Probabilità e statistica

La statistica ed il calcolo delle probabilità sono due discipline matematiche di importanza cruciale non solo per il pensiero scientifico ma anche perché ci possono aiutare, nella vita di tutti i giorni, a farci un'opinione o a prendere una decisione "informata". Probabilmente non ti sei meravigliato nel trovarle abbinate. Per cominciare a capire in che cosa consistano, diamo uno sguardo d'insieme alle relazioni che esistono tra di esse, e soprattutto alle loro differenze.

	Calcolo delle probabilità	**Statistica**
area scientifica	matematica	matematica
aiuta a	giudicare/scegliere razionalmente su base quantitativa in un contesto incerto	giudicare/scegliere razionalmente su base quantitativa in un contesto incerto
quando/come è emersa come moderna disciplina	nel XVII secolo, dallo studio di problemi relativi ai giochi di azzardo	nei secoli XVII-XVIII, per le necessità informative degli stati nazionali
si usa principalmente per	predire il futuro	interpretare il passato, imparare da esso, interpretare il presente
supporta il ragionamento	deduttivo	induttivo
è una disciplina eminentemente	teorica; si può sviluppare a partire da pochi assiomi	pratica; ci aiuta a ipotizzare leggi osservando regolarità nei dati
consente di	valutare le conseguenze delle leggi di un certo mondo ideale	misurare quanto il nostro mondo si discosta da un mondo ideale

Ci si può fare un'idea della distinzione tra le due discipline considerando i possibili processi di pensiero di un matematico al primo incontro con il gioco dei dadi [2]

- se il matematico fosse un esperto di probabilità, in gergo un "probabilista", osservando un dado penserebbe: dadi con 6 facce? presumibilmente ogni faccia ha le stesse probabilità di atterrare all'insù; ora, supponendo che ogni faccia abbia una probabilità di 1/6, posso capire quali sono le mie possibilità di vincere puntando su una combinazione in una sequenza di lanci
- uno statistico, invece, vedrebbe i dadi e penserebbe: quei dadi possono sembrare ok, ma come faccio a sapere che non siano truccati? per un po' di tempo li guarderò e terrò traccia di quanto spesso esce ogni numero: poi potrò decidere se le mie osservazioni confermano l'assunzione che le facce abbiano uguali probabilità; quando ne sarò sicuro, chiederò ad un probabilista di consigliarmi come giocare.

Il calcolo delle probabilità

La probabilità si presenta in molte varianti e associata a molti termini diversi, come "attesa", "caso", "possibilità", "quotazione" (di un concorrente in una gara), ecc. La probabilità è una nozione utile quando non c'è *certezza*, cioè nei processi che sono intrinsecamente *casuali* o che tali consideriamo perché non conosciamo bene le leggi a cui ubbidiscono.

Il mondo è pieno di incertezza

Chi vincerà il campionato?

L’indice di borsa oggi crescerà o calerà?

Quale sarà il tempo oggi?

Questo film mi piacerà?

Noi - noi personalmente o un sistema “intelligente” a cui facciamo ricorso per ottenerne una previsione o un consiglio - interagiamo continuamente con il mondo, e dobbiamo farcene un’idea: che cosa è quel che vediamo o sentiamo e come potrebbe comportarsi o evolvere. E’ raro che possiamo provare che qualcosa è vero, ma ci chiediamo comunque quanto sia verosimile un certo evento o quale sia la più ragionevole spiegazione di un certo fatto.

Anche se comunemente si parla di “certezze” scientifiche, gli addetti ai lavori sanno bene che di certezze dimostrate con lo stesso rigore di un teorema di matematica ce ne sono ben poche. Restando tra gli addetti ai lavori, per esempio tra fisici sperimentali, due fonti di incertezza sono: i cosiddetti “errori di misura”, che ci lasciano incerti sul valore dei dati osservati, e la difficoltà di individuare la teoria (più) compatibile con i fenomeni osservati. [9]

Le difficoltà dello scienziato dipendono anche dal fatto che, in prima approssimazione esistono due tipi di leggi:

- leggi probabilistiche; un esempio semplice è quello del lancio di una moneta; un esempio più complesso è quello della genetica; in questi campi gli eventi osservabili non sono una pura conseguenza logica della teoria
- leggi deterministiche; l’esempio “classico” è fornito dalle leggi di Newton (meccanica classica), che peraltro sono valide solo ad una certa scala di osservazione e con qualche limitazione.

