, delle quali conosciamo le distribuzioni di probabilità. Vogliamo disporre tali variabili in un grafo causale per poter studiare il sistema statistico in senso dinamico. In primo luogo dobbiamo estendere la nozione di compatibilità che abbiamo introdotto per giustificare il principio di Reichenbach ad una struttura causale arbitraria. Come nel caso in 2 variabili, per spiegare in termini di un grafo causale le dipendenze tra le variabili che osserviamo nei dati dobbiamo introdurre delle variabili che rappresentano il rumore, come le λ delle sezioni precedenti. Se siamo in grado di riprodurre il sistema di indipendenze statistiche tramite l'introduzione di queste variabili, allora il sistema può essere interpretato in senso dinamico. Possiamo formalizzare tale intuizione estendendo il concetto di compatibilità 2d della sezione precedente:

Definizione 4.3.1 (Compatibilità con un grafo causale)**.** Dato un modello causale con delle variabili $X_1, ..., X_n$ associate a un grafo causale G, e una distribuzione di probabilità sulle variabili $P(X_1, ..., X_n)$, diremo che P è *compatibile* con il grafo causale G se possiamo trovare un grafo causale G' ottenuto da G aggiungendo ulteriori nodi $\lambda_1, ..., \lambda_n$ in modo che, per ogni nodo i, ogni λ_i ha una sola freccia causale che punta a X_i, e si può trovare una distribuzione $P(\lambda_i)$ e una funzione $f : (\lambda_i, PA_i) \to X_i$ tale per cui:

$$P(X_1, ..., X_n) = \sum_{\lambda_1,...,\lambda_n} \left[\prod_{i=1}^{n} \delta(X_i, f_i(\lambda_i, PA_i)) P(\lambda_i) \right] \tag{4.12}$$

Come nel caso in 2 variabili, la definizione significa semplicemente che la distribuzione di probabilità emerge dalla dipendenza dinamica tra le variabili assieme a termini di rumore tra di loro non correlati. Se un grafo è classicamente compatibile con la distribuzione di probabilità che osserviamo, possiamo riscrivere l'equazione (4.12):

$$P(X_1, ..., X_n) = \prod_{x_i} \sum_{x_i} \delta(X_i, f_i(\lambda_i, PA_i)) P(\lambda_i) \tag{4.13}$$

Quindi, utilizzando la dilatazione classica (equazione (4.4)):

$$P(X_1, ..., X_n) = \prod_{i=1}^{n} P(X_i | PA_i) \tag{4.14}$$

Quando tale proprietà vale, la distribuzione di probabilità si dice *markoviana*. Formalmente:

Definizione 4.3.2 (Distribuzione markoviana per un grafo causale - definizione per *fattorizzazione*)**.** Una distribuzione di probabilità $P(X_1, ..., X_n)$ si dice *markoviana* per un grafo causale G e un insieme di variabili $X_1, ..., X_n$, se e solo se può essere scritta nella forma:

$$P(X_1, ..., X_n) = \prod_{i=1}^{n} P(X_i|PA_i) \tag{4.15}$$

(Ricordiamo che $P(X_i|PA_i)$ può essere dedotta dalla distribuzione osservazionale.)

Vediamo ora come tale definizione proprietà deriva direttamente dalla definizione di causa che stiamo utilizzando. Assumendo che il grafo e la distribuzione di probabilità siano *markoviani*, calcoliamo la probabilità $P(X|Pa(X), ND(X))$, dove ND(X) sono le variabili non discendenti di X. Ci aspettiamo che questa sia uguale a $P(X|Pa(X))$, cioè che gli effetti non determinino la causa. In effetti:

$$\begin{aligned} P(X|PA(X), ND(X)) &= \frac{P(X, PA(X), ND(X))}{P(PA(X), ND(X))} = \\ &= \frac{P(X|PA(X)) \prod_{Y \in PA(X), ND(X)} P(Y|PA(Y))}{\prod_{Y \in PA(X), ND(X)} P(Y|PA(Y))} = \\ &= P(X|PA(X)) \end{aligned} \tag{4.16}$$

come volevasi dimostrare. Si può anche dimostrare il contrario, cioè che la relazione 4.16 implica la 4.15), e quindi che le due condizioni sono logicamente equivalenti (Koller e Friedman (2009)). In altre parole, avremmo potuto enunciare la condizione markoviana in termini di schermatura di una variabile dai suoi figli causali e derivarne il fatto che la distribuzione di probabilità deve fattorizzare, per ogni variabile, sui suoi genitori causali. Formalmente:

Definizione 4.3.3 (Distribuzione markoviana per un grafo causale - definizione per *schermatura*)**.** Una distribuzione di probabilità $P(X_1, ..., X_n)$ si dice *markoviana* per un grafo causale G e un insieme di variabili $X_1, ..., X_n$, se e solo se è tale per cui, per ogni i:

$$P(X_i|pa(X_i), ND(X_i)) = \prod_{i=1}^{n} P(X_i|PA_i) \tag{4.17}$$

In modo analogo al caso in 2 variabili, si dimostra che vale il seguente teorema:

Teorema 4.3.1. Dato un modello causale su un grafo causale G con variabili $X_1, ..., X_n$ e distribuzione di probabilità $P(X_1, ..., X_n)$, sono equivalenti le seguenti asserzioni:

1. $P(X_1, ..., X_n)$ è compatibile con la struttura causale di G (definizione 4.3.1);
2. $P(X_1, ..., X_n)$ è markoviana per G (definizione 4.3.2).

Come per il principio di Reichenbach, discutiamo le implicazioni:

— la $1 \rightarrow 2$ indica che, se possiamo introdurre delle variabili di rumore che permettono di riprodurre le relazioni di correlazioni che osserviamo come avverrebbe con un sistema dinamico, allora tale sistema è descritto da una distribuzione di probabilità che fattorizza come in equazione (4.14);

— La $2 \to 1$ è importante per l'inferenza causale. Asserisce che, se si osserva una distribuzione $P(X_1, ..., X_n)$ che fattorizza come in equazione (4.14), allora questa è descrivibile mediante dei grafi che riproducono tali correlazioni semplicemente aggiungendo dell variabili di rumore per ogni variabile del sistema. Come nel caso in due variabili, ciò non significa che tale grafo è unico: più grafi riproducono le stesse fattorizzazioni. Comprendere quale di questi sia il grafo corretto è l'obbiettivo dei metodi di inferenza causale, che discuteremo nel prossimo capitolo.

Abbiamo quindi giustificato formalmente il motivo per cui, in analisi causale, si assegnano alle distribuzioni di probabilità solo dei grafi per i quali esse siano markoviane.

4.4 I modelli a equazioni strutturali

Per dare concretezza alle formule introdotte in questo capitolo, è utile specificare le relazioni tra le variabili del sistema causale che stiamo analizzando. Le equazioni associate a tali relazioni hanno una semplice interpretazione: rappresentano le dipendenze dinamiche tra le variabili del sistema in esame. Queste relazioni funzionali sono dette *strutturali*, e vanno intese in senso adirezionale. Consideriamo, ad esempio, due variabili X e Y, disposte in uno schema causale in cui la prima è causa della seconda (figura 4.1b). Se scrivessimo semplicemente $P(Y) = f(X, \lambda)$, dal punto di vista matematico potremmo sempre invertire la relazione scrivendo $P(X) = f^{-1}(Y, \lambda)$. Tuttavia, l'irreversibilità del processo descritta nel capitolo 3 impedisce (dal punto di vista termodinamico) che Y possa causare X. Pertanto, il processo dinamico non può avvenire in senso opposto, cioè la relazione non può essere invertita. Per indicare questa direzionalità utilizzeremo la notazione:

$$Y \coloneqq f(X) \tag{4.18}$$

In questo modo stiamo fornendo un'informazione *causale*: Y è una variabile statistica determinata in modo deterministico dalla variabile X, ed il processo scorre solo in una direzione: $X \to Y$. Notiamo che c'è una relazione biunivoca tra l'insieme di equazioni strutturali e il grafo ad esso associato. Ad esempio, le relazioni:

$$\begin{cases} X_1 \coloneqq f_{X_1}(X_2, \lambda_1) \\ X_3 \coloneqq f_{X_2}(X_1, X_3, \lambda_3) \end{cases} \tag{4.19}$$

esprimono le stesse assunzioni codificate nel grafo di figura 6.5a. Generalmente, i termini di rumore sono indicati nelle equazioni strutturali ma non nei grafi causali. Questo significa che, da un punto di vista strutturale, il grafo di figura 6.5a e quello di figura 4.3 rappresentano la stessa struttura.

Notiamo che non stiamo fornendo alcuna espressione algebrica esplicita. Per questa ragione, questo tipo di analisi si dice *non-parametrica*. Se assumessimo che tutte le relazioni sono lineari, come nel caso di Wright, l'analisi sarebbe *parametrica*, e avremmo, per il modello di figura 6.5a:

$$Y \coloneqq \beta X + \lambda_Y \tag{4.20}$$

dove β è il parametro di influenza causale che esprime l'influenza di X su Y e λ_Y il rumore associato a Y. Nei prossimi capitoli, utilizzeremo i sistemi ad equazioni strutturali ogni volta che vorremo rendere esplicita l'analisi di un sistema. Inoltre, quando il contesto è sufficientemente chiaro, torneremo ad utilizzare il simbolo $=$ al posto di $\coloneqq$, per alleggerire la notazione.

Nome	**Definizione**	**Esempio**
Nodo.	Rappresentazione grafica della variabile di un sistema.	X
Arco.	Relazione tra due variabili di un sistema.	$\rightarrow, \leftarrow, \leftrightarrow, -$
Grafo.	Insieme di nodi e di archi.	$X - Y, X \rightarrow Y, ...$
Direzione di un arco.	Direzione della freccia dell'arco.	In $X \rightarrow Y$, l'arco è diretto da X a Y.
Forma dell'arco.	Relazione funzionale che lega le due variabili che collega l'arco	In $X \rightarrow Y$, la forma dell'arco potrebbe essere $Y = \beta X$.
Nodi adiacenti.	Nodi connessi da un arco.	X e Y in $X \rightarrow Y$.
Arco diretto.	Arco con una direzione.	$\rightarrow$ in $X \rightarrow Y$.
Arco adirezionale.	Arco senza una direzione.	$-$ in $X - Y$.
Arco bidirezionale.	Arco che punta in entrambe le direzioni.	$\leftrightarrow$ in $X \leftrightarrow Y$.
Scheletro di un grafo.	Grafo ottenuto rendendo adirezionali tutti gli archi di quello originale.	$X - Y$ è lo scheletro di $X \rightarrow Y$.
Grafo diretto.	Grafo con solo archi diretti.	$X \rightarrow Y \leftarrow Z$
Percorso.	Sequenza di archi tali per cui ognuno inizia con il nodo con cui era terminato il precedente e termina con il nodo con cui inizia il successivo.	$X \leftrightarrow Y \leftarrow Z$
Percorso causale.	Percorso che contiene solo archi diretti.	$X \rightarrow Y \rightarrow Z$
Percorso anticausale.	Percorso che contiene almeno un arco non diretto.	$X \rightarrow Y \leftarrow Z$
Nodi connessi.	Nodi che possono essere congiunti con un percorso.	X e Z in $X \rightarrow Y \leftarrow Z$
Loop	Percorso causale che inizia e finisce con la stessa variabile.	$X \rightarrow Y \rightarrow X$.
Grafo aciclico.	Grafo che non contiene *loop*.	$X \rightarrow Y \leftrightarrow Z$
Grafo causale.	Grafo diretto e aciclico.	$X \rightarrow Y \rightarrow Z$
Genitori causali di una variabile.	Nodi nel grafo da cui parte un arco che punta alla variabile.	X è genitore causale di Y in $X \rightarrow Y$
Antenati di una variabile.	Nodi del grafo per i quali esiste un percorso causale che li congiunge alla variabile.	X è antenato di Z in $X \rightarrow Y \rightarrow Z$
Figli causali di una variabile.	Nodi del grafo per i quali esiste un arco causale dalla variabile al nodo.	Y è figlio causale di X in $X \rightarrow Y$.
Discendenti causali di una variabile.	Nodi del grafo per i quali esiste un percorso diretto che parte dalla variabile e arriva al nodo.	Z è discendente causale di X in $X \rightarrow Y \rightarrow Z$.

Tabella 4.1: Definizioni utilizzate per descrivere un grafo causale.

Capitolo 5

L'inferenza causale

Solo dopo aver conosciuto la superficie delle cose
ci si può spingere a cercare quel che c'è sotto.
Ma la superficie delle cose è inesauribile.
I. Calvino, *Palomar*, 1983

Nel capitolo precedente abbiamo illustrato come sia possibile introdurre diagrammi causali compatibili con una prospettiva deterministica, consentendoci di descriverli come sistemi dinamici. In tal modo, abbiamo scoperto che le relazioni funzionali tra le variabili del grafo causale devono soddisfare una proprietà nota come *condizione causale markoviana*. Abbiamo accennato al fatto che ciò non è sufficiente per dedurre la struttura causale corretta di un sistema dalle correlazioni tra le variabili che lo compongono. In questa sezione vogliamo mostrare quali sono le difficoltà da affrontare quando cerchiamo di *dedurre* uno schema causale da un insieme di correlazioni. Tale deduzione costituisce l'obiettivo dei metodi di *inferenza causale*, che discuteremo brevemente in questo capitolo, illustrando alcuni contesti in cui essi vengono applicati in fisica.

5.1 Le fondamenta dei grafi causali e l'indipendenza statistica

In questa sezione analizziamo le strutture primitive dei grafi causali per comprendere le relazioni di associazione che essi implicano. Inizialmente, si esaminano le strutture con due variabili, per poi passare a quelle con tre variabili. Infine, si dimostra che la generalizzazione a grafi arbitrari è immediata, in quanto essi possono sempre essere scomposti nelle loro strutture primitive.

5.1.1 Strutture a due variabili

Consideriamo un sistema con due variabili indipendenti X_1 e X_2. Il grafo associato a tale sistema sarà costituito da due variabili prive di alcun nesso causale, come illustrato in figura 5.1a e discusso nel capitolo precedente. Per passare dal lessico grafico alle distribuzioni di

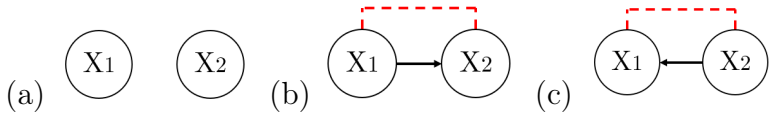


Figura 5.1: Grafi causali con due variabili, X_1 e X_2, nei seguenti casi: a) X_1 e X_2 non sono connesse; b) X_1 causa X_2; c) X_2 causa X_1. Nel caso a), le due variabili sono indipendenti, mentre nei casi b) e c) risultano associate.

probabilità associate, possiamo applicare la condizione markoviana:

$$P(x_1, x_2) = \prod_i P(x_i|pa_i) = P(x_1|pa_1)P(x_2|pa_1) = P(x_1)P(x_2) \tag{5.1}$$

Quindi, l'assenza di connessione tra i due nodi del grafo implica che le due variabili siano statisticamente indipendenti. Per indicare indipendenza statistica tra due variabili utilizzeremo la notazione:

$$X_1 \perp\!\!\!\perp_P X_2 \tag{5.2}$$

dove il pedice P indica che tale indipendenza è una relazione di indipendenza probabilistica. Consideriamo ora il caso rappresentato in figura 5.1b, in cui X_1 è causa di X_2. Analizzando il grafo, osserviamo che X_1 è genitore causale di X_2 quindi, applicando la condizione markoviana:

$$P(x_1, x_2) = \prod_i P(x_i|pa_i) = P(x_1|pa_1)P(x_2|pa_1) = P(x_1)P(x_2|x_1) \neq P(x_1)P(x_2) \tag{5.3}$$

dove la disuguaglianza è dovuta al fatto che $P(x_2) = \sum_{x_1} P(x_2|x_1)$. Ciò indica che le due variabili sono associate. Indichiamo relazioni di associazione tramite frecce tratteggiate rosse, come in figura 5.1b. Tale rapporto di associazione è simmetrico: se X_1 è associato a X_2, allora vale anche il contrario. Infatti, la stessa relazione vale nel caso illustrato in figura 5.1c, nonostante i due grafi siano differenti da un punto di vista causale. Infatti, in modo analogo:

$$P(x_1, x_2) = P(x_1|pa_1)P(x_2|pa_1) = P(x_1|x_2)P(x_2) \neq P(x_1)P(x_2) \tag{5.4}$$

Per esprimere tale simmetria, le frecce che rappresentano proprietà di associazione, sono adirezionali. Da un punto di vista associativo, dunque, i grafi di figura 5.1b e 5.1c sono equivalenti. In generale:

Due nodi adiacenti sono sempre associati.

dove ricordiamo che due nodi si dicono adiacenti se sono connessi da un arco (tabella 4.1).

