1.1 Ragionare con la probabilità e la statistica

1.2 Una domanda per iniziare

Immagina di osservare per un certo tempo una persona che lancia una moneta, annotando ogni volta se esce testa o croce.

Dopo un numero piuttosto alto di tiri, ecco il risultato.

A quel punto ti chiedono di scommettere, puntando una certa somma su testa o croce.

Cosa fai?

In realtà non c'è una sola risposta esatta. Ma la scelta dice molto di te:

- Se hai puntato su "testa", con l'idea che, essendo in ritardo, la sua probabilità è aumentata, vuol dire che non pensi che si tratti di una successione di eventi indipendenti! È la tipica fallacia del giocatore d'azzardo, per cui ti sconsiglierei di frequentare i casinò.
- Se punti indifferentemente su "testa" o "croce", pensi come i matematici esperti di probabilità, in gergo i "probabilisti". Visto che una moneta ha due facce, è facile dedurre che testa e croce hanno entrambe il 50 di probabilità di uscire. Una probabilità che non viene influenzata dai lanci precedenti.
- Se hai puntato su "croce", invece, pensi come gli statistici. Visto che "croce" esce così spesso, vuol dire che la moneta è truccata. E quindi, continuerà a uscire più spesso "croce"!

1.3 Probabilità e statistica

Forse non ti sorprenderà trovare abbinate statistica e calcolo delle probabilità. Molti tendono a confonderle, perché sono entrambe discipline matematiche di importanza cruciale non solo per il pensiero scientifico, ma anche per la vita di tutti i giorni.

Aiutano a farci un'opinione e a prendere decisioni "informate" in molti campi.

Per cominciare a capire in che cosa consistono, diamo uno sguardo d'insieme alle relazioni che legano statistica e calcolo delle probabilità.

E soprattutto alle loro differenze!

1.4 La probabilità

Noi - noi personalmente o un sistema "intelligente" a cui ricorriamo per ottenerne una previsione o un consiglio - interagiamo continuamente col mondo.

E abbiamo bisogno di farcene un'idea, di sapere cos'è quello che vediamo o sentiamo e come si comporterà.

Anche la scienza, vuole certezze. Anzi, le vorrebbe, perché è difficile provare che qualcosa è assolutamente vero.

La scienza conosce leggi deterministiche, come quelle della meccanica classica formulate da Newton. Ma sono valide solo a una certa scala di osservazione e con qualche limitazione.

In pratica, è impossibile predire l'esito del lancio di una moneta: sono troppe le variabili in gioco e, anche se volessimo misurarle tutte, gli errori di misura sono sempre in agguato.

Ed è impossibile, questa volta teoricamente impossibile, individuare la posizione di un elettrone in un determinato momento.

È in casi come questo che interviene la nozione di probabilità, che ci aiuta a ragionare su processi che sono intrinsecamente casuali o che consideriamo tali perché non conosciamo bene le leggi a cui ubbidiscono.

Il calcolo della probabilità ci aiuta a capire quanto è verosimile un certo evento o qual è la spiegazione più ragionevole di un certo fatto.

Così, anche la scienza fa ricorso a parole come "attesa", "caso", "possibilità", e formula leggi probabilistiche, come quelle che governano il lancio di una moneta, la genetica, la fisica delle particelle.

Sono tutti campi in cui gli eventi osservabili non sono una pura conseguenza logica della teoria.

Allo stesso modo, nella vita di tutti i giorni ci affidiamo, per esempio, a alcoli probabilistici per prevedere chi vincerà il campionato, se l'indice di borsa domani crescerà o calerà, quale sarà il tempo…

1.5 Qualche definizione di probabilità

Ecco una definizione relativamente semplice, anche se un po' ambigua, di probabilità:

"La probabilità di un evento è quante volte ci si aspetta, nel lungo termine, che tale evento si presenti come risultato di un processo casuale, rispetto a tutti gli altri possibili risultati.

Da questa definizione ricaviamo alcuni punti fermi, che applichiamo al lancio di un dado.

Innanzitutto, la probabilità di un evento, per esempio, l'uscita del numero "cinque" è un rapporto: al numeratore il numero di volte in cui ci si aspetta che esca "cinque", al denominatore il numero di volte in cui esce un numero qualunque. È facile verificare che, se il dado non è truccato, il "cinque" esce una volta su sei….