Una definizione e due assiomi

Tra le definizioni di probabilità che abbiamo incontrato, questa ci è sembrata abbastanza semplice:

> 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.

Anche se, come noteremo più avanti, questa definizione è un po’ ambigua, ne ricaviamo già alcuni punti fermi: a) la probabilità è un *rapporto*; b) il suo valore è un *numero* compreso tra 0 e 1; a volte tale valore è espresso come una percentuale o come una frazione in cui il numeratore rappresenta il numero di “successi” attesi rispetto al numero totale delle prove. I due estremi, zero e uno, rappresentano casi limite, perché in realtà corrispondono ad una certezza.

Il processo di cui si parla nella definizione potrebbe consistere nel lancio ripetuto di uno o più dadi, o nel giocare a testa (T) o croce (C); in quest’ultimo caso, se si giocasse 3 volte, un evento potrebbe consistere nella sequenza [T T C]; i restanti eventi corrispondono alle altre 8 possibili combinazioni: [T T T], [T C T], [C T T], ecc.

E con questo siamo già in grado di introdurre due dei tre *assiomi di base* del calcolo delle probabilità: [6]

> Il *primo assioma*, detto anche assioma *della positività*, ci dice che la probabilità di un generico evento E è un numero reale maggiore o uguale a 0; in simboli:
>
> $P(E) >= 0$
>
> Il *secondo assioma*, detto anche *assioma della certezza*, ci dice che la probabilità dell’evento certo I è 1; in simboli:
>
> $P(I) = 1$
>
> una formulazione alternativa è questa: la probabilità dell’intero *spazio campione*, simbolizzato da Ω, cioè la somma delle probabilità di tutti i possibili eventi tra loro alternativi, è 1; in simboli:
>
> $P(\Omega) = 1$

Ritornando alla definizione di probabilità, quello che può lasciare perplessi è l'espressione "nel lungo termine"; che cosa significa? Qualcuno la sostituisce con l'espressione "in un numero infinito di esperimenti"; ma questo rende la definizione non operativa, cioè impossibile da usare per determinare la probabilità di un evento in modo sperimentale.

Una definizione alternativa di probabilità, che forse piacerebbe di più ad un probabilista teorico, è la seguente:

> La probabilità di un evento è il rapporto fra il numero dei casi favorevoli ed il numero dei casi possibili supposti tutti *ugualmente possibili*.

Anche questa definizione ci lascia un po' insoddisfatti, perché non è facile individuare dei domini applicativi nei quali i casi siano tutti ugualmente possibili. Di solito si tratta di domini piuttosto artificiali, come quelli dei giochi d'azzardo.

D'altra parte, una definizione diretta di probabilità non è necessaria; abbiamo già visto che il calcolo delle probabilità è una disciplina teorica; che si può sviluppare a partire da pochi assiomi: sono tali assiomi e tutti i teoremi che ne possiamo derivare che costituiscono la vera definizione (indiretta) di probabilità.

Il terzo assioma

A questo punto dobbiamo introdurre il terzo assioma, che è leggermente più complicato degli altri. E per farlo ripassiamo prima la terminologia già incontrata e ne introduciamo dell'altra: [7]

- *processo casuale*: è un fenomeno naturale o un esperimento di cui possiamo osservare una serie di risultati (o uscite) che non siamo in grado di prevedere
- *evento*: è una specifica serie di risultati all'interno di un processo casuale
- *evento semplice*: è un evento che contiene un solo risultato, per esempio il lancio di un dado
- *evento composto*: è un evento che contiene più risultati, per esempio una sequenza di lanci di dado
- *spazio campione* (o *spazio campionario*): è l'insieme di tutti i possibili eventi in un processo casuale
- *eventi incompatibili*: eventi che non possono occorrere contemporaneamente; per esempio due eventi semplici di lancio di dado, in uno dei quali esce il 3 e nell'altro il 4.