5.1.2 Strutture a tre variabili

Consideriamo le strutture primitive a 3 variabili dei grafi causali, rappresentate in figura 5.3. Il grafo in figura 5.2a è detto *catena*, quello in figura 5.2b *forchetta* e quello in figura 5.2c *collider*. Data l'importanza di tali strutture, è opportuno affiancare degli esempi concreti alla loro trattazione formale.

Cominciamo considerando la struttura a catena (figura 5.2a). Una struttura del genere potrebbe essere la seguente: pioggia → strada bagnata → numero di incidenti d'auto. La pioggia (variabile X_1) bagna la strada (variabile X_2) la quale provoca un maggior numero di incidenti

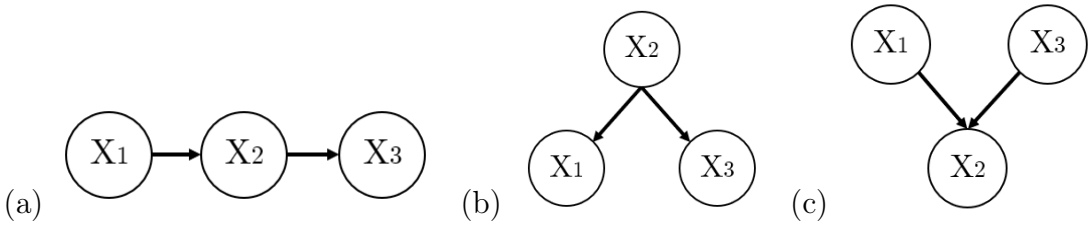


Figura 5.2: Strutture bayesiane a 3 variabili X_1, X_2 e X_3 nel caso in cui la struttura causale sia: a) a catena; b) a forchetta; c) un *collider*. In catene e forchette X_1 e X_3 sono associate nei dati grezzi e indipendenti se condizioniamo su X_2, nei *collider* sono indipendenti nei dati grezzi e associati se condizioniamo su X_2.

d'auto (variabile X_3). In questo caso, X_2 (la strada bagnata) è nota come *mediatore*, ed ha la funzione di trasmettere ad X_3 (il numero di incidenti d'auto) l'informazione che eredita da X_1 (la pioggia).

Nel caso in 2 variabili, abbiamo detto che due nodi adiacenti sono sempre associati. Infatti, la pioggia è associata con la strada bagnata, che a sua volta è associata con il numero di incidenti d'auto. Cosa possiamo dire dell'associazione tra la pioggia e il numero di incidenti? Intuitivamente, ci aspettiamo di trovare che anche queste sono associate, ossia di trovare un numero maggiore di incidenti d'auto quando piove. Ciò può essere mostrato formalmente traducendo in lessico probabilistico le assunzioni espresse nel grafo in figura 5.2a. Utilizzando la condizione markoviana per tale struttura a catena:

$$P(x_1, x_3) = P(x_1)P(x_3|x_2) \neq P(x_1)P(x_3) \tag{5.5}$$

infatti $P(x_3) = \sum_{x_2} P(x_3|x_2)$. Dunque, come ci aspettavamo, le due variabili sono associate.

Ci chiediamo ora cosa accade se condizioniamo sul mediatore, cioè la strada bagnata (variabile X_2)? Assumendo che l'unico modo in cui la pioggia può aumentare il numero di incidenti sia bagnare la strada, ci aspettiamo che, se guardiamo solo a casi in cui la strada è asciutta o bagnata non dovremmo trovare associazione tra le due variabili. Più precisamente, stiamo supponendo che, se consideriamo solo i casi in cui la strada è bagnata, il numero di incidenti stradali non dipende dal fatto che stia piovendo o meno. Le macchine scivolano a causa dell'acqua, a prescindere che questa sia caduta dal cielo o emersa da un tombino rotto. In questi casi, diremo che il mediatore X_2 *isola* X_1 e X_3. Anche tale intuizione è formalizzata con successo dal grafo di figura 5.2a. Infatti, applicando la condizione markoviana e il teorema di Bayes ad una struttura a catena condizionata su X_2 si ottiene:

$$P(x_1, x_3|x_2) = \frac{P(x_1, x_2, x_3)}{P(x_2)} = \frac{P(x_1)P(x_2|x_1)P(x_3|x_2)}{P(x_2)} \tag{5.6}$$

Ma:

$$P(x_1|x_2) = \frac{P(x_1, x_2)}{P(x_2)} = \frac{P(x_1)P(x_2|x_1)}{P(x_2)} \tag{5.7}$$

Dunque:

$$P(x_1, x_3|x_2) = P(x_1|x_2)P(x_3|x_2) \tag{5.8}$$

In una struttura a catena, quindi, le variabili X_1 e X_3 sono associate, ma un condizionamento su X_2 le rende indipendenti. Formalmente:

$$X_1 \perp\!\!\!\perp_P X_3 | X_2 \tag{5.9}$$

Consideriamo ora la struttura a forchetta, rappresentata in figura 5.1b. In questi casi, X_2 si dice *causa comune* o *variabile confondente* di X_1 e di X_2. Una variabile confondente rende X_1 e X_3 statisticamente associati anche se non sono causalmente legati. Un esempio è quello già discusso nel capitolo 6.1 con l'associazione che si trovava tra la vendita di gelati e il numero di attacchi di squali. In tal caso, la struttura è: numero di gelati venduti ← mese dell'anno → numero di attacchi di squali. Formalmente:

$$P(x_1, x_3) = P(x_1|x_2)P(x_3|x_2) \neq P(x_1)P(x_3) \tag{5.10}$$

perché $P(x_1) = \sum_{x_2} P(x_1|x_2)$ e $P(x_3) = \sum_{x_2} P(x_3|x_2)$. Anche in questo caso ci aspettiamo che X_1 e X_3 siano indipendenti se condizioniamo su X_2. Abbiamo già visto nel capitolo 6.1 che questo è il caso, ma l'abbiamo visto nel contesto ristretto dell'analisi lineare. Il formalismo dei grafi causali ci permette di mostrare che tale relazione vale per qualsiasi struttura descrivibile tramite tali nessi causali. Infatti, applicando ancora una volta la condizione markoviana e il teorema di Bayes:

$$P(x_1, x_3|x_2) = \frac{P(x_1, x_2, x_3)}{P(x_2)} = \frac{P(x_1|x_2)P(x_2)}{P(x_2)} P(x_3|x_2) = P(x_1|x_2)P(x_3|x_2) \tag{5.11}$$

Ossia X_1 e X_3 sono indipendenti se condizioniamo su X_2, come volevamo dimostrare, cioè:

$$X_1 \perp\!\!\!\perp_P X_3 | X_2 \tag{5.12}$$

Infine, consideriamo il grafo di figura 5.2c. Chiameremo X_2 un *collider*. Un esempio è il seguente: preparazione per un esame (X_1) → esito dell'esame (X_2) ← qualità del sonno la notte prima dell'esame (X_3). Prepararsi bene nei mesi precedenti e dormire bene la notte prima dell'esame, cioè i fattori X_1 e X_3, non sono, in generale, due variabili correlate. (Anche qui non ci soffermiamo su questioni del tipo: studiare meglio rende più tranquilli e di conseguenza migliora il sonno.) Tuttavia, se supponiamo di sapere che un esame è andato male, e che la notte prima lo studente abbia dormito bene, siamo portati a sospettare che non si sia preparato abbastanza nei mesi precedenti. Quindi, condizionare sul *collider* ha introdotto un'associazione tra le altre due variabili che prima non era presente. Anche questa volta, tali proprietà sono formalizzate correttamente dal grafo causale. Possiamo scrivere la probabilità congiunta di X_1 e X_3 sommando sulle probabilità di tutte le possibili istanze di X_2 (tale procedura è nota come *marginalizzazione*):

$$\begin{aligned} P(x_1, x_3) = \sum_{x_2} P(x_1, x_2, x_3) = \sum_{x_2} P(x_1)P(x_2|x_1, x_3)P(x_3) = \\ P(x_1)P(x_3) \sum_{x_2} P(x_2|x_1, x_3) \end{aligned} \tag{5.13}$$

dove, ancora una volta, si è utilizzata la condizione markoviana. Ma $P(x_2|x_1, x_3) = P(x_2)$, quindi $\sum_{x_2} P(x_2|x_1, x_3) = \sum_{x_2} P(x_2) = 1$, dunque:

$$P(x_1, x_3) = P(x_1)P(x_3) \tag{5.14}$$

cioè:

$$X_1 \perp\!\!\!\perp_P X_3 \tag{5.15}$$

Tuttavia, se condizioniamo su X_2, si perde tale indipendenza, infatti:

$$\begin{aligned} P(x_1, x_3|x_2) = \frac{P(x_1)P(x_2|x_1, x_3)P(x_3)}{P(x_2)} = \\ P(x_1)P(x_3)\frac{P(x_2|x_1, x_3)}{\sum_{x_1,x_3} P(x_2|x_1, x_3)} \neq P(x_1)P(x_3) \end{aligned} \tag{5.16}$$

Quindi, condizionare su X_2 rende X_1 e X_3 parzialmente dipendenti, come ci aspettavamo. Inoltre, lo stesso vale se condizioniamo su discendenti di un *collider*. Nel nostro esempio, un discendente causale della brutta prestazione all'esame potrebbe essere un basso voto finale. Se sappiamo che il voto finale è basso, le variabili sonno e preparazione sono correlate. Infatti, poiché il voto è basso, probabilmente lo studente ha fatto una brutta prestazione all'esame, e quindi, se la notte prima ha dormito bene, probabilmente ha studiato male nei mesi precedenti. Tale situazione è rappresentata nel grafo rappresentato in figura 5.5a, dove X_1 e X_3 sono indipendenti, ma perdono tale proprietà se condizioniamo su X_4. Infatti:

$$\begin{aligned} P(x_1, x_2, x_3|x_4) = \frac{P(x_1, x_2, x_3, x_4)}{P(x_4)} = \\ = P(x_1)P(x_2)\frac{P(x_3|x_1, x_2)P(x_4|x_3)}{\sum_{x_3} P(x_4|x_3)} \end{aligned} \tag{5.17}$$

da cui, marginalizzando su x_3:

$$\begin{aligned} P(x_1, x_2|x_4) = P(x_1)P(x_2)\left(\sum_{x_3} P(x_3|x_1, x_2)\right)\frac{P(x_4|x_3)}{\sum_{x_3} P(x_4|x_3)} = \\ = P(x_1)P(x_2)\frac{P(x_4|x_3)}{\sum_{x_3} P(x_4|x_3)} \neq P(x_1)P(x_2) \end{aligned} \tag{5.18}$$

dove nel primo passaggio si è utilizzato il fatto che $P(x_3)$ è una distribuzione di probabilità (quindi normalizzata a 1).

5.2 La *d-separazione* e l'equivalenza statistica

Abbiamo studiato in modo approfondito le strutture primitive a 2 e a 3 variabili perché qualsiasi grafo è decomponibile in unioni di tali strutture. Ad esempio, il grafo riportato in figura 5.3a è a 3 variabili, ma non è né una catena, né una forchetta, né un *collider*. Tuttavia, può essere decomposto in una struttura a forchetta ($X_1 \leftarrow X_2 \rightarrow X_3$ e $X_1 \rightarrow X_3$) e in una a catena ($X_2 \rightarrow X_1 \rightarrow X_3$). In modo analogo, il grafo di figura 5.3b può essere decomposto in una struttura a forchetta ($X_1 \leftarrow X_0 \rightarrow X_3$), un *collider* ($X_1 \rightarrow X_2 \leftarrow X_3$) e un'altra struttura a catena ($X_1 \rightarrow X_2 \rightarrow X_4$). Ovviamente, tali decomposizioni non sono uniche, ma questo non ha rilevanza per i nostri scopi. Le considerazioni sulle strutture fatte in questa sezione possono così essere estese a grafi di arbitraria complessità.

Studiare le strutture presenti nel percorso da una variabile ad un'altra ci permette di capire se queste sono statisticamente indipendenti o meno. In generale, diremo che un percorso è *bloccato* da un insieme di nodi se contiene un *collider* su cui non si sta condizionando (e se non si sta condizionando su alcun suo discendente), oppure se in tale insieme è presente una catena o una forchetta nella quale viene condizionato il mediatore o la variabile confondente, rispettivamente per una catena e per una forchetta. In entrambi i casi, infatti, l'associazione

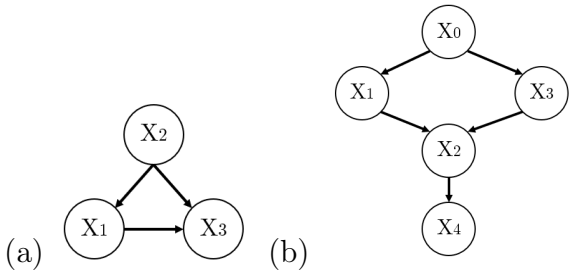


Figura 5.3: Esempi di grafi causali per mostrare come qualsiasi grafo può essere decomposto in strutture primitive.

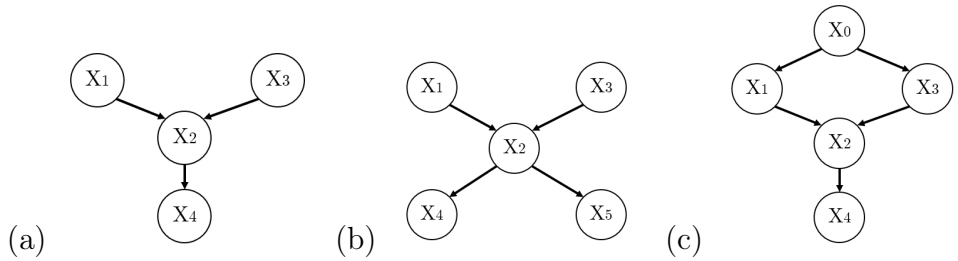


Figura 5.4: Esempi di strutture bayesiane. a) Se si condiziona su X_4, poiché è discendente di X_2, che è un *collider*, X_1 e X_3 diventano correlate; b,c) altri esempi per comprendere il concetto di *percorso bloccato*.

non viene trasportata perché *bloccata* dal *collider* o dal condizionamento. Ad esempio, nel grafo in figura 5.5b, se si condiziona su X_2, si blocca il percorso che connette X_4 e X_5, ma si apre quello che connette X_1 e X_3, che era bloccato da X_2 prima del condizionamento. In modo analogo, i percorsi $X_0 \leftrightarrow X_4$ in figura 5.5c ($X_0 \to X_1 \to X_2 \to X_4$ e $X_0 \to X_3 \to X_2 \to X_4$) sono entrambi bloccati da X_2, ma vengono aperti se si condiziona su X_2.