Come vedi, la probabilità può essere espressa con una frazione, ma possiamo usare anche un numero compreso tra 0 e 1…

o una percentuale.

Ritornando alla definizione di probabilità, quello che può lasciare perplessi è l'espressione "nel lungo termine" …

Che significa?

Significa che perché si verifichi l'evento atteso, nel nostro caso il "cinque" del dado, con la frequenza prevista, i lanci devono essere molti.

Quanti?

Qualcuno risolve il problema sostituendo "nel lungo termine" con "in un numero infinito di esperimenti".

Ma così diventa impossibile determinare la probabilità in modo sperimentale.

Una definizione alternativa di probabilità, che piacerebbe a un probabilista teorico, è questa:

"La probabilità di un evento è il rapporto fra il numero dei casi favorevoli e il numero dei casi possibili, supposti tutti ugualmente possibili."

Anche questa definizione, però, non ci soddisfa, perché di solito i possibili eventi non sono ugualmente possibili. A meno che non si tratti di domini artificiali, come quelli del gioco d'azzardo.

D'altra parte, una definizione diretta di probabilità non è necessaria. Possiamo usare la definizione indiretta che si sviluppa a partire da pochi assiomi…

1.6 Due assiomi

Partendo da questo esempio, siamo in grado di introdurre i primi due assiomi di base del calcolo delle probabilità.

Il primo, detto anche "assioma della positività", ci dice che la probabilità di un generico evento E è un numero reale maggiore o uguale a 0. Si esprime così…

Da notare che se la probabilità di un evento è zero, siamo di fronte a una certezza, perché l'evento non si verificherà mai.

Un'altra certezza è descritta del secondo assioma, chiamato appunto "assioma della certezza":

la probabilità di un evento certo è 1.

Questo assioma si trova anche formulato così: la somma delle probabilità di tutti i possibili eventi tra loro alternativi, cioè dell'intero spazio campione, simboleggiata dalla lettera greca omega, è 1.

1.7 Qualche definizione

Prima di proseguire con il terzo assioma, fermiamoci a definire, o a definire meglio, alcuni termini:

- Processo casuale: è un fenomeno naturale o un esperimento di cui possiamo osservare una serie di risultati che non siamo in grado di prevedere.
- Evento: è un risultato all'interno di un processo casuale. Può essere un
  - evento semplice, quando il risultato è uno solo (come nel lancio di un dado); o un
  - evento composto, quando comprende più risultati (come nel lancio di due o più dadi, contemporaneamente o in sequenza).
- Spazio campione (o spazio campionario): è l'insieme di tutti i possibili eventi in un processo casuale, come i sei possibili esiti del lancio di un dado.
- Eventi incompatibili: eventi che non possono occorrere contemporaneamente. Per esempio, se lanciamo un solo dado non possono uscire contemporaneamente il 3 e il 4.

Visto che il concetto di eventi incompatibili è essenziale per introdurre il terzo assioma della probabilità, prima di proseguire rispondi a questa domanda…!

1.8 Il terzo assioma

Per introdurre il terzo assioma, torniamo a una partita a dadi.

Se il nostro avversario ha tirato ottenendo un "quattro", quante probabilità abbiamo di vincere con un punteggio superiore?

Il terzo assioma, detto anche assioma dell'unione ci dice che:

se A e B sono eventi incompatibili, la probabilità che si verifichi uno dei due (o, in termini matematici, la probabilità dell'unione dei due eventi) è uguale alla somma della probabilità dei singoli eventi.

Ecco come si rappresenta formalmente questo assioma …

Qui la prima parte dell'uguaglianza si legge "A unione B".

Nel nostro caso, la probabilità di vincere è data dalla probabilità di tirare un "cinque" più la probabilità di tirare un "sei", …

…cioè una probabilità su tre!

Ma se i dadi sono due, il calcolo si complica…

1.9 La probabilità composta

Ritorniamo ancora una volta alla nostra partita a dadi, questa volta però i dadi sono due e il nostro avversario, piuttosto fortunato, ha tirato un “cinque” e un “sei”.

In tutto 11.

Che probabilità abbiamo di superarlo con un 12?

Estendendo il ragionamento fatto con un solo dado, potremmo pensare che, essendo 12 i possibili punteggi totali, la probabilità di ottenere proprio 12 sia, appunto, un dodicesimo.