Ed ecco il *terzo assioma*, detto anche *assioma dell'unione*:

> Se A e B sono eventi incompatibili, allora la probabilità dell'unione dei due eventi è uguale alla somma della probabilità dei singoli eventi. In simboli:
>
> $P(A \cup B) = P(A) + P(B)$
>
> dove possiamo leggere "$A \cup B$" come "A unione B

Esempio: se lanciamo un dado una sola volta, la probabilità che "esca il 3 o il 4" è la somma della probabilità che "esca il 3" e della probabilità che "esca il 4". Appare troppo ovvio che diverse facce di un dado corrispondano ad eventi tra loro incompatibili; altri tipi di esperimento, come l'estrazione di una carta da un mazzo, offrono esempi più significativi; lasciamo al lettore di individuare, tra le seguenti coppie di eventi, l'unica che comprende eventi incompatibili: [8]

a) estraendo una carta da un mazzo: E1 – "esce una carta di spade", E2 – "esce una carta di coppe"

b) giocando alla roulette: E1 – "esce un numero compreso tra 10 e 30", E2 – "esce il colore rosso"

c) estraendo una carta da un mazzo: E1 – "esce una figura", E2 – "esce una carta di denari".

Ragionare con la probabilità condizionata

Eventi indipendenti e probabilità composta

Consideriamo i due eventi di questo esempio:

lanciando una moneta 2 volte: A – “esce testa al primo lancio”, B – “esce coda al secondo lancio”

probabilmente già sappiamo che i due eventi sono *indipendenti*, che cioè il risultato del primo lancio non influenza quello del secondo; il ritenere il contrario costituisce la tipica *fallacia del giocatore d’azzardo*.

Il *teorema della probabilità composta*, afferma che

Nel caso di eventi indipendenti, la probabilità di un evento composto è uguale al prodotto delle probabilità degli eventi componenti.
In simboli:

$$P(A \cap B) = P(A) * P(B)$$

dove possiamo leggere “A ∩ B” come “A intersezione B

Usando questo teorema possiamo calcolare la probabilità composta per il nostro esempio:

$$P(A \cap B) = P(A) * P(B) = 1/2 * 1/2 = 1/4$$

dato che, nel caso di moneta ideale, simmetrica, la probabilità che esca ciascuna delle due facce è 1/2.

Eventi dipendenti e probabilità condizionata

Talvolta capita di dover calcolare la probabilità del verificarsi di un evento sapendo che un altro evento, collegato al primo per motivi logici o di successione temporale, si è già verificato o non potrà non verificarsi.

Con la notazione $P(A|B)$ si indica la *probabilità condizionata* di A rispetto a B, definita come la probabilità che ha l’evento A se l’evento B è certo.

Dai tre assiomi di base del calcolo delle probabilità, si può derivare il *teorema della probabilità condizionata*, che è espresso dalla seguente formula

$$P(A|B) = P(A \cap B) / P(B)$$

esso afferma che

La probabilità condizionata di un evento A rispetto ad un evento B possibile è uguale al rapporto tra la probabilità composta dei due eventi e la probabilità assoluta di B.

Il teorema della probabilità condizionata è utile nel caso di eventi tra loro dipendenti, mentre se provassimo a calcolare la probabilità condizionata nel caso di eventi indipendenti, ci renderemmo conto che essa coincide con la probabilità assoluta. Viceversa, le due uguaglianze

$$P(A|B) = P(A) \quad e \quad P(B|A) = P(B)$$

costituiscono congiuntamente una possibile definizione dell’indipendenza reciproca degli eventi A e B.

Consideriamo ora queste due coppie di eventi:

lanciando un dado: A – “esce il 5”, B – “esce un numero dispari”

lanciando un dado: A – “esce il 5”, B – “esce un numero pari”

dato che 5 è una delle tre facce dispari, nel primo caso

$P(A|B) = 1/3$, ovvero la probabilità condizionata di A rispetto a B è un terzo;

dato che tra i numeri pari il 5 non figura, nel secondo caso

$P(A|B) = 0$, ovvero la probabilità condizionata di A rispetto a B è nulla

mentre in entrambi i casi la probabilità assoluta di A è 1/6; quindi in entrambi i casi i due eventi della coppia sono *dipendenti*.

La statistica

Come abbiamo visto nell'unità precedente, la formalizzazione delle discipline sociali, quali la psicologia e l'economia, è più problematica di quella delle cosiddette scienze "hard", quali la chimica e la fisica; tuttavia, anche nelle discipline "soft" non si è rinunciato a far uso della matematica.