Quando tutti i percorsi possibili da due variabili (o insiemi di variabili) X_1 e X_3 sono bloccati da una terza variabile (o un insieme di altre variabili) X_2, si dice che X_1 e X_3 sono *d-separati* da X_2. Indicheremo tale concetto con la seguente notazione:

$$X_1 \perp\!\!\!\perp_G X_3 \tag{5.19}$$

Notiamo che una *d-separazione* può valere sia nel grafo semplice, sia a seguito di un condizionamento su alcune variabili (quando, per bloccare un percorso, è necessario chiudere una catena o una forchetta). Nel secondo caso, se la variabile (o l'insieme di variabili) su cui stiamo condizionando è X_2, scriveremo:

$$X_1 \perp\!\!\!\perp_G X_3 | X_2 \tag{5.20}$$

Nella sezione precedente abbiamo visto che, nelle strutture causali primitive, ogni volta che c'è *d-separazione* c'è anche indipendenza statistica. Dato che ogni grafo si può scomporre in tali strutture, il principio vale in generale. In questo modo si introduce una corrispondenza biunivoca tra il concetto grafico di *d-separazione* e quello statistico di indipendenza probabilistica. Ovviamente, tale proprietà dipende dall'assunzione che il grafo sia markoviano. Si può dimostrare che vale anche il contrario: se vale indipendenza statistica ogni volta che c'è *d-separazione*, allora il grafo è markoviano. (Per una dimostrazione formale si rimanda a Koller e Friedman (2009).) Il teorema che riassume questi risultati è il seguente:

Teorema 5.2.1 (*Condizione markoviana per d-separazione.*)**.** Una distribuzione di probabilità P è markoviana rispetto al grafo G se e solo se, se X_1 e X_2 sono *d-separate* in G condizionando su X_2 (che può anche essere vuoto), allora X_1 e Y_3 sono indipendenti probabilisticamente condizionando su X_2, cioè:

$$P \text{ markoviana per } G \qquad \Leftrightarrow \qquad (X_1 \perp\!\!\!\perp_G X_3 | X_2 \Leftrightarrow X_1 \perp\!\!\!\perp_P X_3 | X_2) \tag{5.21}$$

Tale teorema è quindi un'altra possibile formalizzazione della condizione markoviana, equivalente logicamente alla condizione per fattorizzazione (definizione 4.3.2) e a quella per schermatura (definizione 4.3.3). In virtù di tale teorema, la nozione grafica di *d-separazione* e quella probabilistica di *indipendenza* sono equivalenti quindi, da adesso, utilizzeremo la notazione:

$$\perp\!\!\!\perp_P = \perp\!\!\!\perp_G \equiv \perp\!\!\!\perp \tag{5.22}$$

Dato un grafo causale, la sua struttura implicherà una serie di assunzioni di indipendenza probabilistica, condizionate e non. In generale, quando due grafi causali implicano lo stesso insieme di relazioni di indipendenza statistica, si dice che appartengono alla stessa *classe di equivalenza*. Poiché le strutture a forchetta e a catena implicano le stesse relazioni di indipendenza, se cambiassimo alcune strutture a forchetta (o anche tutte) in strutture a catena, il grafo implicherebbe le stesse relazioni di indipendenza, e lo stesso varrebbe se cambiassimo strutture a catena in strutture a forchetta. Ovviamente, non possiamo fare lo stesso con i *collider*: introdurre un nuovo *collider* o eliminarne uno che prima era presente modificherebbe le relazioni di indipendenza statistica tra le nostre variabili. Quindi qualsiasi grafo equivalente a quello di partenza dovrà contenere gli stessi *collider*. In generale, si dimostra il seguente teorema (Judea Pearl (2009b)):

Teorema 5.2.2. Due grafi sono equivalenti se e solo se hanno lo stesso scheletro e contengono gli stessi *collider*.

Finora abbiamo mostrato in quale modo un grafo causale contiene delle proprietà di indipendenza statistica. Tuttavia, generalmente, quello che accade è il processo opposto. Abbiamo dei dati sui quali misuriamo un insieme di indipendenze probabilistiche, e ci chiediamo quale sia il grafo che rappresenti nel modo migliore tali dati. Quando cerchiamo di costruire il diagramma causale dobbiamo farlo in modo coerente con tali dati. Non possiamo modellizzare un sistema con una struttura a forchetta (figura 5.2b) se X_1 e X_3 non sono indipendenti quando condizioniamo su X_2. Più in generale, se due variabili sono indipendenti statisticamente sotto alcune condizioni (che dipendono dalla struttura del grafo), allora i rispettivi nodi devono essere *d-separati* nel grafo causale sotto gli stessi condizionamenti. Tuttavia, come abbiamo visto, possiamo assegnare diversi schemi causali a stesse dipendenze statistiche. Di conseguenza, quando possediamo dati osservativi su un certo numero di variabili e troviamo delle relazioni di indipendenza statistica, avremo un certo numero di grafi possibili che rispettano la condizione markoviana e le relazioni di indipendenza che abbiamo trovato. Comprendere quale di questi sia quello giusto è spesso molto complesso, e costituisce l'obbiettivo dei metodi di *inferenza causale*, che discuteremo nella prossima sezione.

5.3 L'inferenza causale

Gli algoritmi di scoperta causale, partendo dalle correlazioni osservate nei dati, intendono scoprire lo schema causale che descrive correttamente il sistema. In primo luogo, tali metodi

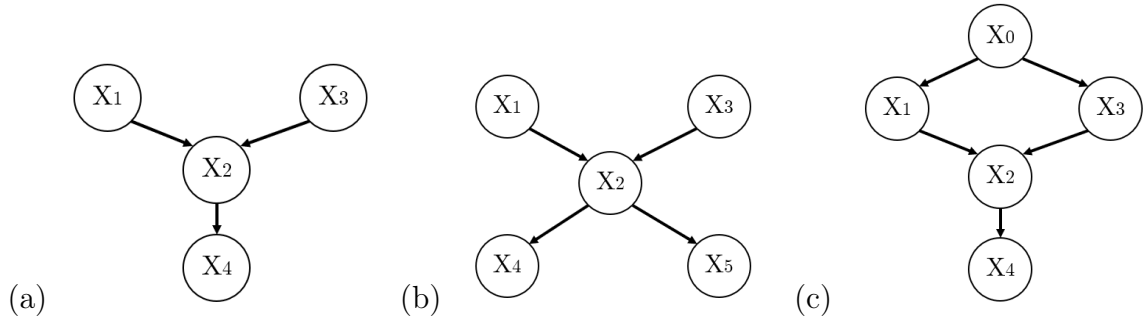


Figura 5.5: Esempi di strutture bayesiane. a) Se si condiziona su X_4, poiché è discendente di X_2, che è un *collider*, X_1 e X_3 diventano correlate; b,c) altri esempi per comprendere il concetto di *percorso bloccato*.

partono sempre concentrandosi sulle relazioni di indipendenza statistica che si rilevano nei dati. Tuttavia, spesso è problematico estrapolare tali relazioni quando i dati sono limitati. In questo elaborato non ci soffermeremo su tale problematica, ed applicheremo i metodi a distribuzioni di probabilità già note.

5.3.1 *No fine-tuning*

Supponiamo di trovare nei nostri dati che due variabili X_1 e X_3 sono indipendenti senza alcun condizionamento, ma che sono entrambe correlate con una terza variabile X_2. Questo è quanto accade, ad esempio, con un *collider* (figura 5.2c). Tuttavia, le stesse relazioni possono essere realizzate con una struttura come quella di figura 5.3a. Cerchiamo di comprenderlo con un esempio numerico, utilizzando i metodi dei coefficienti di percorso di Wright. Ad esempio, in un grafo come quello in figura 5.3a, se scegliamo i parametri in modo che:

$$\alpha_2 = -0.5 \quad \alpha_2 = 0.25 \quad \alpha_3 = 0.5 \tag{5.23}$$

Allora, applicando il metodo dei coefficienti di percorso:

$$r_{x_1,x_3} = \alpha_2 + \alpha_1\alpha_3 = -0.25 + 0.5 \times 0.5 = -0.25 + 0.25 = 0 \tag{5.24}$$

Quindi si riproducono le stesse relazioni di indipendenza statistica. Tuttavia, quando modellizziamo tali correlazioni con un *collider*, se cambiamo i parametri del primo modello *in qualsiasi modo*, tali relazioni rimangono invariate. Al contrario, *quasi tutte* le variazioni dei parametri nel grafo di figura 5.3a, eliminano l'indipendenza statistica tra X_1 e X_3. Quindi, la prima struttura causale spiega le relazioni di indipendenza in modo più robusto a cambiamenti, e i metodi di inferenza causale preferiscono scegliere questa opzione. Questa assunzione è nota nella letteratura di analisi causale con il nome di *fedeltà* (*Faithfulness*) (Judea Pearl (2009b)). Quando utilizzata in fisica, è più opportuno indicarla come *no fine-tuning* (Wood e Spekkens (2015), Allen et al. (2017)):

Definizione 5.3.1 (Criterio *no fine-tuning*.)**.** All'interno della classe di equivalenza di grafi causali che riproducono le stesse indipendenze statistiche, si preferiscono quelli tali per cui tali indipendenze continuano a valere per qualsiasi variazione dei parametri statistici.

Ciò significa che tutte le indipendenze statistiche devono essere conseguenza della struttura causale, piuttosto che dei suoi parametri.

Soffermiamoci ancora sull'esempio appena discusso. Notiamo che, in tale contesto, lo spazio dei parametri ha 3 dimensioni. Poiché abbiamo anche il vincolo $\alpha_1 + \alpha_2\alpha_3 = 0$, i parametri che riproducono l'indipendenza statistica sono rappresentati da un piano in uno spazio a 3 dimensioni, che ha misura nulla. Tale questione vale in modo del tutto generale: dato che, se lasciamo tutti i parametri possibili liberi, ogni relazione di indipendenza introduce un vincolo, lo spazio dei parametri che riproduce tale indipendenza avrà *sempre* misura nulla. Al contrario, se una relazione di indipendenza è rappresentata da una *d-separazione* nel grafo causale, questa non introduce alcun nuovo vincolo, e i parametri possono essere variati in modo arbitrario. Tale criterio è molto utile negli algoritmi di inferenza causale, di cui mostreremo un esempio di applicazione nella sezione successiva.

5.3.2 Un esempio di inferenza causale

Vediamo ora come il criterio *no fine-tuning* si applica negli algoritmi di inferenza causale. Supponiamo di voler rispondere alla domanda causale "il fumo causa il cancro ai polmoni?" Per ogni membro della popolazione statistica che possediamo, il valore della variabile S (da *smoking*), cioè il fatto che l'individuo fumi o meno, è noto, così come è noto quello della variabile C (da *cancer*), cioè se l'individuo contrae il cancro o meno. Supponiamo di osservare una correlazione tra S e C, e di aver accesso ad una terza variabile T (da *tar*), che indica se l'individuo ha catrame nei suoi polmoni o meno. Infine, supponiamo che S e T sono indipendenti quando condizioniamo su T, cioè:

$$S \perp\!\!\!\perp C|T \tag{5.25}$$

(ricordiamo che, come abbiamo visto nella sezione precedente, il simbolo $\perp\!\!\!\perp$ indica sia *d-separazione* che indipendenza statistica, in quanto nel teorema 5.2.1 si è dimostrato che le due nozioni sono equivalenti.) Infine, supponiamo che le 3 variabili siano le uniche rilevanti al problema, cioè che tutti i termini di rumore siano non correlati.[1] Vogliamo capire il modo in cui gli algoritmi di inferenza causale agiscono in questi casi.

Per farlo, considereremo ogni ipotesi possibile sull'*ordinamento causale* delle variabili. Un'assunzione di ordinamento causale implica che una variabile può agire causalmente su un'altra solamente se è più in alto nell'ordinamento. Consideriamo l'ordinamento causale $S < T < C$. La struttura causale più generale per tale ordinamento è riportata in figura 5.6a. Tale grafo ha una semplice interpretazione: il fumo causa il cancro sia introducendo catrame nei polmoni (percorso $S \to T \to C$) sia tramite altri processi biologici (percorso $S \to C$). La distribuzione di probabilità associata a tale modello si ottiene, come sempre, applicando la condizione markoviana:

$$P(S, C, T) = P(S)P(T|S)P(C|S, T) \tag{5.26}$$

Qualsiasi distribuzione di probabilità associata a tale ordinamento causale può essere scritta in questo modo, per una scelta appropriata dei parametri. Utilizziamo ora l'informazione che $S \perp\!\!\!\perp C|T$. Tale relazione di indipendenza implica che $P(C|S, T) = P(T)$. Quindi, per impedire *fine-tuning* dei parametri, eliminiamo la freccia tra S e C, ottenendo il grafo di figura 5.6b. La nuova struttura non può generare qualsiasi struttura con tale ordinamento causale, ma può

[1] Le uniche variabili nascoste con conseguenze non banali sono le cause comuni, ossia le variabili che introducono correlazione spuria tra due variabili osservate. Infatti, introdurre una variabile nascosta che sia mediatrice di un nesso causale non ha alcuna ripercussione nelle relazioni di indipendenza statistica, infatti non possiamo condizionare su di essa (perché non possediamo dati) e dunque non possiamo introdurre nuove relazioni di *d-separazione* nel grafo causale. Il discorso è analogo se la variabile nascosta è figlia causale di due variabili osservate, cioè se è un *collider*.

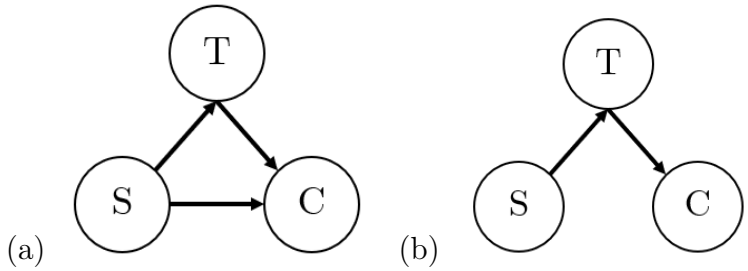


Figura 5.6: Esempi di grafi causali che legano le variabili S, T e C assumendo un ordinamento causale $S < T < C$. a) Grafo causale più generale; b) grafo causale ottenuto applicando il criterio del *no fine-tuning* e l'informazione statistica $S \perp\!\!\!\perp C|T$ al grafo di figura 5.6a.

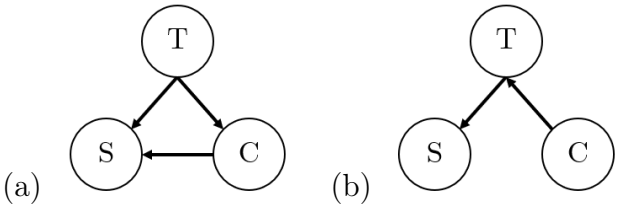


Figura 5.7: Esempi di grafi causali che legano le variabili S, T e C assumendo un ordinamento causale $C < T < S$. a) Grafo causale più generale; b) grafo causale ottenuto applicando il criterio del *no fine-tuning* e l'informazione statistica $S \perp\!\!\!\perp C|T$ al grafo di figura 5.7a.

generare qualsiasi struttura con la relazione di indipendenza probabilistica richiesta, ed è quindi una candidata per essere il grafo che correttamente interpreta il sistema. Il grafo di figura 5.6b si interpreta osservando che l'unico percorso che rende correlati S e C è mediato da T, e quindi che l'unico modo in cui il fumo causa il cancro ai polmoni è introducendo catrame in essi.

Consideriamo ora l'ordinamento $C < T < S$. La struttura causale più generale per tale ordinamento è quella riportata in figura 5.7a. La distribuzione di probabilità associata è:

$$P(S, T, C) = P(C)P(T|C)P(S|C, T) \tag{5.27}$$

Ma l'informazione di indipendenza statistica $S \perp\!\!\!\perp C|T$ implica che $P(S|C, T) = P(S|T)$. Quindi, come prima, possiamo eliminare la freccia tra C e S, ottenendo il grafo di figura 5.7b, che è un altro candidato come struttura causale.