È la classica risposta istintiva che però, come accade spesso, è sbagliata e di molto!

Infatti, i risultati dei due dadi sono eventi indipendenti, perché il risultato del primo non influenza il secondo. Sarebbe lo stesso se si lanciasse lo stesso dado due volte di seguito.

Per conoscere la probabilità che si verifichino insieme due eventi indipendenti, nel nostro caso “sei” sul primo dado e “sei” sul secondo, usiamo il “teorema della probabilità composta”:

la probabilità di un evento composto è uguale al prodotto delle probabilità degli eventi componenti.

La prima parte della formula si legge “A intersezione B”.

E ora applichiamola al nostro caso…

Come vedi, la probabilità di vincere la partita è una su 36 non una su 12.

È una bella differenza, che ci fa capire quanto è importante, per ragionare correttamente, saper calcolare le probabilità di due eventi indipendenti.

Ma cosa succede se, invece, i due eventi non sono indipendenti?

Ma le complicazioni del discorso non finiscono qui…

1.10 La probabilità condizionata

È capitato a tutti di assistere all’estrazione di una lotteria di paese, di quelle in cui i biglietti hanno una serie e un numero.

Sappiamo che sono stati vendute venti serie complete, per un totale di 2.000 biglietti.

Qual è la nostra probabilità di vincere il premio, avendo acquistato un biglietto?

Semplice: una su 2.000!

Senza sperarci troppo, eccoci al momento dell’estrazione.

Dal palco, qualcuno estrae un biglietto dall’urna, lo apre e inizia a leggere: “Serie D, numero…”

Ecco: fermiamoci un attimo. A quel punto, sapendo che la nostra serie è proprio “D”, percepiamo immediatamente che la probabilità di vincita è aumentata enormemente. E la nostra attenzione, giustamente, si risveglia.

Infatti, lo spazio campionario non è più di 2.000 eventi, ma solo di 100!

In casi come questo, l’evento A (l’uscita del nostro biglietto) è legato a un altro evento B (l’uscita della nostra serie).

La probabilità condizionata di A rispetto a B è la probabilità che ha l'evento A se l'evento B è certo.

E la scriviamo così…

Il teorema della probabilità condizionata, che deriva dai tre assiomi di base del calcolo delle probabilità, ci dice che la probabilità condizionata di un evento A rispetto a un evento B è uguale al rapporto tra la probabilità composta dei due eventi e la probabilità assoluta di B…

Questo teorema ha due limitazioni:

· si può applicare solo se l'evento B è possibile (cioè la sua probabilità è maggiore di zero);

· è utile solo se i due eventi sono dipendenti tra loro (altrimenti la probabilità condizionata coincide con quella assoluta).

1.11 La statistica

La scienza difficilmente può fare a meno della matematica.

Ma non è sempre la stessa matematica. Per trattare gli aspetti e i modelli della realtà, in molti campi si usa la statistica che, come il calcolo della probabilità, punta a ottenere, se non la certezza assoluta, almeno una ragionevole riduzione dell'incertezza.

La statistica è un ramo della matematica che si è sviluppato per tentare di interpretare i dati, cercarvi regolarità e tendenze, formulare previsioni.

Si usa in campi applicativi disparati, come:

- la genetica e la demografia;
- le scienze ambientali e l'analisi delle risorse;
- l'econometria e le indagini di mercato;
- l'epidemiologia e la sperimentazione di nuovi farmaci;
- la sociologia e la psicologia sperimentale;
- la linguistica computazionale;
- la fisica delle particelle subatomiche.

1.12 Una storia di vedove e orfani

La nascita della statistica moderna risale al XVIII secolo, sotto la spinta di diverse esigenze, tra cui la nascita delle assicurazioni sulla vita.

L'idea viene attribuita a Robert Wallace e Alexander Webster, due ministri della chiesa di Scozia, appassionati di matematica. In anni in cui la morte del capofamiglia significava miseria, erano preoccupati per la sorte delle vedove e degli orfani dei loro colleghi, cui la Riforma protestante aveva permesso di sposarsi.

Nel 1744, Wallace e Webster misero a punto e fecero approvare per legge un fondo di previdenza, alimentato, come oggi, da un versamento periodico.