Per trattare gli aspetti e i modelli della realtà che si affrontano nelle scienze sociali, il ramo della matematica che si è maggiormente sviluppato è la statistica, ramo che è difficile distinguere nettamente dal calcolo della probabilità, dato che anche in esso la riduzione dell'incertezza costituisce uno degli obbiettivi centrali.

La statistica è un ramo della matematica abbastanza ampio e variegato, che si è sviluppato principalmente per rispondere alle necessità dell'interpretazione dei dati, della ricerca in essi di regolarità e tendenze, della formulazione di previsioni in campi applicativi disparati come, a solo titolo di esempio

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

Una storia di vedove e orfani

La statistica, come la conosciamo oggi, è nata un po' in tutto il mondo 2-3 secoli fa, spinta anche dalle necessità degli uffici che gestiscono le assicurazioni sulla vita. Una storia interessante in proposito venne raccontata in una conferenza di attuari accademici nel 1972 [11] ed è ripresa e sintetizzata anche in [3] e [4]. Qui ne facciamo una sintesi ulteriore da [3].

Sembra che gli scozzesi siano i creatori del moderno business delle assicurazioni sulla vita, un business che nacque dall'amicizia di due ministri della chiesa, Robert Wallace and Alexander Webster, appassionati di matematica e preoccupati per la sorte delle le vedove e degli orfani dei loro colleghi; infatti, dal momento in cui la chiesa riformata di Scozia permise ai suoi ministri di sposarsi, alla morte di questi vedove e orfani cadevano spesso in grave miseria.

Tra una bevuta e l'altra, perché erano noti ubriaconi, Wallace e Webster nel 1744 misero a punto, e fecero approvare per legge, un piano per creare un fondo di previdenza perpetuo; il problema più difficile era capire quanto ogni ministro dovesse versare periodicamente a tale fondo. Per capirlo, raccolsero da varie fonti, come i registri parrocchiali, dati utili a stimare l'aspettativa di vita dei ministri scozzesi, nonché dati sui tassi di rendita medi degli investimenti.

Il lavoro di Wallace e Webster portò all'elaborazione delle prime *tavole di mortalità*; in questo si servirono di tavole attuariali pubblicate mezzo secolo prima dal matematico inglese Edmond Halley, della consulenza di un professore dell'Università di Edimburgo e dei recenti progressi del calcolo delle probabilità, tra i quali il più notevole era la formulazione della *legge dei grandi numeri* da parte del matematico svizzero Jakob Bernouilli.

Tutto questo consentì ai due di calcolare quanti ministri sarebbero stati in vita e quanti sarebbero morti *mediamente* nel corso di ogni anno, quanti anni *in media* le vedove sarebbero sopravvissute e con quanti figli, eccetera eccetera; alla fine decisero che il fondo pensioni si sarebbe sostenuto se ogni ministro avesse versato almeno 10 sterline l'anno. Stimarono inoltre che da allora al 1765 il fondo avrebbe accumulato un capitale di 58.438 sterline; la loro previsione si rivelò precisa in modo sorprendente: quell'anno il capitale del fondo raggiunse proprio quella cifra, a meno di una sterlina!

La legge dei grandi numeri

E’ constatazione diffusa tra gli sperimentatori il fatto che, quando si ha a che fare con fenomeni casuali, la frequenza relativa di un evento, cioè il numero delle volte che quell’evento si presenta rispetto al numero totale delle prove, tende a *stabilizzarsi* all'aumentare del numero delle prove; questo fatto viene chiamato *legge empirica del caso*.

Ai primi del ‘700 la legge empirica del caso fu riformulata da Jakob Bernouilli, che la ribattezzò come *legge dei grandi numeri*, nei seguenti termini:

> Se p è *probabilità* di successo di un evento E, cioè la probabilità del verificarsi di E in una prova, allora la *frequenza relativa* dei successi in n prove indipendenti tende a p, quando n tende a infinito.

La legge dei grandi numeri, detta anche *Teorema di Bernouilli* è importante perché mette chiaramente in relazione tra loro due concetti

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

essa non è di nessuna utilità per prevedere il successo o meno di un evento volta per volta (anche se alcuni ci provano, a loro rischio e pericolo!); dice invece che, se è difficile prevedere l’esito del singolo evento, per esempio, il *tempo di vita* di un nucleo radioattivo, può essere facile stimare la durata media di vita di n nuclei radioattivi, con approssimazione tanto migliore quanto più grande è il numero n.