Notiamo che, a volte, differenti ordinamenti causali generano la stessa struttura causale. Ad esempio, sia l'ordinamento $T < S < C$ che l'ordinamento $T < C < S$ conducono alla struttura causale di figura 5.8. Alcuni ordinamenti, invece, come $S < C < T$ e $C < S < T$ sono tali per cui l'osservazione di indipendenza non comporta alcuna semplificazione della struttura. Ad

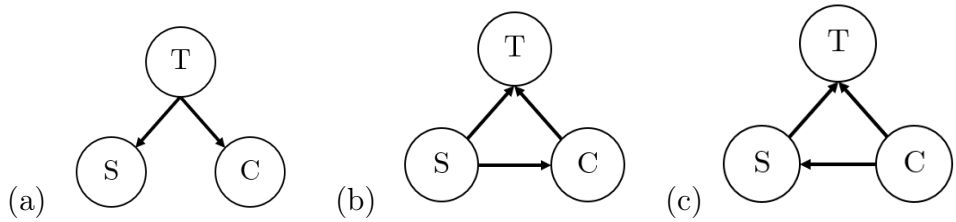


Figura 5.8: Caption.

esempio, per $S < C < T$ la distribuzione di probabilità congiunta è:

$$P(S, C, T) = P(S)P(C|S)P(T|C, S) \tag{5.28}$$

ma nessuno di questi termini può essere semplificato dall'informazione $S \perp\!\!\!\perp C|T$. $S < C < T$ è rappresentato dal grafo in figura 5.8b, mentre $C < S < T$ da quello in figura 5.8c.

Quindi, in conclusione, abbiamo 5 candidati per la struttura causale del nostro sistema, riportati nelle figure 5.6b, 5.7b, 5.8a, 5.8b e 5.8c. Tuttavia, i grafi di delle figura 5.8b e 5.8c non rispettano il criterio *no fine-tuning*, quindi li escludiamo. Infine, supponiamo di avere delle informazioni aggiuntive, che ci permettono di escludere alcuni ordinamenti causali. Ad esempio, supponiamo di sapere che il catrame (variabile T) compare sempre *dopo* il fumo (variabile S). Dato che, come discusso nel capitolo 6.1, una causa avviene sempre dopo il suo effetto, dobbiamo escludere qualsiasi ordinamento causale tale per cui $T < S$. Di conseguenza escludiamo anche i grafi in figure 5.7b e 5.8a. In conclusione, l'unico candidato che spieghi le indipendenze statistiche osservate senza *fine-tuning* dei parametri e rispettando l'ordinamento temporale noto è quello riportato in figura 5.6b. È stato dimostrato che l'algoritmo qui presentato è corretto, nel senso che, se esiste un insieme di strutture causali che rappresentano le correlazioni osservate, allora l'algoritmo riporterà quelle che non presentano *fine-tuning* dei parametri statistici.[2]

Ovviamente, non è detta che le variabili che osserviamo siano tutte quelle rilevanti per descrivere il nostro sistema. Ad esempio, potrebbe esserci un fattore genetico non osservato che introduce associazione spuria tra il fumare e il contrarre il cancro. (Le aziende di vendita del tabacco si sono appellate per anni a tale giustificazione.) Con il tempo, ci si è però resi conto che l'associazione tra fumo e cancro ai polmoni era troppo forte per essere spiegata in termini di una variabile confondente e, ad oggi, la comunità scientifica concorda sul fatto che il fumo *causi* il cancro ai polmoni (J. Pearl e Mackenzie (2018), capitolo 5). Gli algoritmi che permettono l'esistenza di variabili confondenti non osservate sono molto più complessi, e non li discuteremo in questo elaborato.

5.4 Inferenza causale e disuguaglianze di Bell (cenni)

Nel 2014 i fisici Christopher J. Wood e Robert W. Spekkens hanno pubblicato un *paper* intitolato "The lesson of causal discovery algorithms for quantum correlations: Causal explanations of Bell-inequality violations require fine-tuning" (Wood e Spekkens (2015)). In tale articolo, i metodi di inferenza causali sono stati applicati allo studio delle correlazioni tra osservabili quantistiche in esperimenti di tipo Bell, e si è mostrato che, per spiegare le correlazioni osservate, è *necessario* fare *fine-tuning* sui parametri statistici. Nelle ultime righe dell'articolo si è sottolineato che tali risultati dipendono dal fatto che i modelli causali utilizzati sono quelli sviluppati in un contesto di deterministico, e che, probabilmente, occorre modificare alcuni punti chiave per dar ragione delle differenti caratteristiche paradigmatiche della meccanica quantistica. Tale articolo ha aperto le porte ad una linea di ricerca che sta tentando di estendere il formalismo introdotto in questo elaborato al contesto teorico della meccanica quantistica.

Abbiamo visto nel capitolo 4 che il cuore della modellizzazione causale è il principio di Reichenbach, il quale permette di introdurre una condizione di compatibilità classica che, se generalizzata, porta alla condizione causale markoviana, cuore di tutti i metodi di analisi causale.

[2] Inoltre, l'algoritmo genera distribuzioni di probabilità "minimali", ossia che rappresentano il modello nel modo più semplice possibile. Questa è un'altra delle assunzioni più utilizzate in inferenza causale, che non discutiamo in questa sede in quanto concerne principalmente l'inferenza in presenza di variabili confondenti non osservate.

Tale condizione non è altro che un vincolo sui grafi compatibili con un insieme di indipendenze statistiche. Tali indipendenze, in grafi arbitrari, sono espresse dal concetto di *d-separazione* il quale, di conseguenza, andrebbe rivisitato se fossimo costretti a modificare il principio di Reichenbach.

In Pienaar e Brukner (2015) è stato sottolineato che, quando il principio di causa comune viene calato nel contesto teorico della meccanica quantistica, la sua parte quantitativa diviene problematica, e non è più logicamente equivalente alla parte qualitativa. In tale articolo viene proposta una regola di separazione grafica analoga al concetto di *d-separazione* che non dipenda dalla proprietà di fattorizzazione richiesta dal principio di Reichenbach. In Allen et al. (2017) e Costa e Shrapnel (2016) si va ancora oltre: riprendendo le osservazioni di Pienaar e Brukner (2015) sulle problematiche del principio di Reichenbach, si generalizza l'intero concetto di struttura causale nel caso quantistico, portando anche a modifiche nella formulazione della condizione markoviana.

Il tema di ricerca è ancora aperto, e non verrà ulteriormente analizzato in questo elaborato. In ogni caso, mette bene in risalto il fatto che la fisica, in alcuni ambiti, può trovare nel lavoro svolto in analisi causale delle nozioni che possono rivelarsi cruciali.

Capitolo 6

Il *do-calculus*: dalle correlazioni alla causalità

$$P(y|do(t)) = P(y|t)$$
(o "Correlation *is* causation".)

Nel capitolo 4 abbiamo introdotto i grafi causali, strutture matematiche che ci permettono di analizzare un sistema statistico in senso dinamico. Nel capitolo 5 abbiamo evidenziato che tali strutture non sono generalmente deducibili dai dati. In questo capitolo metteremo da parte questa difficoltà, assumendo che il grafo del sistema sia noto. Questo ci consentirà di introdurre metodi analitici che permettono di quantificare l'effetto *causale* di una variabile su un'altra.

6.1 Quando "Correlation *is* causation"

Nel capitolo 2 abbiamo mostrato in che modo l'analisi statistica fallisce nel distinguere associazioni spurie da associazioni causali. Quello che manca a tale scienza, in effetti, è il formalismo dei grafi causali che abbiamo sviluppato nei precedenti capitoli. Una volto noto il grafo causale che rappresenta un sistema, definire e distinguere associazioni spurie e associazioni causali è immediato.

Per analizzare l'associazione tra due variabili è necessario considerare tutti i percorsi che le connnettono nel grafo causale, siano essi causali (composti esclusivamente da archi causali) o anticausali (che includono anche archi anticausali) (tabella 4.1). Definiamo:

> Associazione *totale*: associazione dovuta a tutti i percorsi non bloccati che congiungono due variabili.
> Associazione *causale*: associazione dovuta ad ogni percorso causale non bloccato.
> Associazione *spuria*: associazione dovuta ad ogni percorso anticausale non bloccato.

Ad esempio, l'associazione tra l'altezza di un padre e quella del figlio vista nel capitolo 2, che può essere rappresentata graficamente con il grafico di figura 6.1a, è interamente causale, mentre quella tra la vendita dei gelati e il numero di attacchi di squali (figura 6.1b) è interamente

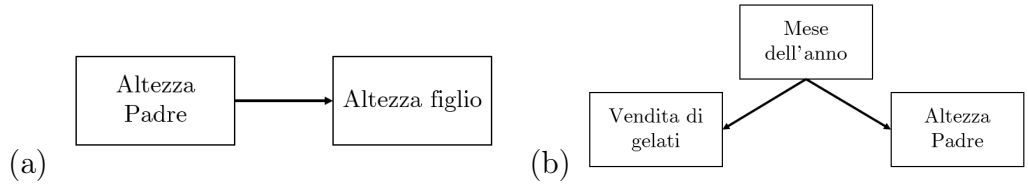


Figura 6.1: Grafi causali associati agli esempi del capitolo 2: a) processo di eredità dell'altezza da un padre a un figlio; b) correlazione tra attacchi di squali e vendita di gelati, interpretata con una variabile confondente, il mese dell'anno.

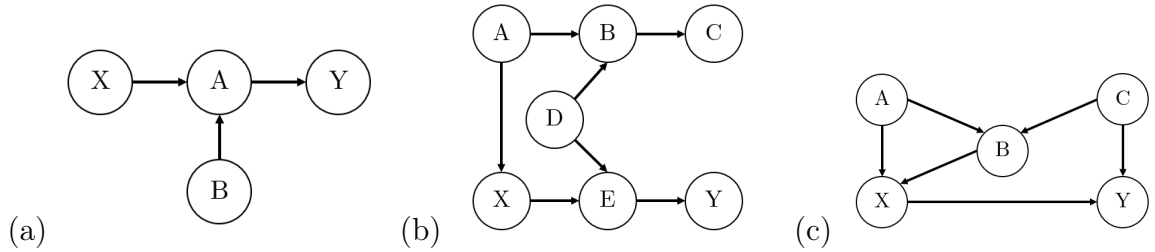


Figura 6.2: Esempi di grafi causali per comprendere il concetto di *d-separazione* e comprendere come ottenere informazione causale da dati osservazionali.

spuria. Pertanto, la causalità è una sottocategoria dell'associazione: alcune associazioni sono causali, mentre altre non lo sono, e questa distinzione dipende dal grafo che rappresenta il nostro sistema. Questo significa che è vero che spesso "correlation is not causation", ma, a volte, misurare un'associazione esprime la forza di un nesso causale, cioè "correlation *is* causation" (J. Pearl e Mackenzie (2018)).

Ovviamente, l'associazione può essere in parte causale e in parte spuria. Il problema fondamentale dell'analisi causale è quello di cancellare l'associazione spuria da quella totale, per ottenere solo quella causale, in modo da misurare la forza di un nesso causale dai dati osservazionali (Neal (2020)). Per comprendere se l'associazione tra due variabili è causale, è sufficiente verificare se, nel grafo che rappresenta il sistema, queste sono connesse da percorsi causali o meno, o se i percorsi anticausali che le collegano solo bloccati. Se tale condizione è soddisfatta, misurare la loro associazione equivale a misurare la forza del loro legame causale. In caso contrario, ciò non è sufficiente, poiché parte di essa potrebbe essere spuria. Tuttavia, potrebbero esserci variabili tali per cui, se condizioniamo su di esse, possiamo bloccare tutto lo scorrere di associazione anticausale, rendendo esclusivamente causale quella residua. In tal caso, le leggi della probabilità condizionata ci permettono di ottenere la forza di un nesso causale dalla semplice misura dell'associazione tra il trattamento e l'effetto, questa volta, però condizionando su alcune variabili.

Consideriamo, ad esempio, i grafi in figura 6.2, e chiediamoci come valutare la forza del nesso causale tra X e Y basandoci solo sulle distribuzioni di probabilità osservazionali. Nel grafo di figura 6.2a non occorre condizionare su alcuna variabile, infatti non ci sono percorsi anticausali che legano X e Y. Di conseguenza, l'associazione tra le due variabili *è* il loro nesso causale. Nel grafo di figura 6.2b, invece, la freccia che collega A e X punta ad X, potenzialmente introducendo un'associazione non causale. Tuttavia, l'unico percorso che congiunge X e Y tramite tale freccia è bloccato dal *collider* B. Quindi, anche in questo caso, non è necessario condizionare su alcuna variabile: tutta l'associazione nei dati tra X e Y è di natura causale. Consideriamo infine il caso della figura 6.2c: in questo grafo ci sono due percorsi anticausali

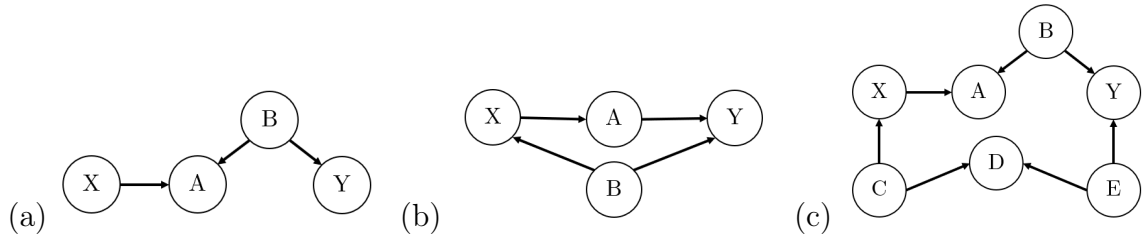


Figura 6.3: Esempi di grafi causali per comprendere l'origine dell'associazione spuria.

che collegano X e Y: $X \leftarrow B \leftarrow C \rightarrow Y$ e $X \leftarrow A \rightarrow B \leftarrow C \rightarrow Y$. Per rendere causale tutta l'associazione tra X e Y dobbiamo bloccare entrambi questi percorsi, ad esempio condizionando su C. Potremmo anche considerare di condizionare su B, ma in tal caso, essendo B un *collider*, apriremmo il percorso $X \leftarrow A \rightarrow B \leftarrow C \rightarrow Y$, che dovremmo chiudere condizionando su A *o* su C. Quindi, sia il condizionamento $\{C\}$ che i condizionamenti $\{B, A\}$ o $\{B, C\}$ permettono di eliminare le associazioni spurie tra X e Y. Nelle prossime sezioni formalizzeremo queste intuizioni.

Cerchiamo adesso di comprendere da dove emerge l'associazione spuria. Un percorso causale è composto solo da archi causali, quindi deve necessariamente cominciare con un arco causale. Consideriamo ora un percorso anticausale che congiunge due variabili e che inizia con un arco causale. Per avere associazione spuria, è necessario che almeno uno degli archi sia anticausale. Di conseguenza, appena si introduce un arco anticausale, si sta introducendo un *collider*, e quindi il percorso è bloccato. Ad esempio, se il percorso anticausale tra le variabili X e Y è $X \rightarrow Z_1 \rightarrow ... \rightarrow Z_k \leftarrow Z_{k+1} \rightarrow Z_{k+2} \rightarrow ... \rightarrow Y$, Z_k è un *collider*. Di conseguenza, un percorso che introduce associazione spuria deve necessariamente cominciare con un arco anticausale, cioè una freccia che punta alla variabile di trattamento. Quindi:

> L'associazione spuria può essere introdotta esclusivamente da percorsi che cominciano con un arco che punta alla variabile di trattamento.

Quando vogliamo calcolare l'effetto di una variabile su di un'altra, dunque, è su questi percorsi che dobbiamo concentrare la nostra attenzione.