In questi casi, il problema da risolvere è: a quanto devono ammontare i versamenti per garantire un reddito adeguato alle famiglie?

Per capirlo, raccolsero una gran mole di dati sulle aspettative di vita dei ministri del culto e delle loro vedove, sul numero di figli, sulle rendite degli investimenti e via dicendo.

In questo lavoro si servirono delle tavole attuariali, pubblicate mezzo secolo prima dal matematico inglese Edmond Halley, della consulenza di un professore dell'Università di Edimburgo e dei progressi del calcolo delle probabilità, tra i quali la "legge dei grandi numeri" del matematico svizzero Jakob Bernouilli.

Ne risultarono strumenti nuovi, come le "tavole di mortalità" e una previsione: un fondo pensioni sostenibile richiedeva almeno 10 sterline l'anno da ogni ministro del culto.

Stimarono, inoltre, che nel 1765 il fondo avrebbe accumulato un capitale di 58.438 sterline. Una stima sorprendentemente precisa: nel 1765 il capitale raggiunse proprio quella cifra, con l'approssimazione di una sterlina!

1.13 La legge dei grandi numeri

Quello che permise (e permette ancora) la sostenibilità dei fondi pensione è un'osservazione piuttosto diffusa: quando si ha a che fare con fenomeni casuali ripetitivi (come il lancio di un dado o la morte), la frequenza relativa di un evento tende a stabilizzarsi all'aumentare del numero delle prove.

Tirando una volta un dado, può uscire un numero qualunque, ma tirandolo moltissime volte, alla fine ogni numero avrà una frequenza relativa prossima a 1/6.

Questa è la "legge empirica del caso", che, ai primi del '700, Jakob Bernouilli riformulò, ribattezzandola come "legge dei grandi numeri", nei seguenti termini:

"Se P è probabilità di un evento E, allora la frequenza relativa dei successi in N prove indipendenti tende a P, quando N tende a infinito."

La legge dei grandi numeri (o teorema di Bernoulli) non è di nessuna utilità per prevedere il successo volta per volta: chi ci prova (per esempio, puntando sui numeri ritardati del lotto) lo fa a suo rischio e pericolo!

Ma questa legge ci permette di stimare con precisione tutti i fenomeni casuali, mettendo chiaramente in relazione tra loro due concetti:

- la probabilità teorica di un evento, concetto centrale della teoria della probabilità;
- la sua frequenza relativa, concetto centrale della statistica.

1.14 Statistica descrittiva e inferenziale

Nella statistica si usa distinguere due filoni principali, che hanno finalità diverse:

- La statistica descrittiva, raccoglie i dati e li elabora per fornire una "fotografia" delle caratteristiche significative di un insieme di “unità statistiche” (persone, animali, piante, oggetti) chiamato “popolazione”.

• La statistica inferenziale, utilizza i dati per ricavare induttivamente regole generali e, quindi, fare previsioni su situazioni incerte.

Com’è facile intuire, la statistica descrittiva viene “prima”, cioè è la base per qualunque operazione di statistica inferenziale.

Ma in cosa consiste l’elaborazione della statistica descrittiva?

In primo luogo, si specifica quali sono le grandezze che ci interessano. Per esempio, il sesso e l’età delle persone, l’altezza delle piante, il valore ottenuto tirando i dadi, il colore dei fiori.

Qui dobbiamo fare attenzione a scegliere grandezze osservabili e, in qualche modo, misurabili.

Possono essere grandezze:

• qualitative, come il sesso e il colore;

• o quantitative.

Queste ultime si distinguono in:

o grandezze “continue”, come l’altezza e l’età, che possono assumere qualunque valore intermedio entro certi limiti;

o grandezze “discrete”, che non consentono valori intermedi; è il caso del lancio dei dadi, che può valere “cinque” o “sei”, ma non “cinque e mezzo”.

È evidente che le operazioni possibili nei diversi casi non sono le stesse…

1.15 Statistica descrittiva: la distribuzione di frequenza

L’operazione più elementare consiste nel contare il numero di elementi della popolazione che, per una certa grandezza, presentano determinati valori.

Si può disegnare così un istogramma.

Per esempio, nel caso di una grandezza qualitativa come il sesso, possiamo contare i maschi e le femmine.