Un semplice quiz

Abbiamo appena accennato ad una delle nozioni fondamentale della statistica, la “media”; proviamo allora a fare piccolo test su di esso. Per ottenere il numero medio di bambini per famiglia in un villaggio, un insegnante ha contato il numero totale di bambini. Ha poi diviso per 50 perché c'erano 50 famiglie. Il numero medio di figli per famiglia era di 2,2. Quale delle seguenti affermazioni è sicuramente vera? [1]

(a) La metà delle famiglie del villaggio ha più di due figli

(b) Più famiglie del villaggio hanno 3 figli che 2 figli

(c) Nel villaggio ci sono in totale 110 bambini

(d) Ci sono 2,2 bambini nel villaggio per ogni adulto

(e) Il numero più comune di figli in una famiglia è 2

(f) Nessuna delle affermazioni precedenti è vera.

Certamente sarai capace di trovare da solo la risposta giusta; ma, in quanto tempo?

Termini e concetti di base della statistica

Nella statistica si usa distinguere due principali filoni:

- *la statistica descrittiva*; raccoglie i dati, li sintetizza e li commenta per descrivere le caratteristiche di una *popolazione statistica*; una popolazione può essere un insieme di persone, animali, piante, ecc.; gli elementi della popolazione vengono detti *unità statistiche*
- la *statistica inferenziale*; utilizza i dati per indurre regole e fare previsioni.

[prendere da qualche parte uno schemino su statistica descrittiva e inferenziale?]

Con riferimento all’insieme dei valori numerici di una certa proprietà di una *popolazione*, la statistica considera certe grandezze derivate o *parametri*; i due parametri più importanti e noti sono:

- il *valore medio*, detto anche *media aritmetica*; è il rapporto tra la somma dei valori della proprietà considerata ed il numero di tali valori, cioè delle unità statistiche a cui i valori si riferiscono

$(v^1 + v^2 + ... + v^n)$ / n

nel nostro esempio, la popolazione è costituita dalle 50 famiglie del villaggio, ogni famiglia costituisce un'unità, la proprietà che consideriamo è il numero dei figli, la somma dei valori di tale proprietà è 110 e il valore medio è 110/50 = 2,2

- il *valore mediano*, detto anche la *mediana*; la mediana è il valore centrale, tale che metà della popolazione presenta un valore inferiore ad esso per la proprietà considerata, e l'altra metà della popolazione un valore superiore; nell'esempio sulle famiglie del villaggio, non disponiamo di dati sufficienti a individuare la mediana.

Campionamento e inferenza statistica

Censimenti e campionamento

In linea di principio è possibile ricavare tutti i parametri di interesse di una popolazione statistica effettuando un'*indagine censuaria*, che coinvolge l'intera popolazione. La più tipica indagine censuaria è il censimento della popolazione di una regione: notizie di censimenti, spesso limitati ad alcune categorie di persone, si hanno presso quasi tutti gli antichi popoli.

In considerazione della complessità, della durata e del costo e delle indagini censuarie, in tempi moderni si è affermata la prassi delle *indagini campionarie*, effettuate cioè su un campione limitato della popolazione; mentre le prime forniscono direttamente i parametri di interesse, nel caso delle indagini campionarie, quali sono i sondaggi elettorali e le indagini di mercato, quei parametri sono stimati.

In molti casi un'indagine campionaria ben progettata può fornire risultati più accurati che non un'indagine censuaria svolta con mezzi e preparazione limitati. In Italia il metodo campionario è stato introdotto nella prassi dell'ISTAT nel 1925.

Tecniche statistiche sofisticate sono usate

- per *scegliere* un campione rappresentativo; esiste infatti un'ampia varietà di metodi di campionamento e non necessariamente un campionamento "casuale" è il migliore
- per *inferire*, dai parametri calcolati sul campione, i parametri dell'intera popolazione e stimare l'entità dei conseguenti errori di approssimazione.