Vediamo un esempio concreto nei grafi di figura 6.3. Nel grafo di figura 6.3a il percorso $T \rightarrow A \leftarrow B \rightarrow Y$ che congiunge X e Y è anticausale. Tuttavia, questo comincia con un arco anticausale, quindi ci aspettiamo che sia bloccato. In effetti questo è il caso: il *collider* A blocca il flusso di associazione, quindi X e Y sono *d-separati*, e dunque statisticamente indipendenti (teorema 5.2.1). Nel caso di figura 6.3b, oltre al percorso già analizzato nel caso precedente, è presente quello anticausale $Z \leftarrow C \rightarrow Y$. Questa volta, tale percorso inizia con un arco anticausale, quindi *potrebbe* introdurre associazione spuria. In effetti, il percorso non è bloccato. Poiché dall'altro percorso non scorre alcuna associazione, in questo caso l'associazione tra X e Y è interamente spuria. Infine, nel grafo di figura 6.3c, ci sono due percorsi anticausali. Uno è quello già discusso per il caso di figura 6.3a, che è bloccato dal *collider* A. L'altro è il percorso anticausale $X \leftarrow C \rightarrow D \leftarrow E \rightarrow Y$, che comincia con un arco anticausale e, dunque, *potrebbe* introdurre associazione spuria. Tuttavia, questo è bloccato dal *collider* C, quindi tale associazione non scorre, e, anche in questo caso, X e Y sono *d-separati*. In generale, dunque:

> Il fatto che il percorso anticausale che connette due variabili inizi con un arco anticausale è condizione necessaria, ma non sufficiente, affinché tale percorso introduca

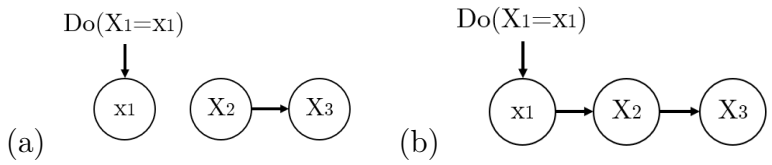


Figura 6.4: a) Grafo causale di una struttura a forchetta. b) Grafo causale di una struttura a forchetta a seguito dell'intervento sulla variabile X_2.

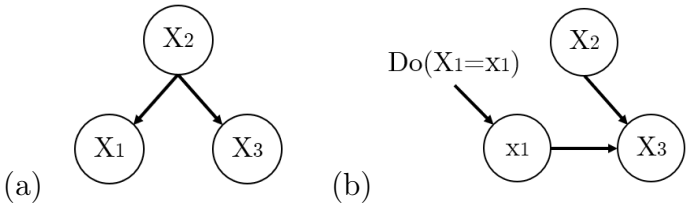


Figura 6.5: a) Grafo causale di una struttura a forchetta. b) Grafo causale di una struttura a forchetta a seguito dell'intervento sulla variabile X_1.

associazione spuria.

6.2 Il *do-operator*

Nella precedente sezione abbiamo visto che tutta l'associazione spuria tra due variabili è introdotta da percorsi anticausali che iniziano con una freccia che punta alla variabile di trattamento. Se non ci fossero tali frecce, dunque, tutta l'associazione che scorre tra le due variabili sarebbe causale. Ovviamente, questa potrebbe comunque essere nulla, come nel caso di figura 6.3a. Di conseguenza, per esprimere una domanda causale del tipo "qual è l'effetto della variabile T sulla variabile Y?" (per uniformarci alla notazione utilizzata in questi ambiti, da adesso chiameremo sempre T la variabile di *trattamento*, e Y la variabile su cui vogliamo vedere l'effetto di T), possiamo introdurre un operatore che modifica il grafo causale eliminando tutte le frecce che puntano alla variabile di trattamento. Tale azione è analoga a ciò che viene fatto nella pratica sperimentale: si modifica il valore della variabile di trattamento e, se non cambia nulla in quella su cui vogliamo misurare l'effetto, diciamo che non c'è un nesso causale tra le due.

Tuttavia, in analisi statistica, generalmente non possiamo fare esperimenti sul nostro sistema. Ad esempio, come potremmo mostrare che l'associazione tra la vendita dei gelati e gli attacchi di squali è spuria? Se chiudessimo tutte le gelaterie, in modo da impedire a chiunque di mangiare gelati, potremmo mostrare che, comunque, ci sono attacchi di squali, e quindi che l'associazione era spuria. Ovviamente, però, tale strada non è generalmente percorribile. Il formalismo causale sviluppato nei capitoli precedenti, ciononostante, ci permette di *simulare* l'effetto di un esperimento e dedurre come si comporterebbe il nostro sistema sotto l'effetto di un *intervento* (Woodward (2005)). Questo non significa altro che *isolare i percorsi di associazione causale*. Tale intuizione può essere formalizzata tramite il *do-operator*, che introduciamo in questa sezione. Tale operatore ci permette di precisare il significato della domanda: "quale è l'effetto causale di una variabile su un'altra?"

È conveniente introdurre il *do-operator* discutendo come questo agirebbe nelle due strutture primitive equivalenti dal punto di vista statistico, cioè la catena e la forchetta, riportate rispettivamente in figura 6.4a e 6.5a. Supponiamo di *intervenire* in entrambi i grafi su X_1 fissando

il suo valore a x_1. In tal modo, X_1 non dipende più dai suoi genitori causali: la sua probabilità sarà pari a 1 per il valore a cui lo fissiamo e pari a 0 per tutti gli altri valori. Questo significa modificare il grafo causale del sistema. La struttura che si ottiene per il grafo a forchetta è riportata in figura 6.5b, quella che si ottiene con il grafo a catena in figura 6.4b. Le nuove distribuzioni di probabilità, sotto l'assunzione di markovianità, sono:

$$\begin{aligned} P_{\text{forchetta}}(x_2, x_3|do(X_1 = x_1)) &= P(x_2)P(x_3|x_2) \\ P_{\text{catena}}(x_2, x_3|do(X_1 = x_1)) &= P(x_2|x_1)P(x_3|x_2) \end{aligned} \tag{6.1}$$

L'utilizzo dell'operatore *do* ha introdotto una nuova distribuzione di probabilità, detta *interventionale*, che differisce da quella originaria, detta *osservazionale*. Il fatto che queste siano differenti riflette le diverse assunzioni causali che abbiamo introdotto nel grafo. Per vederlo in modo immediato, consideriamo la come varia la probabilità di X_3 a seguito dell'intervento nei due grafi. Intuitivamente, ci aspettiamo che in quello a forchetta questa non sia modificata, perché X_1 non è una causa di X_3, mentre in quello a forchetta sì. In effetti, marginalizzando su X_2:

$$\begin{aligned} P_{\text{fork}}(x_3|do(X_1 = x_1)) &= \sum_{x_2} P(x_2)P(x_3|x_2) = P(x_3|x_2) \\ P_{\text{chain}}(x_3|do(X_1 = x_1)) &= \sum_{x_2} P(x_2|x_1)P(x_3|x_2) = P(x_3|x_2) \end{aligned} \tag{6.2}$$

Nel caso della forchetta:

$$P_{\text{fork}}(x_1|x_3) = P_{\text{fork}}(x_1|do(X_3 = x_3)) \tag{6.3}$$

Quindi X_3 *non* ha effetto causale su X_1 nella struttura a forchetta. Al contrario, nel caso della catena:

$$P_{\text{chain}}(x_1|X_3) \neq P^{\text{chain}}(x_1|do(X_3 = x_3)) \tag{6.4}$$

Quindi X_3 *ha* effetto causale su X_1 nella struttura a catena. Diremo dunque che:

> Una variabile ha effetto causale su un'altra se la probabilità della seconda varia nella distribuzione di probabilità ottenuta intervenendo sulla prima.

Cerchiamo adesso di strutturare il discorso in modo più formale. In primo luogo, assumiamo esplicitamente che, a seguito di un intervento, il grafo sia ancora markoviano. Ciò è naturale perché significa semplicemente che il sistema non cessa di essere un sistema dinamico a seguito dell'intervento, e implica che le probabilità $P(x_i|pa_i)$ dei nodi su cui non abbiamo effettuato un intervento non cambiano. Inoltre, come già discusso, $P(x_i|pa_i) = 1$ per tutti i nodi oggetto di intervento (possiamo intervenire anche su più nodi allo stesso momento). Chiaramente, ciò vale solo se x_i assume proprio il valore di intervento: per tutti gli altri valori, la probabilità sarà nulla. In una struttura come quella in figura 6.6a, un'operazione di *do* sulla variabile E modifica il grafo come in figura 6.6b. Se si effettua anche un'operazione di *do* sulla variabile B, si ottiene il grafo di figura 6.6c. Generalizzando, se interveniamo su un insieme di nodi S, la distribuzione di probabilità post-intervento sarà:

$$P\left(x_1, x_2, ..., x_n|do(S = s)\right) = \prod_{i \notin S} p(x_i|pa_i) \tag{6.5}$$

Ad esempio, nel caso di figura 6.6b:

$$P\left(a, b, c, d, x, e, y|do(E = e)\right) = P(a)P(b|d)P(c|b)P(d)P(x|a)P(y|e) \tag{6.6}$$

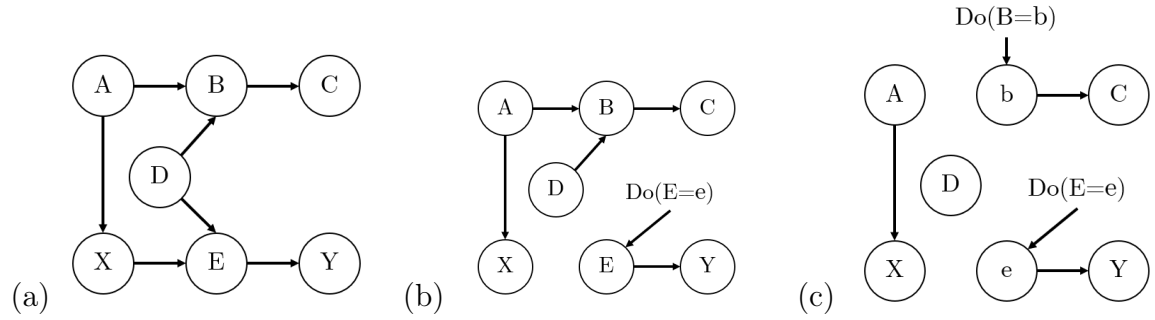


Figura 6.6: Grafo causale di una struttura complessa a) prima di un intervento sulla variabile E; b) dopo un'operazione $do(E = e)$.

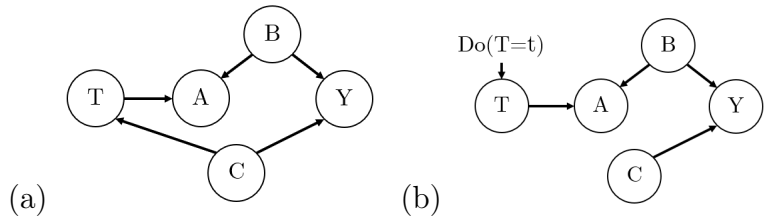


Figura 6.7: Esempio di grafo causale a) prima dell'intervento; b) dopo l'intervento. Notiamo che l'associazione tra T e Y è interamente spuria.

e in quello di figura 6.6c:

$$P\left(a, c, d, x, e, y|do(E = e), do(B = b)\right) = P(a)P(c|b)P(d)P(x|a)P(y|e) \tag{6.7}$$

Quando studiamo le probabilità post-intervento eliminiamo tutte le frecce che puntavano alla variabile di trattamento, cancellando le associazioni spurie. In tale distribuzione, dunque, tutto lo scorrere di associazione è causale. Se riusciamo a calcolare $P(y|do(t))$, avremo calcolato l'effetto causale di T su Y. Chiaramente, la risposta può essere anche che non c'è alcun effetto causale. Ad esempio, consideriamo il grafo di figura 6.7, e interveniamo sulla variabile T per studiarne l'effetto causale su Y. Intuitivamente, è semplice capire che la probabilità di Y non cambia a seguito dell'intervento. Infatti, non ci sono percorsi causali che connettono le due variabili. Nel grafo post-intervento di figura 6.7b, l'unico percorso che unisce T e Y è $T \to A \leftarrow B \to Y$, che è anticausale a causa dell'arco anticausale $A \leftarrow B$. Di conseguenza, ci aspettiamo che la distribuzione di probabilità di Y non vari. Possiamo vedere tale invarianza esplicitamente. In primo luogo, introduciamo la distribuzione di probabilità post-intervento:

$$P(a, b, y, c|do(t)) = P(a|t)P(b)P(y|b, c)P(c) \tag{6.8}$$

marginalizzando su a, b e c:

$$\begin{aligned} P(y|do(t)) &= \sum_a \sum_b \sum_c P(a|t)P(b)P(y|b, c)P(c) = \\ &= \left(\sum_a P(a|t)\right)\left(\sum_b P(b)\right)\left(\sum_c P(c)\right) P(y|b, c) = \\ &= P(y|b, c) = P(y) = P(y|t) \end{aligned} \tag{6.9}$$

dove nel secondo passaggio si è utilizzato che $P(a|t) = P(a) \Rightarrow \sum_a P(a|t) = 1$ e che, poiché Y non è figlia di T, $P(y) = P(y|t)$. In questo modo abbiamo dimostrato formalmente che T non ha effetto causale su Y, date le assunzioni del grafo G.

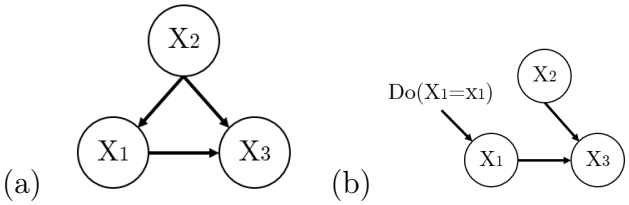


Figura 6.8: Grafo causale composto da 3 variabili X_1, X_2, X_3: a) prima di un intervento; b) dopo un intervento $do(X_1 = x_1)$.

Adesso vediamo come l'operatore *do* ci permette di calcolare l'effetto causale di una variabile su un'altra quando questo non è nullo. Analizziamo il grafo di figura 6.8. Siamo interessati all'effetto causale di T su Y. La probabilità $P(Y|T)$ non fornisce questa informazione, poiché contiene anche un'associazione spuria dovuta al percorso $T \leftarrow X \rightarrow Y$. Consideriamo ora il modello post-intervento, dove si effettua l'operazione $do(T = t)$, riportato in figura 6.8b. La nuova distribuzione di probabilità è:

$$P(x, y|do(t)) = P(x)P(y|x, t) \tag{6.10}$$

dove si è usata la notazione:

$$P(y|do(t)) \equiv P(t|do(T = t)) \tag{6.11}$$

Per ottenere la nuova distribuzione di Y, è sufficiente mediare su X:

$$P = (y|do(t)) = \sum_x P(x)P(y|x, t) \tag{6.12}$$

Notiamo che, se le relazioni fossero lineari, ciò equivarrebbe a studiare la correlazione tra T e Y con X costante, cioè utilizzare il coefficiente dell'equazione (2.6). Al contrario, considerare la semplice associazione significherebbe esaminare il coefficiente di correlazione tra T e Y senza alcun condizionamento (equazione 2.2). Infatti, da un punto di vista probabilistico (senza fare assunzione sulle relazioni tra le variabili), applicando la condizione markoviana e condizionando su X:

$$P(y|t) = \sum_y P(t|x)P(y|x, t) \tag{6.13}$$

che è differente dall'espressione trovata prima.

In generale, non si utilizza semplicemente la probabilità post-intervento come indice di influenza causale. Un indicatore molto utilizzato è l'ITE (*individual treatment effect*), che mostra quanto varia *in media* il valore della variabile effetto ad un cambiamento di quella di trattamento. Ad esempio, se T è una variabile binaria $T = 0 \vee T = 1$, tale indicatore si può calcolare con la formula:

$$\text{ITE}\,(Y|do(T)) = \mathbb{E}\,[P(Y/do(T = 1)) - P(Y/do(T = 0))] \tag{6.14}$$

Dove $\mathbb{E}$ indica il valore di aspettazione della probabilità in questione.[1] Nel caso di figura 6.8, ad esempio, l'ITE sarebbe nullo, perché T non ha alcun effetto causale su Y, cioè:

$$\text{ITE}\,(T|do(T)) = \mathbb{E}\,[P(Y/do(T = 1)) - P(Y/do(T = 0))] = 0 \tag{6.15}$$

[1]Tale notazione è molto utilizzata nel formalismo dei *potential outcome*, che non discutiamo il questo elaborato (Imbens e Rubin (2015)). Infatti, in tale contesto teorico, spesso ci si interessa all'effetto di fenomeni quali l'attuazione di una politica o di un trattamento medico sulla popolazione. Semplificando, in questi casi, la variabile di trattamento è binaria: la procedura politica o si attua o non si attua, e quindi si utilizzano espressioni quali quella in equazione (6.14) per calcolarne l'effetto causale.