Possiamo fare la stessa operazione con una grandezza quantitativa discreta, come il valore che esce ripetendo mille volte il lancio di due dadi.

Se la grandezza quantitativa è continua, dobbiamo contare gli elementi facendo riferimento non a specifici valori, ma a intervalli, in modo da tener conto di tutte le possibilità intermedie.

Ecco, per esempio, la distribuzione di frequenza dell’altezza di 500 alunni delle scuole superiori…

In questa distribuzione puoi notare un aspetto interessante: collegando gli istogrammi con una linea, si vede una forma “a campana”, simile alla nota distribuzione di Gauss, che viene seguita da molti eventi che hanno valori continui.

Una distribuzione di frequenza ci dà molte informazioni utili su una popolazione. Ma nella maggior parte dei casi, è ancora più utile fotografare la situazione in forma sintetica.

Per esempio con la media…

1.16 Statistica descrittiva: media, mediana e moda

Sappiamo tutti come si calcola una media, ma… quali informazioni possiamo ricavarne?

Proviamo a fare piccolo test.

Come vedi, la media ci dà un'idea di uno stato di cose, ma non dice tutto. Ci servono altri calcoli. In primo luogo quelli della cosiddetta "tendenza centrale" di una certa proprietà.

Le misure della tendenza centrale, che si applicano solo a dati quantitativi, sono in tutto tre:

• La media aritmetica è, come sappiamo, il rapporto tra la somma dei valori della proprietà considerata e il numero di unità statistiche.

In questo caso, 110 bambini diviso 50 famiglie.

• La mediana è quel valore che divide in due la distribuzione. Metà della popolazione ha valore inferiore, l'altra metà della popolazione un valore superiore.

• La moda è il valore che ha la frequenza maggiore.

Nel caso del nostro test sui bambini del villaggio, non possiamo rispondere "A" o "E", perché non conosciamo, rispettivamente, la mediana e la moda, che potremmo ottenere dalla distribuzione di frequenza.

Ma a questo punto sembra di sentire i versi di un poeta…

1.17 Statistica descrittiva: la varianza

"Me spiego: da li conti che se fanno

seconno le statistiche d'adesso

risurta che te tocca un pollo all'anno:

e, se nun entra nelle spese tue,

t'entra ne la statistica lo stesso

perch'è c'è un antro che ne magna due."

Trilussa, in questi famosi versi, ci sta dicendo che, per capire un fenomeno, la media e le altre misure della tendenza centrale sono insufficienti.

La stessa media può riferirsi a una situazione di sostanziale omogeneità (per esempio, tutte le famiglie hanno 2 o 3 figli) o di fortissima sperequazione (come nel caso paradossale di una sola famiglia che avesse 102 figli)!

Per questo, le notizie che, soprattutto in campo economico, partono con la media sparata in prima pagina, senza aggiungere altro, vanno prese con le molle.

Ci direbbe di più una distribuzione di frequenza, con tanto di istogrammi, ma potrebbe essere di difficile lettura.

Esistono, però, alcuni indici che ci danno velocemente un'idea della dispersione dei valori. Tra queste, la varianza che è la media dei quadrati delle deviazioni dei singoli valori dalla media.

Il calcolo non è difficile con un foglio elettronico, ma per ora è importante sapere che:

· Una varianza vicina allo zero ci dice che il valore è pressoché simile in tutta la popolazione.

Una varianza molto alta ci fa capire che ci sono notevoli differenze.

1.18 Il campionamento

Finora abbiamo parlato della rilevazione dei valori come se fosse facile.

In realtà lo è solo quando la popolazione è limitata e possiamo coinvolgerla tutta in un'indagine "censuaria".

Se, invece, si tratta di migliaia o milioni di unità, i tempi, i costi e le difficoltà logistiche aumentano a dismisura.

Censimenti della popolazione vengono effettuati fin dall'antichità, ma in tempi moderni si è affermata la prassi delle indagini campionarie, come i sondaggi elettorali e le indagini di mercato, effettuate su una parte limitata della popolazione: il campione, appunto. In Italia, il metodo campionario è stato introdotto dall'Istat nel 1925.

Se un'indagine censuaria fornisce direttamente i parametri di interesse, in un'indagine campionaria questi vengono stimati.

Ma è meglio un'indagine campionaria ben progettata di un'indagine censuaria svolta con mezzi e preparazione limitati.