L'inferenza statistica

L'inferenza statistica è un insieme di metodi con cui si cerca di trarre conclusioni su una popolazione in base ad informazioni ricavate da un campione. Essa può essere svolta in modo autonomo, per studiare senza "preconcetti" le caratteristiche di un'intera popolazione, o essere svolta in modo ausiliario, per costruire l'*evidenza scientifica* su un'ipotesi di ricerca, soprattutto quando si studiano fenomeni ad alta variabilità quali possono essere i fenomeni biologici.

Nel secondo caso, il procedimento, variante del più generale ciclo conoscitivo induttivo, prevede i seguenti passi

- 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 campionarie (esempio: media sul campione)
- stima dei parametri nella popolazione in base ai risultati forniti dal campione (inferenza).

L’approccio bayesiano all’apprendimento

Il teorema di Bayes

Il *teorema di Bayes*, noto anche come *teorema della probabilità delle cause*, può essere considerato il teorema fondamentale *dell’abduzione*. Proposto dal reverendo Thomas Bayes nel ‘700, viene impiegato per calcolare la probabilità di un evento A quale causa scatenante di un evento B verificato. Esso “rovescia” la probabilità condizionata passando da quella “a posteriori” a quella “a priori” . [13]

L’enunciato del *teorema di Bayes* non è particolarmente complicato:

> La probabilità condizionata di un evento A rispetto ad 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.
>
> $P(A|B) = P(B|A) * P(A) / P(B)$

Anche la sua dimostrazione non è difficile, ma la omettiamo come nel caso degli altri teoremi.

In realtà, più che il *teorema di Bayes* quale sopra esposto, a noi interessa una sua estensione al caso in cui sappiamo che gli eventi A^1, A^2, … A^n tra loro disgiunti ricoprono completamente lo spazio dei campioni di A e conosciamo le probabilità assolute di ciascuno di essi; in tal caso esso ci consente di calcolare la probabilità di ciascun evento A^i come causa dell’evento verificato B.

Invece di mostrare la formulazione della versione estesa del teorema, descriviamo con un esempio il tipico problema che esso consente di risolvere. [13]

> Una ditta produce un componente per computer in tre diversi stabilimenti, che chiameremo S1, S2 e S3. I componenti prodotti possono o meno essere difettosi. Sappiamo che la ditta produce il 30% del totale dei componenti nello stabilimento S1, il 25% nello stabilimento S2 e il restante 45% nello stabilimento S3. Indagini statistiche, condotte dall’azienda, confermano che il 2% dei componenti prodotto nel primo stabilimento è difettoso, nel secondo stabilimento il difetto si presenta con un’incidenza del 1,8%, mentre il terzo stabilimento produce solo l’1,33% di componenti difettosi. Un cliente ordina un componente e riceve un componente difettoso. Calcolare la probabilità che il componente ricevuto provenga dal secondo stabilimento.

La soluzione è: il componente difettoso proviene con una probabilità del 27,3% dallo stabilimento S2. Non ne diamo la dimostrazione, anche perché non abbiamo esposto la versione estesa del teorema di Bayes su cui essa si può basare; d’altronde ci interessava solo dare un’idea del suo possibile utilizzo.

Inferenza bayesiana e apprendimento automatico

Il teorema di Bayes è considerato una pietra angolare nella costruzione di modelli di come si possa apprendere dall’esperienza, facendo buon uso dei dati a disposizione; allora non deve sorprendere il fatto che la cosiddetta *inferenza bayesiana*, una metodologia ad esso ispirata, sia così popolare nel campo del *machine learning*, o *apprendimento automatico*.

Il contesto applicativo tipico è quello dei *sistemi complessi*, i sistemi in cui numerosi *agenti* interagiscono in modo non lineare, tale cioè che gli effetti non dipendono dalle cause in modo semplice, proporzionale. Quando essi non sono troppo complessi, si prova a rappresentarne il comportamento mediante modelli *parametrici*; la definizione di un modello richiede l’identificazione dei valori dei *parametri* che figurano in esso; spesso esistono motivi teorici per supporre che tali valori seguano *distribuzioni statistiche* ben studiate, come per esempio la nota *distribuzione di Bernouilli* per eventi a valori *discreti* (tipo lancio di dati) o la *distribuzione di Gauss* (la tipica curva gaussiana, che ricorda una campana) per eventi a valori continui.