Quando le variabili utilizzate sono continue, in generale si utilizza come indicatore causale la derivata dell'effetto rispetto alla causa:

$$\frac{\partial}{\partial t}\mathbb{E}\left[P(Y/do(t))\right] \tag{6.16}$$

utilizzeremo tale indicatore quando mostreremo un'applicazione dell'operatore *do* in sistemi ad equazioni strutturali con relazioni lineari.

In conclusione, notiamo che intendere la nozione di causalità in termini di azione su un sistema avvicina ulteriormente il formalismo che abbiamo sviluppato alla fisica. Ad esempio, in Aurell e Del Ferraro (2016), nel contesto della teoria statistica della risposta lineare, un *do* su un sistema viene interpretato come una funzione di risposta del ad un intervento esterno. Nell'ambito della teoria della risposta lineare, queste funzioni di risposta sono semplici da calcolare, e gli autori suggeriscono che questi metodi della fisica possono essere utilizzati per risolvere alcune distribuzioni di probabilità post-intervento. Tale similitudine è stata sottolineata anche in Baldovin, Cecconi e Vulpiani (2020), dove questa viene utilizzata per discutere metodi di inferenza causale su sistemi markoviani, dove con "markoviano" si intende che lo stato del sistema al tempo t dipende solo dal suo stato al tempo $t-1$. In realtà, i quantificatori di causalità che sono stati proposti in fisica negli ultimi decenni sono vari. I metodi di analisi causale hanno permesso di far chiarezza sul significato di tali quantificatori. In Smirnov (2022), ad esempio, viene proposta un'unificazione di tali metodi che segue le indicazioni epistemologiche dei metodi discussi in questo elaborato. Non ci soffermeremo ulteriormente su questa tematica.

6.2.1 Il metodo *back-door*

In questa sezione introdurremo il criterio più semplice per stabilire se una distribuzione di probabilità post-intervento sia deducibile da dati osservazionali. Nelle sezioni precedenti abbiamo visto che, se tutti i percorsi anticausali, che da adesso chiameremo anche *back-door*, che connettono due variabili sono bloccati (dalla presenza di un *collider* o dall'aggiustamento su una variabile), allora tutta l'associazione che scorre tra le due è causale. Di conseguenza, quando possiamo bloccare tutti questi percorsi, possiamo dedurre una quantità causale dai dati osservazionali. Questo è quanto abbiamo fatto nella sezione precedente: abbiamo eliminato l'associazione spuria dovuta a X_2 semplicemente condizionando su tale variabile. In questa sezione generalizzeremo questa operazione a grafi arbitrari, dove possono essere presenti più percorsi *back-door*.

In prima istanza, considereremo il caso in cui non ci sono variabili intermedie tra quella su cui interveniamo e quella di cui vogliamo analizzare l'effetto causale. Un esempio di grafo in cui ci sono più percorsi anticausali tra T e Y è riportato in figura 6.9a. In questo caso, i percorsi *back-door* sono $T \leftarrow X_1 \rightarrow Y$, $T \leftarrow X_2 \rightarrow X_4 \rightarrow Y$, $T \leftarrow X_2 \leftarrow X_3 \rightarrow X_4 \rightarrow Y$ e $T \leftarrow X_2 \rightarrow X_5 \rightarrow X_4 \rightarrow Y$. Diremo che un insieme di variabili W *soddisfa il criterio back-door relativamente a* T *e* Y se valgono entrambe le seguenti condizioni:

— W blocca tutti i percorsi *back-door* tra T e Y;

— W non contiene discendenti causali di T.

(Ricordiamo che una variabile si dice discendente causale di un'altra se esiste un percorso causale che le lega.) Nel nostro caso, ad esempio, W potrebbe essere $\{X_1, X_2\}$ o $\{X_1, X_4\}$. Notiamo, in particolare, che: a) se condizioniamo su X_1 o X_4 non è necessario condizionare anche su X_3, infatti il percorso $T \leftarrow X_2 \leftarrow X_3 \rightarrow X_4 \rightarrow Y$ è comunque bloccato; b) il percorso per $T \leftarrow$

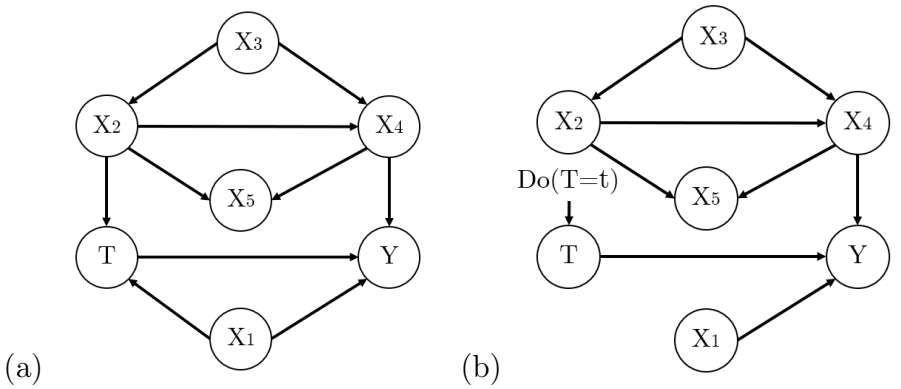


Figura 6.9: Esempio di grafo causale in cui ci sono più percorsi *back-door* che introducono associazione spuria tra T e Y.

$X_2 \to X_5 \to X_4 \to Y$ è già bloccato perché X_5 è un *collider*. Di conseguenza, condizionando su tali variabili otteniamo l'effetto causale di T su Y. La distribuzione di probabilità post intervento, riportata in figura 6.9b sarà:

$$\begin{aligned} P(x_1, x_2, x_3, x_4, x_5, y|do(t)) = \\ P(x_1)P(x_2|x_3)P(x_3)P(x_4|x_3)P(x_5|x_2, x_4)P(y|x_4, x_1) \end{aligned} \tag{6.17}$$

Marginalizzando su X_1 e X_2 (come possibile scelta di W):

$$P(y|do(t)) = \sum_{X_1, X_2} P(x_1)P(x_2|x_3, |do(t))P(x_3)P(x_4|x_3)P(x_5|x_2, x_4)P(y|x_4, x_1) \tag{6.18}$$

che può essere riscritta:

$$P(y|do(t)) = \sum_{w} P(y|do(t), w)P(w, do(t)) \tag{6.19}$$

Ma abbiamo detto che W blocca tutti i percorsi anticausali tra T e Y, quindi "correlation *is* causation", il che significa:

$$P(y|do(t), w) = P(y|t, w) \tag{6.20}$$

Inoltre, nessun elemento di W può dipendere da T in quanto, per definizione, è parte di un percorso che contiene una freccia anticausale che punta a T. Di conseguenza:

$$P(w|do(t)) = P(w) \tag{6.21}$$

E, in conclusione:

$$P(y|do(t)) = \sum_{w} P(y|t, w)P(w) \tag{6.22}$$

che è la formalizzazione delle intuizioni della sezione precedente. Tale formula significa che, quando il criterio *back-door* è soddisfatto dall'insieme W, per calcolare l'effetto di T su Y è sufficiente condizionare la probabilità su tale insieme. Se le relazioni tra le variabili fossero lineari, quindi studiate dalla teoria della regressione lineare, applicare il metodo *back-door* significherebbe che per trovare l'effetto causale di T su Y è sufficiente fare la regressione tenendo costanti tutti i fattori di W.

Vediamo ora come possiamo utilizzare il metodo *back-door* quando l'influenza di T su Y è mediata da un'altra variabile, indicata con M. Se il grafo associato è come quello in figura

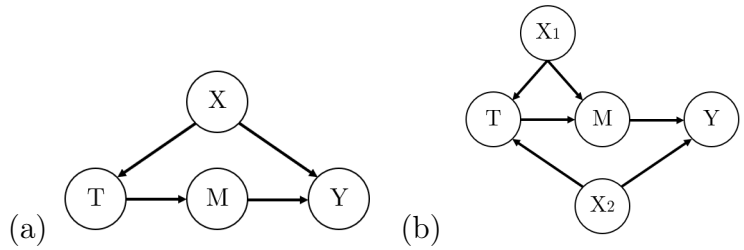


Figura 6.10: Si mostrano: a) un grafo causale in cui l'influenza causale di T su Y è mediata dalla variabile M, e c'è una variabile confondente X tra T e Y; b) lo stesso grafo in cui è aggiunta un'ulteriore variabile confondente X_2 tra T e M (e X viene chiamata X_1).

6.10, il procedimento è semplice: basta eliminare M dal grafico e utilizzare T come genitore causale di Y. In questo modo:

$$P(y|do(t)) = \sum_x P(y|t,x)P(x) \tag{6.23}$$

In alternativa, possiamo studiare l'influenza di T su M, e successivamente quella di M su Y. In primo luogo, notiamo che il percorso *back-door* tra T e M ($T \leftarrow X \rightarrow Y \leftarrow M$) è bloccato dal *collider* in Y. Quindi tutta l'informazione che scorre è causale, cioè:

$$P(m|do(t)) = P(m|t) \tag{6.24}$$

Per quanto riguarda $P(y|do(m))$, osserviamo che c'è un percorso *back-door* che può essere chiuso condizionando su X ($M \leftarrow T \leftarrow X \rightarrow Y$). Quindi:

$$P(y|do(m)) = \sum_x P(y|m,x)P(x) \tag{6.25}$$

Per ottenere il passaggio di informazione causale da T a Y dobbiamo mediare su tutti i possibili valori di M:

$$\begin{aligned} P(y|do(t)) = \sum_m P(m|do(t))P(y|do(m)) = \sum_m P(m|t) \sum_x P(y|m,x)P(x) = \\ = \sum_x \left(\sum_m P(m|t) \right) P(y|m,x)P(x) \end{aligned} \tag{6.26}$$

Ma $P(m) = P(m|t)$, quindi $\sum_m P(m|t) = \sum_m P(m) = 1$, dunque:

$$P(m|do(t)) = \sum_x P(y|m,x)P(x) \tag{6.27}$$

Che è quanto trovato precedentemente.

Esaminiamo ora il caso in cui ci sono variabili confondenti anche tra T e M, come in figura 6.10b. Ancora una volta, cominciamo studiando l'influenza di T su M, e successivamente quella di M su Y. Applicando l'aggiustamento *back-door* per la connessione $T \rightarrow M$ si ottiene:

$$P(m|do(t)) = \sum_{x_1} P(y|t,x_1)P(x_1) \tag{6.28}$$

Consideriamo ora l'effetto di M su Y, cioè $P(y|do(m))$. Notiamo che M e Y sono connessi dai percorsi anticausali $M \leftarrow X_1 \rightarrow T \leftarrow X_2 \rightarrow Y$ e $M \leftarrow T \leftarrow X_2 \rightarrow Y$, il primo dei quali

è bloccato dal collider T. Per chiudere il secondo percorso, non possiamo condizionare su T perché altrimenti riapriremmo il primo. Tuttavia, se condizioniamo su X_2, chiudiamo entrambi i canali. Di conseguenza:

$$P(y|do(m)) = \sum_{x_2} P(y|m, x_2)P(x_2) \tag{6.29}$$

Infine, mediando su M:

$$\begin{aligned} P(y|do(t)) = \sum_{m} P(m|do(t))P(y|do(m))P(m) = \\ \sum_{m}\sum_{x_1} P(y|t, x_1)P(x_1) \sum_{x_2} P(y|m, x_2)P(x_2) \end{aligned} \tag{6.30}$$

Tale distribuzione di probabilità è calcolabile direttamente dai dati se si hanno misure di tutte le variabili. Tuttavia, questo non è sempre il caso. Nella prossima sezione generalizzeremo ulteriormente questi metodi alle casistiche dove non possediamo i dati di alcune variabili. Prima, però, discutiamo come il discorso si cala nel contesto dei sistemi ad equazioni strutturali, in modo da poter fare esempi più quantitativi.

6.2.2 Interventi nei sistemi ad equazioni strutturali

È semplice estendere il concetto di *intervento* introdotto per i grafi causali ai sistemi ad equazioni strutturali. Quello che occorre fare è semplicemente sostituire l'espressione della variabile su cui si interviene (che era prima funzione dei suoi genitori causali) con il valore a cui la fissiamo. Ad esempio, consideriamo nuovamente il grafico della figura 6.8a, e supponiamo di agire sulla variabile X_1 fissandola a x_1. Dal punto di vista grafico, si ottiene la struttura riportata in figura 6.8b. In termini di sistemi ad equazioni strutturali, si ottengono le relazioni:

$$\begin{cases} X_1 = x_1 \\ X_3 = f(x_1, X_2) \end{cases} \tag{6.31}$$

(Si è utilizzato il simbolo $=$ al posto del più appropriato $:=$ per alleggerire la notazione, dato che il contesto rende chiaro che tali relazioni vanno intese in senso adirezionale.) Vediamo adesso come questi sistemi ci permettono di rendere quantitativa l'analisi delle sezioni precedenti. Consideriamo il grafo in figura 6.11, e supponiamo che le relazioni siano lineari, cioè che il sistema di equazioni strutturali sia:

$$\begin{cases} T = \alpha_1 X \\ Y = \alpha_2 T + \beta X \end{cases} \tag{6.32}$$

dove α_1 è il parametro che rappresenta la forza dell'influenza causale di X su T, e analogamente per α_2 e β. Esprimeremo le variabili normalizzandole sulla loro deviazione standard. La struttura causale associata al sistema di equazioni è rappresentata in figura 6.11. In primo luogo, calcoliamo l'associazione statistica tra le due variabili, che contiene sia l'associazione causale dovuta al percorso $T \to Y$ sia quella spuria rappresentata dal percorso anticausale $T \leftarrow X \to Y$, cioè la probabilità $P(y|t)$:

$$\mathbb{E}\left[Y|T = t\right] = \mathbb{E}\left[\beta T + \alpha_2 X\right] = \beta t + \alpha_2 \mathbb{E}\left[X|T = t\right] \tag{6.33}$$

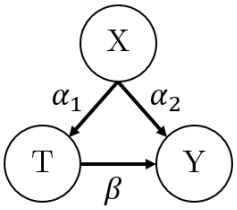


Figura 6.11: Grafo causale associato al sistema di equazioni strutturali $T = f_T(X), Y = f_Y(X, T)$. Si è assunto che il sistema sia lineare, e che le equazioni strutturali si possano scrivere nella forma: $T = \alpha_1 X, Y = \alpha_2 T + \beta X$

Poiché $X = \alpha_1 T$, condizionando su T e studiando la regressione lineare di T su X si ottiene che $\mathbb{E}[X|T = t] = \alpha_1 t$, dove si è usato il fatto che il coefficiente di regressione è simmetrico quando utilizziamo le variabili normalizzate. Di conseguenza:

$$\mathbb{E}[Y|T = t] = \beta t + \alpha_1 \alpha_2 t = (\beta + \alpha_1 \alpha_2)\, t \tag{6.34}$$

In questo contesto, dove il modello è continuo, utilizziamo come indicatore causale la derivata dell'effetto rispetto alla causa (in realtà, in questo caso, l'interpretazione non è causale, infatti stiamo solamente studiando un'associazione statistica). Quindi:

$$\frac{\partial}{\partial t}\mathbb{E}[Y|t] = \beta + \alpha_1 \alpha_2 \tag{6.35}$$

Notiamo che ciò coincide con quanto si otterrebbe utilizzando il metodo dei coefficienti di percorso: la correlazione tra le due variabili è la somma dei prodotti dei coefficienti di percorso lungo i cammini che le legano nel grafo causale.