In un'indagine campionaria, ci sono alcuni punti critici che richiedono metodi statistici sofisticati:

- La scelta di un campione veramente rappresentativo della popolazione (e qui non è detto che un campionamento "casuale" sia il migliore).
- L'inferenza dei parametri della popolazione a partire da quelli calcolati sul campione.
- La stima degli errori di approssimazione.

1.19 L'inferenza statistica

L'inferenza statistica è un insieme di metodi per trarre conclusioni sull'intera popolazione in base alle informazioni ricavate da un campione.

L'inferenza statistica può essere svolta in modo autonomo, per studiare senza "preconcetti" le caratteristiche di una popolazione, o in modo ausiliario, per fornire supporto a un'ipotesi di ricerca.

In questo secondo caso, utile soprattutto con fenomeni ad alta variabilità come quelli biologici, il procedimento è una variante del ciclo conoscitivo induttivo.

I passi sono i seguenti:

· formulazione dell'ipotesi;

· valutazione della probabilità di ottenere certi risultati nella popolazione se l'ipotesi fosse vera;

· estrazione di una parte della popolazione (campionamento);

- calcolo delle statistiche sul campione (distribuzione di frequenza, media, varianza, ecc.);
- stima dei parametri nella popolazione in base ai risultati forniti dal campione (inferenza).

Le modalità di inferenza sono moltissime, ma ce n'è una che è interessante approfondire qui…

1.20 Statistica e abduzione: l'approccio bayesiano

Il teorema di Bayes o "teorema della probabilità delle cause" è una pietra angolare per costruire modelli dell'apprendimento dall'esperienza, facendo buon uso dei dati a disposizione.

Il teorema fornisce strumenti al ragionamento induttivo e a quello abduttivo (cioè quella forma di inferenza che consente di risalire dagli effetti alle probabili cause).

La domanda a cui voleva rispondere, nel XVIII secolo, il reverendo Thomas Bayes era:

"Se si verifica un evento B, qual è la probabilità che la sua causa sia A?"

Il teorema di Bayes, in un certo senso, "rovescia" la probabilità condizionata, passando da quella "a posteriori" a quella "a priori".

Il suo enunciato è questo:

La probabilità condizionata di un evento A rispetto a un evento B è uguale alla probabilità condizionata di B rispetto ad A moltiplicata per il rapporto tra le probabilità assolute di A e B.

In realtà, quello che ci interessa è un'estensione del teorema di Bayes ai casi in cui conosciamo tutte le possibili cause di un effetto che abbiamo verificato e ci chiediamo quale sia la più probabile.

È il tipico problema del medico di fronte a un certo tipo di dolori articolari, che (semplificando) potrebbero essere dovuti solo a tre cause: artrite, artrosi o influenza.

Per orientare la sua diagnosi, il medico tiene conto dei dati epidemiologici a disposizione:

- le probabilità relative di artrite, artrosi o influenza per quel tipo di paziente;
- la probabilità che ciascuna delle tre malattie abbia come sintomo dolori articolari di quel genere.

Poi applicando il teorema di Bayes, può stabilire la causa più probabile e, con quella ipotesi, richiedere esami più accurati.

1.21 Inferenza bayesiana e apprendimento automatico

L'inferenza bayesiana è un metodo di inferenza statistica che applica il teorema di Bayes per aggiornare la probabilità di un'ipotesi mano a mano che si rendono disponibili nuove informazioni.

Il contesto applicativo tipico è quello dei sistemi complessi, i sistemi in cui numerosi agenti interagiscono in modo non lineare e, quindi, gli effetti non sono proporzionali alle cause.

L'inferenza bayesiana differisce da approcci più tradizionali, perché i parametri del modello di sistema complesso vengono appresi sotto forma di distribuzioni di probabilità. Ne risulta un modello flessibile, facilmente aggiornabile in base a nuove osservazioni, senza ricominciare ogni volta da capo.

È per questo che l'inferenza bayesiana è così popolare nel machine learning o apprendimento automatico, le cui tecniche sono largamente impiegate in campi che ci interessano da vicino. In particolare l'analisi e l'interpretazione dei testi, come nel Natural Language ToolKit e in spaCy, e altri compiti di linguistica computazionale.

1.22 Il percorso si conclude qui