L'inferenza bayesiana è un metodo di inferenza statistica in cui il teorema di Bayes viene utilizzato per aggiornare la probabilità di un'ipotesi, cioè dei parametri del modello che la rappresenta, a mano a mano che si rendono disponibili ulteriori prove o informazioni. [15]

Nell'inferenza bayesiana, a differenza che in approcci più tradizionali, i parametri del modello vengono appresi sotto forma di *distribuzioni di probabilità*, cosa che rende il modello stesso flessibile e che consente di aggiornare la sua forma, quando disponiamo di nuove osservazioni sul sistema, senza dover buttar via tutto quanto si è "appreso" fino ad allora. Inoltre sembra che l'inferenza bayesiana consenta, entro certi limiti, di trattare modelli approssimati di sistemi non lineari, cosa praticamente impossibile con altri approcci. [14]

In un precedente percorso abbiamo notato come tecniche di apprendimento automatico siano largamente impiegate in compiti di analisi e interpretazione dei testi. Abbiamo brevemente descritto il *Natural Language ToolKit* (*NLTK*), una libreria Python ad orientamento didattico: essa dedica un'intera sezione al machine learning; *spaCy*, una libreria Python un po' più moderna e di taglio più professionale, poggia su algoritmi aggiornati di *deep learning*. E abbiamo visto come algoritmi di apprendimento automatico siano largamente applicati in compiti di linguistica computazionale quali

- segmentazione dei testi in frasi e in token
- classificazione dei documenti e *sentiment analysis*
- calcolo della somiglianza stilistica e/o semantica tra documenti
- *allineamento* frase per frase e, entro certi limiti, parola per parola di testi paralleli
- costruzione di *memorie di traduzione* e traduzione automatica su base statistica (SMT = statistical machine translation).

WEBOGRAFIA

[1] Critical thinking about average. From top drawer teachers - resources for teachers of mathematics

https://topdrawer.aamt.edu.au/Statistics/Assessment/Assessment-tasks/Critical-thinking-about-average

[2] Steve Skiena, Probability versus Statistics. In Calculated Bets: Computers, Gambling, and Mathematical Modeling to Win!

https://www3.cs.stonybrook.edu/~skiena/jaialai/excerpts/node12.html

[3] A very Scottish history of insurance

http://sonsofscotland.com/scottish-history-insurance/

[4] Yuval Noah Harari, *Sapiens, Da animali a dèi – Breve storia dell'umanità*, Bompiani, 2018

[6] Università di Bologna – Progetto matematica, Definizione assiomatica o la teoria unificata di probabilità.

http://progettomatematica.dm.unibo.it/ProbElem/8definiz.html

[7] Nelson Education Ltd., Probability Events. 2009

http://www.nelson.com/school/elementary/mathK8/quebec/0176237879/documents/NM8SB_12B.pdf

[8] YouMath - Scuola di matematica e fisica, Eventi compatibili, incompatibili e complementari.

https://www.youmath.it/lezioni/probabilita/probabilita-discreta/1200-eventi-compatibili-e-incompatibili.html

[9] Giulio D'Agostini, Incertezze in Fisica e nelle altre scienze naturali. Istituto Nazionale di Fisica Nucleare

https://www.roma1.infn.it/~dagos/PRO/node5.html

[10] Giancarlo Ragozini, Università di Napoli Federico II, Principali teoremi del calcolo delle probabilità.

http://www.federica.unina.it/sociologia/statistica/principali-teoremi-del-calcolo-della-probabilita/

[11] J.B. Dow, Early actuarial work in eighteenth-century Scotland. Lecture delivered to the Faculty of Actuaries in Scotland, 16th October 1972.

https://www.researchgate.net/publication/305950556_Early_actuarial_work_in_eighteenth-century_Scotland

[12] Inferenza statistica

https://ccrma.stanford.edu/~apinto/Inferenza_statistica.pdf

[13] Giovanni Barazzetta, Il teorema di Bayes e la probabilità condizionata.

https://library.weschool.com/lezione/bayes-teorema-formula-esercizi-probabilita-condizionata-14965.html

[14] When machine learning meets complexity: why Bayesian deep learning is unavoidable

https://medium.com/neuralspace/when-machine-learning-meets-complexity-why-bayesian-deep-learning-is-unavoidable-55c97aa2a9cc

[15] Wikipedia, Bayesian inference.

https://en.wikipedia.org/wiki/Bayesian_inference