Adesso studiamo l'effetto causale di T su Y, isolandolo dall'associazione spuria trasmessa da X. Notiamo che X blocca tutti i percorsi *back-door* (l'unico) che trasferiscono informazione spuria tra T e Y. Quindi possiamo utilizzare l'aggiustamento *back-door*:

$$P(y|do(t)) = \mathbb{E}_x \mathbb{E}[\beta T + \alpha_2 X|T = t, X] = \mathbb{E}_X[\beta t + \alpha_2 X] = \beta t + \alpha_2 \mathbb{E}[X] \tag{6.36}$$

dove $\mathbb{E}[X]$ è il valore di aspettazione di X su tutti i valori che può assumere, quindi una costante. Di conseguenza:

$$\frac{\partial}{\partial t}\mathbb{E}P(y|do(t)) = \frac{\partial}{\partial t}\mathbb{E}_x \mathbb{E}[\beta T + \alpha_2 X|T = t, X] = \beta \tag{6.37}$$

che è quanto ci aspettavamo di trovare. Inoltre, notiamo che tale valore differisce da quello trovato precedentemente, ottenuto studiando l'associazione tra le due variabili.

Per ottenere questo risultato utilizzando il metodo dei coefficienti di percorso, avremmo dovuto conoscere almeno 3 correlazioni, in quanto abbiamo 3 coefficienti incogniti. Poiché le variabili sono 3, ciò significa che dobbiamo avere misure di tutte le variabili. Tuttavia, questo non è sempre il caso: potremmo non avere dei dati su alcune variabili del nostro sistema, proprio come Wright nel caso dei porcellini d'india per quanto riguarda il tasso di crescita prenatale. Ad esempio, nel grafo di figura 6.11, se non avessimo misurazioni di X, non potremmo ottenere l'effetto causale di T su Y. Ciononostante, Wright riuscì comunque a risolvere il sistema causale calcolando tutti i coefficienti di percorso. Nel prossimo capitolo vedremo in che modo l'analisi causale ci permette di superare le difficoltà delle valutazioni osservative che coinvolgono sistemi nei quali alcune variabili non possono essere misurate.

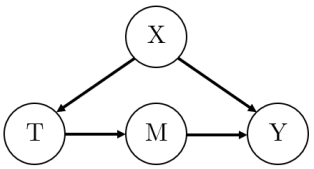


Figura 6.12: Grafo causale rappresentante una variabile T che influenza una variabile Y tramite un mediatore M e una variabile confondente X non osservata che introduce un'associazione spuria tra T e Y.

6.3 Il *do-calculus*

Nella sezione precedente abbiamo mostrato in che modo da alcune correlazioni sia possibile dedurre la forza di un nesso causale. Tuttavia, ciò non sempre è possibile. Quando una domanda causale può essere risposta basandosi semplicemente su dati osservazionali, si dice *identificabile*. Dato che per rispondere a domande causali dobbiamo essere in grado di calcolare la distribuzione di probabilità post-intervento:

> Una domanda causale si dice *identificabile* se la distribuzione di probabilità post-intervento può essere calcolata in termini delle distribuzioni di probabilità osservazionali.

Il problema dell'*identificabilità* è centrale in analisi causale ed è studiato attraverso il *do-calculus*, sviluppato da Pearl negli anni '90. In questa sezione esamineremo tale metodo analitico e ne mostreremo un'applicazione.

6.3.1 Il metodo *front-door*

Consideriamo il grafo causale di figura 6.12, già analizzato nel capitolo precedente. Avevamo dimostrato che possiamo calcolare l'influenza di T su Y condizionando su X:

$$P(y|do(t)) = \sum_x P(y|t,x)P(x) \tag{6.38}$$

Supponiamo adesso di non possedere dati su X. In questo caso, non possiamo utilizzare la formula precedente, perché dipende da X, che non conosciamo. Tuttavia, in questo caso, possiamo calcolare l'effetto causale di T su Y anche se non conosciamo una variabile del grafo.

Il nostro obbiettivo è quello di ottenere un'espressione per $P(y|do(t))$ che non contenga al suo interno X. In modo analogo a quanto fatto nel capitolo precedente, prima calcoliamo l'effetto causale di T su M, poi quello di M su Y. In primo luogo, notiamo che l'unico percorso *back-door* tra M e T è $M \to Y \leftarrow X \leftarrow T$, che è bloccato dal *collider* Y. Quindi:

$$P(m|do(t)) = P(m|t) \tag{6.39}$$

Nel precedente capitolo avevamo condizionato su X per bloccare il percorso *back-door* tra M e Y ($M \leftarrow T \leftarrow X \to Y$). Tuttavia, poiché X ora non è osservata, dobbiamo utilizzare un altro metodo. Notiamo che anche T blocca tale percorso, quindi:

$$P(y|do(m)) = \sum_t P(y|t,m)P(t) \tag{6.40}$$

Mediando su M otteniamo quanto cercato:

$$P(y|do(t)) = \sum_m P(m|do(t))P(y|do(m)) = \sum_m P(m|t) \tag{6.41}$$

In conclusione:

$$P(y|do(t)) = \sum_m P(m|t) \sum_{t'} P(y|t', m)P(t') \tag{6.42}$$

Che ci permette di esprimere l'influenza causale di T su Y esclusivamente in termini di dati osservazionali su variabili note (X non compare l'espressione). Tale aggiustamento è noto come *front-door*. Ogni volta che ci troviamo in una situazione come quella descritta dal grafo in figura 6.12 possiamo utilizzare l'equazione (6.42) per calcolare la probabilità $P(y|do(t))$ in termini di dati osservazionali.

In questo caso, dunque, anche se ci sono delle variabili non osservate, possiamo calcolare l'effetto causale del nostro intervento. Di conseguenza, la domanda causale $P(y|do(t))$, in questo caso, è identificabile. È semplice dimostrare che, quando le variabili sono tutte osservate, gli effetti causali sono sempre identificabili (Shpitser e Judea Pearl (2006)). La questione è più complessa quando ci sono variabili non osservate. Fortunatamente, sono stati introdotti dei metodi analitici che ci permettono di scoprire, in tempi polinomiali, se un effetto causale è identificabile, noto il grafo, e di calcolare la distribuzione di probabilità intervenzionale associata. Tali metodi, sviluppati dalla comunità scientifica negli anni '90 sono una generalizzazione delle formule introdotte nel capitolo precedente, e vanno sotto il nome di *do-calculus* (Judea Pearl (2009b), Spirtes, Glymour e Scheines (2001)).

6.3.2 Le regole del *do-calculus*

Il metodo *front-door* è molto utile perché ci permette di superare il problema della variabile su cui non possediamo dati e di riuscire comunque a calcolare l'effetto di un trattamento sul suo *output*. Tuttavia, stiamo ancora considerando grafi causali molto semplici, dove riusciamo a derivare delle espressioni in termini di osservabili in modo intuitivo. Inoltre, quando alcune variabili non sono note, spesso non si può ottenere una formula per l'effetto causale che dipenda solo da dati osservazionali. In generale, questi calcoli sono molto complessi, e vorremmo possedere un algoritmo che ci permetta di: a) *identificare* se la domanda causale *può* ottenere risposta e b) *calcolare* tale risposta. Il *do-calculus*, introdotto da Pearl in Judea Pearl (1995), consiste in 3 regole alla base di alcuni algoritmi che ci permettono di rispondere a queste domande. Ciò sarà argomento di questa sezione.

Consideriamo degli insiemi di nodi disgiunti (cioè gli insiemi non hanno nodi in comune) X, Y, Z, e W le cui relazioni causali sono rappresentate nel grafo G. Sia $G_{\overline{X}}$ il grafo che si ottiene cancellando da G tutte le frecce che *puntano* a nodi in X. Tale grafo è quello che si ottiene a seguito di un intervento su X, cioè è il grafo post-intervento. Sia $G_{\underline{X}}$ il grafo che si ottiene cancellando da G tutte le frecce che *escono* da nodi in X. Tale grafo elimina qualsiasi effetto causale che possa derivare da X, infatti in esso non ci sono discendenti di X. Ad esempio, per il grafo di figura 6.13a, $G_{\overline{X}}$ è riportato in figura 6.13b, e $G_{\underline{X}}$ in figura 6.13c (notiamo anche che $G_{\underline{X}} = G_{\overline{Z}}$). Le due operazioni possono essere anche effettuate contemporaneamente. Ad esempio, con $G_{\overline{X}\underline{Z}}$ indicheremo il grafo ottenuto eliminando tutte le frecce che puntano a X e tutte quelle che escono da Z, riportato in figura 6.13d. Le regole del *do-calculus*, dimostrate in Judea Pearl (1995), sono le seguenti:

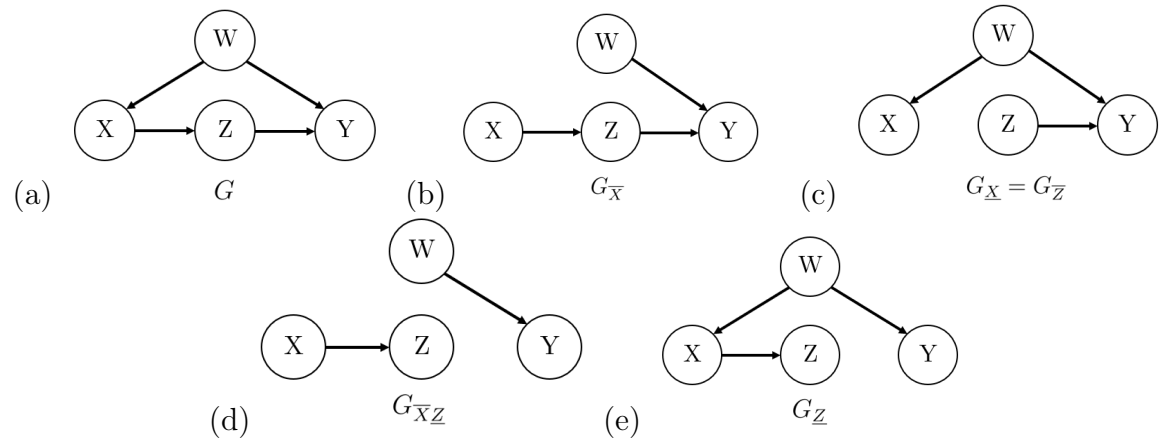


Figura 6.13: Esempi di modifiche ad un grafo G a seguito delle operazioni espresse dalla notazione del *do-calculus*: a) grafo G (non modificato); grafo $G_{\overline{X}}$; c) grafo $G_{\overline{Z}}$ $(= G_{\underline{X}})$.

Regola 1: ignorare una variabile.

$$P(y|do(x), z, w) = P(y|do(x), w) \quad \text{se} \quad Y \perp\!\!\!\perp_{G_{\overline{X}}} Z|X, W \tag{6.43}$$

Regola 2: la correlazione è causalità.

$$P(y|do(x), do(z), w) = P(y|do(x), z, w) \quad \text{se} \quad Y \perp\!\!\!\perp_{G_{\overline{X}\underline{Z}}} Z|X, W \tag{6.44}$$

Regola 3: ignorare un intervento.

$$P(y|do(x), do(z), w) = P(y|do(x), w) \quad \text{se} \quad Y \perp\!\!\!\perp_{G_{\overline{X}\overline{Z(W)}}} Z|X, W \tag{6.45}$$

dove $Z(W)$ è l'insieme dei nodi di Z che non sono antenati di alcun nodo di W in $G_{\overline{X}}$.

Notiamo che tutte le regole contengono un $do(x)$ da entrambi i lati dell'uguaglianza. Questo significa che stiamo studiando il comportamento delle variabili diverse da X nel grafo post-intervento. Per comprendere meglio dal punto di vista intuitivo il significato di queste regole, per adesso studiamo il loro significato non considerando $do(x)$. Tale operazione è legittima perché basta considerare che X sia un insieme vuoto, o aggiungere una variabile ipotetica X che punti alla variabile di intervento, in modo che il grafo non sia modificato.

Regola 1, intuizione. Consideriamo la regola 1 (equazione (6.43)). Senza $do(x)$ questa è:

$$P(y|z, w) = P(y|w) \quad \text{se} \quad Y \perp\!\!\!\perp_G X|W \tag{6.46}$$

Che è semplicemente la relazione di indipendenza statistica sotto condizione di *d-separazione* in un grafo causale, equivalente alla condizione markoviana (teorema 5.2.1). Aggiungendo il $do(x)$ nei condizionamenti, il nuovo grafo a cui dobbiamo riferirci è $G_{\overline{X}}$, e si ottiene nuovamente la regola. Quindi essa significa semplicemente che la proprietà di indipendenza statistica sotto *d-separazione* continua a valere anche nel grafo post-intervento.[2]

Regola 2, intuizione. Eliminando il $do(x)$ dalla regola 2 si ottiene:

$$P(y|do(z), w) = P(y|z, w) \quad \text{se} \quad Y \perp\!\!\!\perp_{G_{\underline{Z}}} Z|W \tag{6.47}$$

[2]È stato dimostrato in Shpitser e Judea Pearl (2006) che, in realtà, tale regola può essere provata assumendo le altre due. Per alleggerire la notazione, tuttavia, è utile considerarla lo stesso.

Notiamo che $G_{\underline{Z}}$ è il grafo da cui sono eliminati tutti gli archi causali che partono da Z (figura 6.13e). Se in tale grafo Y è indipendente da Z quando condizioniamo su W, significa che condizionare su W blocca tutta l'associazione spuria che passa da Z a Y. Di conseguenza, tutta l'associazione tra Z e Y è causale quando condizioniamo su W, e quindi $P(y|do(z), w) = P(y|z, w)$. Quindi la seconda regola del *do-calculus* è semplicemente la generalizzazione del criterio *back-door* a grafi post-intervento.

Regola 3, intuizione. Consideriamo ora la terza regola, eliminando ancora l'operatore $do(x)$:

$$P(y|do(z), w) = P(y|w) \quad \text{se} \quad Y \perp\!\!\!\perp_{G_{\overline{Z(W)}}} Z|W \tag{6.48}$$

L'uguaglianza nell'equazione significa che rimuovere l'operazione $do(z)$ non modifica la distribuzione di probabilità di Y. Dato che $do(z)$ rimuove le frecce che puntano a Z, il grafo associato a $do(z)$ è $G_{\overline{Z}}$. Quindi dobbiamo studiare l'associazione che scorre da Z a Y in $G_{\overline{Z}}$, e studiare sotto quali condizioni questa è la stessa che scorre in G. Poiché tale associazione è causale, ciò significa che non c'è associazione causale tra Z e Y se condizioniamo su W, cioè W blocca tutti i percorsi causali da Z a Y. Ci aspetteremmo quindi che la condizione sia:

$$Y \perp\!\!\!\perp_{G_{\overline{Z}}} Z|W \tag{6.49}$$

Tale equazione, in effetti, vale quando Z è composto da un solo nodo, o da solo strutture a catena o a forchetta. Infatti, potrebbe darsi che Z contiene un *collider* in G, e che W contenga dei loro discendenti. Abbiamo visto che condizionare sul discendente di un *collider* introduce associazione spuria. Quindi nel grafo associato a $do(z)$ ($G_{\overline{Z}}$), potrebbe scorrere dell'associazione spuria. Per avere indipendenza statistica occorre che questa non possa scorrere, quindi, invece del grafo $G_{\overline{Z}}$ occorre utilizzare il grafo $G_{\overline{Z(W)}}$.

È semplice dimostrare che sia il metodo *back-door* che il metodo *front-door* sono derivabili da tali regole (Judea Pearl (2009b), Neal (2020)). In generale, è stato dimostrato in modo indipendente da Huang e Valtorta (2012) e da Shpitser e Judea Pearl (2006)) che il *do-calculus* è *completo* per l'identificazione degli effetti causali. Questo significa che qualsiasi distribuzione di probabilità post-intervento, se è identificabile, può essere ottenuta tramite applicazione ripetuta delle regole del *do-calculus*. Tale prova è anche costruttiva, infatti si basa su un algoritmo dimostrato completo per l'identificazione e il calcolo di distribuzioni di probabilità post-intervento (Huang e Valtorta (2006)). Nel *paper* si dimostra che tale algoritmo è riconducibile ad una applicazione delle 3 regole del *do-calculus*, e che, quindi, un'applicazione ripetuta di tali regole permette di rispondere a qualsiasi domanda causale.

Capitolo 7

Conclusione

È ormai nozione comune che correlazione e causalità siano concetti distinti. La prima è semplicemente una quantificazione dell'interdipendenza tra alcune variabili, e può sempre essere calcolata in modo oggettivo. La seconda, in qualche modo, sembra portare con sé delle assunzioni soggettive sul rapporto tra due eventi. Durante il secolo scorso, coltivando l'ideale neopositivistico di una scienza esente da qualsiasi soggettivismo, la comunità statistica decise così di abbandonare la nozione di causalità *in quanto tale*, escludendola dalla pratica scientifica.

Tuttavia, il concetto di causa è strettamente legato a quello di correlazione: se un evento ne causa un altro, allora i due saranno anche correlati, ma spesso non vale il contrario. Ma soprattutto, la nozione di causalità è spesso decisiva per la nostra comprensione dei fenomeni, e il suo abbandono dogmatico ci preclude l'utilizzo di una parte determinante del nostro sapere. In questo elaborato abbiamo cercato di comprendere il sottile legame che sussiste tra le due nozioni, ripercorrendo il modo in cui la comunità scientifica ha prima abbandonato e poi faticosamente ricostruito la nozione di causalità, facendo chiarezza sul perché questa sia diversa da quella di correlazione e in che modo tale differenza si radichi nella descrizione fisica della realtà. Questo ci ha permesso di introdurre la formulazione contemporanea dei metodi di analisi causale, mostrando come questi siano profondamente radicati nella rappresentazione termodinamica del reale.

Nel capitolo 2 abbiamo discusso in che modo la statistica si è resa conto della differenza tra i concetti di correlazione e causalità. In primo luogo, abbiamo visto che un indice di correlazione non è sufficiente per distinguere una causa dal suo effetto. Inoltre, gli stessi indici di correlazione non permettono di distinguere tra associazioni con significato causale e associazioni di natura meramente correlativa. Dato che la correlazione è un concetto quantificabile matematicamente, quindi oggettivabile, la comunità statistica abbandonò progressivamente la nozione di causalità. Tuttavia, in questo modo, dovette rinunciare ad una componente determinante del nostro sapere, quella, appunto, causale. Il primo studioso che cercò di reintrodurre la conoscenza causale nell'analisi dei sistemi statistici fu il genetista Sewall Wright, con il metodo dei coefficienti di percorso. Wright propose un formalismo in grado di esprimere delle assunzioni causali su un certo sistema, e di utilizzare tale assunzioni per interpretare delle dinamiche insondabili alla semplice analisi statistica. Tuttavia, i metodi di Wright erano ancora limitati. Da un punto di vista tecnico, questi permettevano di studiare esclusivamente sistemi lineari. Da un punto di vista epistemologico, erano ancora definiti in modo troppo vago per essere davvero utili alla pratica scientifica. Nei capitoli successivi abbiamo cercato di comprendere in che modo que-

sti possano essere giustificati e migliorati e, soprattutto, in che modo questi si radichino nella fisica.

Nei capitoli 3 e 4 abbiamo cercato di legittimare, da un punto di vista fisico, la nozione di causa. Nel capitolo 3 ci siamo chiesti in che modo una nozione asimmetrica come quella della causalità possa fondarsi nella fisica, dato che tutte le leggi fondamentali che regolano i fenomeni del nostro mondo sono simmetriche. Abbiamo argomentato che, in realtà, i contesti fenomenologici in cui parliamo di causalità sono sempre macroscopici, e dunque, per analizzarli, occorre adottare una prospettiva termodinamica. Il secondo principio della termodinamica, in effetti, fa emergere un'asimmetria tra passato e futuro che è a fondamento della nostra nozione di causa.

Nel capitolo 4 abbiamo discusso in che modo un sistema statistico può essere studiato in senso dinamico, seguendo le analisi del filosofo tedesco Hans Reichenbach. Abbiamo mostrato che, seguendo un simile approccio, è possibile tradurre proprietà dinamiche di variabili macroscopiche in proprietà statistiche, e spiegare fenomeni come le associazioni spurie in termini causali. In questo modo abbiamo creato un ponte tra la descrizione dinamica di un sistema e le relazioni statistiche che ne discendono per le variabili che lo compongono. Abbiamo visto che possiamo anche fare il contrario: dato un sistema di dipendenze statistiche, possiamo introdurre una struttura causale che sia in grado di darne ragione modellizzandolo in senso dinamico. Tuttavia, abbiamo evidenziato che questo è più problematico, infatti, generalmente, ci sono più spiegazioni dinamiche possibili per gli stessi fenomeni statistici.

Nel capitolo 5 abbiamo discusso in quali casi possiamo avere più spiegazioni causali possibili per gli le stesse osservazioni statistiche. Ciò ci ha condotti al tema dell'inferenza causale, e abbiamo discusso le difficoltà che dobbiamo affrontare quando cerchiamo di dedurre un grafo causale a partire da dati osservazionali. Successivamente, abbiamo accennato a come ciò si declina, nella pratica fisica, nello studio delle disuguaglianze di Bell.

Infine, nel capitolo 6, abbiamo trascurato i problemi dell'inferenza causale, supponendo che il grafo causale sia noto. Abbiamo mostrato che, sotto tale assunzione, è possibile definire e distinguere con successo associazioni spurie e associazioni causali, e calcolare l'effetto causale di una variabile su un'altra. Abbiamo mostrato che per esprimere domande causali è conveniente introdurre un operatore chiamato *do-operator*, il quale *simula* l'intervento su un sistema, in modo analogo a ciò che avviene nella pratica sperimentale delle scienze naturali. Tale operatore permette l'introduzione di un metodo analitico, noto come *do-calculus*, che permette di rispondere a domande causali su sistemi statistici alla luce delle assunzioni che abbiamo fatto su di essi. Tali metodi sono del tutto generali, e non dipendono dalle forme funzionali delle relazioni tra le variabili. Di conseguenza, può essere intesa come una generalizzazione dell'analisi di percorso di Wright.

In conclusione, abbiamo mostrato come la riconduzione dell'analisi causale ai problemi e ai metodi della fisica ne permette una trattazione più rigorosa ed esaustiva di quella generalmente fornita nella letteratura sul tema. Infatti, si è mostrato che molte assunzioni che vengono generalmente presentate in modo dottrinario sono, in realtà, riflesso della descrizione del reale offerta dalla fisica classica. Più precisamente, si è visto che queste nozioni emergono nel limite termodinamico, cioè quando le variabili che trattiamo sono macroscopiche. Nelle *soft sciences*, tutte le variabili hanno tale caratteristica, e ciò spiega il successo dell'utilizzo della nozione di causa in tali contesti. Restano, tuttavia, questioni aperte. Su tutte, si reputa che la più intrigante, e forse la più urgente, riguarda la conciliazione tra analisi causale e indeterminismo quantistico.

Bibliografia

Allen, John-Mark A et al. (2017). «Quantum common causes and quantum causal models». In: *Physical Review X* 7(3), p. 031021.

Anderson, Philip W (1972). «More Is Different: Broken symmetry and the nature of the hierarchical structure of science.» In: *Science* 177(4047), pp. 393–396.

Aristotele (circa 350 a.C.). *Metafisica*. trad. vari.

Aurell, Erik e Gino Del Ferraro (2016). «Causal analysis, correlation-response, and dynamic cavity». In: *Journal of Physics: Conference Series*. Vol. 699. 1. IOP Publishing, p. 012002.

Baldovin, Marco, Fabio Cecconi, Antonello Provenzale et al. (2022). «Extracting causation from millennial-scale climate fluctuations in the last 800 kyr». In: *Scientific Reports* 12(1), p. 15320.

Baldovin, Marco, Fabio Cecconi e Angelo Vulpiani (2020). «Understanding causation via correlations and linear response theory». In: *Physical Review Research* 2(4), p. 043436.

Beebee, Helen et al. (2009). *The Oxford handbook of causation*. Oxford Handbooks.

Boltzmann, Ludwig (1896-1898). *Vorlesungen über Gastheorie*. Traduzione italiana: "Lezioni sulla teoria dei gas". La citazione: "Il concetto di entropia apre uno sguardo più profondo sulla natura rispetto a tutte le altre grandezze fisiche." Johann Ambrosius Barth: Leipzig.

Cartwright, Nancy (1983). *How the laws of physics lie*. OUP Oxford.

Cartwright, Nancy (2007). *Hunting causes and using them: Approaches in philosophy and economics*. Cambridge University Press.

Cartwright, Nancy (2012). «Causal laws and effective strategies». In: *Arguing About Science*. Routledge, pp. 466–479.

Cavalcanti, Eric G e Raymond Lal (2014). «On modifications of Reichenbach's principle of common cause in light of Bell's theorem». In: *Journal of Physics A: Mathematical and Theoretical* 47(42), p. 424018.

Costa, Fabio e Sally Shrapnel (2016). «Quantum causal modelling». In: *New Journal of Physics* 18(6), p. 063032.

Feynman, Richard (1967). «The character of physical law (1965)». In: *Cox and Wyman Ltd., London*.

Fornasini, Paolo et al. (2008). *The uncertainty in physical measurements: an introduction to data analysis in the physics laboratory*. Vol. 995. Springer.

Frenkel, Daan (2014). «Why colloidal systems can be described by statistical mechanics: some not very original comments on the Gibbs paradox». In: *Molecular Physics* 112(17), pp. 2325–2329.

Frisch, Mathias (2014). *Causal reasoning in physics*. Cambridge University Press.

Fritz, Tobias (2016). «Beyond Bell's theorem II: Scenarios with arbitrary causal structure». In: *Communications in Mathematical Physics* 341, pp. 391–434.

Galton, F (1889). «Natural Inheritance Macmillan & Co». In: *London UK*.

Hitchcock, Christopher (1997). «Probabilistic causation». In.

Hitchcock, Christopher e Miklós Rédei (2020). «Reichenbach's common cause principle». In.
Huang, Yimin e Marco Valtorta (2006). «A Study Of Identifiability In Causal Bayesian Networks1 Version 0.3». In.
Huang, Yimin e Marco Valtorta (2008). «On the completeness of an identifiability algorithm for semi-Markovian models». In: *Annals of Mathematics and Artificial Intelligence* 54, pp. 363–408.
Huang, Yimin e Marco Valtorta (2012). «Pearl's calculus of intervention is complete». In: *arXiv preprint arXiv:1206.6831*.
Hume, David (1896). *A treatise of human nature*. Clarendon Press.
Hume, David (2007). «An enquiry concerning human understanding and other writings». In.
Imbens, Guido W. e Donald B. Rubin (2015). *Causal Inference in Statistics, Social, and Biomedical Sciences: An Introduction*. Cambridge University Press.
Ismael, Jenann (2023). «Reflections on the asymmetry of causation». In: *Interface Focus* 13(3), p. 20220081.
Koller, Daphne e Nir Friedman (2009). *Probabilistic graphical models: principles and techniques*. MIT press.
Laplace, Pierre Simon marquis de (1840). *Essai philosophique sur les probabilités*. Bachelier.
Lebowitz, Joel L (2007). «From time-symmetric microscopic dynamics to time-asymmetric macroscopic behavior: An overview». In: *Boltzmann's legacy*, pp. 63–88.
Mach, Ernst (1976). *Knowledge and Error: Sketches on the Psychology of Enquiry*. Trad. da Thomas J. McCormack e Paul Foulkes. D. Reidel Publishing Company: Dordrecht.
Mach, Ernst (1986). *The Principles of the Theory of Heat: Historically and Critically Elucidated*. Trad. da Thomas J. McCormack. D. Reidel Publishing Company: Dordrecht.
Minot, Charles Sedgwick (1891). «Senescence and rejuvenation». In: *The Journal of Physiology* 12(2), p. 97.
Morgan, Stephen L. e Christopher Winship (2007). *Counterfactuals and Causal Inference: Methods and Principles for Social Research*. Cambridge University Press.
Neal, Brady (2020). «Introduction to causal inference». In: *Course Lecture Notes (draft)*.
Neapolitan, Richard E et al. (2004). *Learning bayesian networks*. Vol. 38. Pearson Prentice Hall Upper Saddle River.
Niles, Henry E (1922). «Correlation, causation and Wright's theory of" path coefficients"». In: *Genetics* 7(3), p. 258.
Pearl, J. e D. Mackenzie (2018). *The Book of Why: The New Science of Cause and Effect*. Basic Books. ISBN: 9780465097616. URL: https://books.google.fr/books?id=BzMODwAAQBAJ.
Pearl, Judea (1988). *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Morgan kaufmann.
Pearl, Judea (1995). «Causal diagrams for empirical research». In: *Biometrika* 82(4), pp. 669–688.
Pearl, Judea (2009a). «Causal inference in statistics: An overview». In.
Pearl, Judea (2009b). *Causality*. Cambridge university press.
Pearl, Judea (2012). «The do-calculus revisited». In: *arXiv preprint arXiv:1210.4852*.
Pearson, Karl (1900). *The Grammar of Science*. 2nd. Adam e Charles Black: London.
Pienaar, Jacques e Časlav Brukner (2015). «A graph-separation theorem for quantum causal models». In: *New Journal of Physics* 17(7), p. 073020.
Platone (circa 360 a.C.). *Timeo*. trad. vari.
Provine, William B (1989). *Sewall Wright and evolutionary biology*. University of Chicago press.
Reichenbach, Hans (1991). *The direction of time*. Vol. 65. Univ of California Press.
Rovelli, Carlo (2022a). «Back to Reichenbach». In.

Rovelli, Carlo (2022b). «Memory and entropy». In: *Entropy* 24(8), p. 1022.

Rovelli, Carlo (2023). «How oriented causation is rooted into thermodynamics». In: *Philosophy of Physics* 1(1).

Russell, Bertrand (1912). «On the notion of cause». In: *Proceedings of the Aristotelian society*. Vol. 13. JSTOR, pp. 1–26.

Shpitser, Ilya e Judea Pearl (2006). «Identification of joint interventional distributions in recursive semi-Markovian causal models». In: *AAAI*, pp. 1219–1226.

Smirnov, Dmitry A (2022). «Generative formalism of causality quantifiers for processes». In: *Physical Review E* 105(3), p. 034209.

Spirtes, Peter, Clark Glymour e Richard Scheines (2001). *Causation, prediction, and search*. MIT press.

Tian, Jin e Judea Pearl (2012). «On the testable implications of causal models with hidden variables». In: *arXiv preprint arXiv:1301.0608*.

Virgilio (2009). *Georgiche*. A cura di Vittorio Sermonti. "Felix qui potuit rerum cognoscere causas." Bur Rizzoli: Milano. Cap. 2.

Wood, Christopher J e Robert W Spekkens (2015). «The lesson of causal discovery algorithms for quantum correlations: causal explanations of Bell-inequality violations require fine-tuning». In: *New Journal of Physics* 17(3), p. 033002.

Woodward, James (2001). «Causation and manipulability». In.

Woodward, James (2005). *Making things happen: A theory of causal explanation*. Oxford university press.

Woodward, James (2007). «Causation with a human face». In.

Woodward, James (2014). «A functional account of causation; or, a defense of the legitimacy of causal thinking by reference to the only standard that matters—usefulness (as opposed to metaphysics or agreement with intuitive judgment)». In: *Philosophy of Science* 81(5), pp. 691–713.

Wright, Sewall (1918). «On the nature of size factors». In: *Genetics* 3(4), p. 367.

Wright, Sewall (1920). «The relative importance of heredity and environment in determining the piebald pattern of guinea-pigs». In: *Proceedings of the National Academy of Sciences* 6(6), pp. 320–332.

Wright, Sewall (1921). «Correlation and causation». In: *Journal of agricultural research* 20(7), p. 557.

Wright, Sewall (1923). «The theory of path coefficients a reply to Niles's criticism». In: *Genetics* 8(3), p. 239.

Wright, Sewall (1934). «The method of path coefficients». In: *The annals of mathematical statistics* 5(3), pp. 161–215.