M. Besozzi

ERRORI COGNITIVI, PROBABILITÀ E DECISIONE MEDICHE

Applicazioni e utilità del teorema di Bayes nella diagnostica di laboratorio

probabilità a posteriori

probabilità a priori

$$P(M+|T+) = \frac{P(T+|M+) \cdot P(M+)}{P(T+|M+) \cdot P(M+) + P(T+|M-) \cdot P(M-)}$$

(la causa | dato l'effetto)

(l'effetto | data la causa)

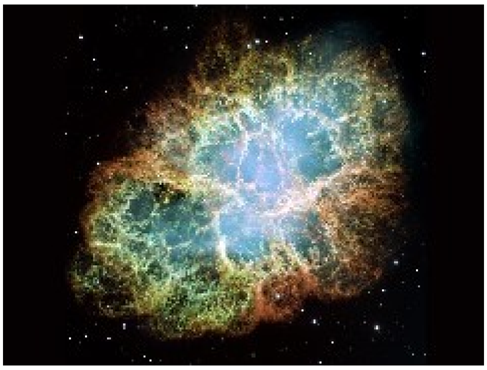
Nota dell'autore:
questi appunti contengono almeno un errore[1].

[1] Non è necessario leggere attentamente tutti questi appunti per scoprire se esiste un errore. Se l'errore esiste, questi appunti contengono almeno un errore. Se non ci fosse nessun errore, l'errore sarebbe rappresentato dall'affermazione che "questi appunti contengono almeno un errore". Pertanto in ogni caso questi appunti contengono almeno un errore. E in ogni caso ringrazio l'autore di questa deliziosa applicazione della logica [1]. Che ci ricorda come, in questo caso, il rigore nella forma racchiuda il rigore della sostanza.

[1] http://www.einaudi.it/libri/libro/piergiorgio-odifreddi/c-era-una-volta-un-paradosso/978880617507

1. Prologo

"Le véritable voyage de découverte ne consiste pas à chercher de nouveaux paysages, mais à avoir de nouveaux yeux."
(Marcel Proust)


La nebulosa del Granchio

Vedere con nuovi occhi. Ecco solo alcuni tra gli esempi più memorabili:
- l'Hubble Space Telescope vede, letteralmente con i "nuovi occhi" della moderna astronomia, un telescopio orbitante a oltre 500 chilometri di altezza, quello che gli astronomi cinesi avevano visto accadere nel 1054[2];
- Galileo, con i nuovi occhi della scienza, vede l'errore commesso da Aristotele, un errore tenacemente tramandato di generazione in generazione per quasi 2000 anni[3];
- Semmelweis vede una cosa che era sotto gli occhi di tutti, ma che evidentemente gli altri non riuscivano a vedere[4];
- Pasteur vede l'esistenza dei microrganismi[5], e in seguito a questo Lister vede le infezioni[6],

---

[2] Era il 4 luglio dell'anno 1054, e nelle cronache degli astronomi cinesi si legge: "*nell'ultimo anno di Chi-ho, della quinta lunazione, il giorno di Chi-ch'ou, una stella brillantissima apparve a circa qualche pollice a SE da Ten Kwan ...*" e qualche mese dopo: "*la stella meravigliosa è diventata invisibile; era stata visibile anche di giorno come Venere al mattino, con raggi puntati in tutte e quattro le direzioni, il colore era bianco e rossastro e rimase visibile per 23 giorni ...*". Si trattava dell'esplosione di una supernova: il suo residuo appare oggi nel cielo come nebulosa del Granchio [1], al cui centro è stata addirittura individuata la stella di neutroni nella quale è concentrata gran parte della massa residua della stella esplosa (nell'immagine, la nebulosa del Granchio fotografata dall'Hubble Space Telescope).

[3] Aristotele (384-322 a.C.) nei suoi scritti di "Fisica" asseriva che lo stato naturale dei corpi è la quiete, ossia l'assenza di moto, e che qualsiasi oggetto in movimento tende a rallentare fino a fermarsi, a meno che non venga spinto a continuare il suo movimento. Dopo quasi 2000 anni Galileo Galilei (1564-1642) scoprì l'errore di Aristotele, esponendo con estrema chiarezza il principio di inerzia, in particolare in due opere, scritte, rispettivamente, nel 1632 e nel 1638: "Il Dialogo sopra i due massimi sistemi del mondo" e "Discorsi e dimostrazioni matematiche intorno a due nuove scienze attenenti alla mecanica et i movimenti locali". Scrive Galileo nel Dialogo: "*il mobile durasse a muoversi tanto quanto durasse la lunghezza di quella superficie, né erta né china; se tale spazio fusse interminato, il moto in esso sarebbe parimenti senza termine, cioè perpetuo*"[2]. Ma questo, scrive ancora Galileo: "*deve intendersi in assenza di tutti gli impedimenti esterni e accidentari*" … e che gli oggetti in movimento siano: "*immuni da ogni resistenza esterna: il che essendo forse impossibile trovare nella materia, non si meravigli taluno, che faccia prove del genere, se rimanga deluso dall'esperienza*". La prima enunciazione formale di questo principio appartiene a Newton, che lo descrisse nella sua famosa opera "*Philosophiae Naturalis Principia Mathematica*" (1687), con la seguente formula: "*Lex prima: Corpus omne perseverare in statu suo quiescendi vel movendi uniformiter in directum, nisi quatenus a viribus impressis cogitur statum illum mutare.*"

[4] Nel 1847 Ignaz Philipp Semmelweis (1818-1865), un medico ungherese, prescrive agli studenti che hanno fatto pratica di dissezione dei cadaveri in sala anatomica, e a tutto il personale del reparto, di lavarsi le mani con una soluzione disinfettante (di cloruro di calcio) prima di visitare le puerpere. La febbre puerperale, che a quei tempi uccideva misteriosamente migliaia di puerpere, soprattutto nei grandi ospedali, nel reparto di maternità del Policlinico di Vienna diretto da Semmelweis scende in due anni dal 12% all'1%. [3]

[5] Louis Pasteur (1822-1895) è universalmente considerato il fondatore della moderna microbiologia. Con alcuni esperimenti cruciali riuscì a dimostrare che l'antica teoria della "generazione spontanea" non aveva alcun fondamento. Questi risultati diedero inizio a un'aspra polemica con il biologo francese Félix Pouchet, che si concluse con l'accettazione dei risultati di Pasteur da parte dell'Académie des Sciences nel 1864. Significativo rilevare che tutte le grandi scoperte dello scienziato francese vengono realizzate affrontando i problemi più gravi, a metà dell'Ottocento, dell'agricoltura, dell'industria agraria, dell'allevamento. La successione delle sue scoperte corrisponde ad una successione di studi su problemi agricoli, agroindustriali, veterinari: anomalie della fermentazione della birra (1854); fermentazione del vino e dell'aceto (1861-1862); pastorizzazione (1862); alterazioni del vino di origine fungina o

fondando l'asepsi;
- Alexander Fleming porta a compimento il lavoro di Pasteur e di Lister, vedendo quello che un paio di generazioni di batteriologici non avevano saputo vedere[7];
- Wegener[8] vede la deriva dei continenti nella straordinaria concordanza delle coste dei continenti affacciati attorno all'Oceano Atlantico, che peraltro era da tempo sotto gli occhi di tutti;
- Einstein vede il tempo in modo diverso e, partendo dalla esigenza che i fenomeni naturali (leggasi "leggi di natura") come la velocità della luce siano identici per tutti gli osservatori, arriva alla teoria della relatività[9];
- Watson e Crick vedono che nel DNA l'informazione *deve* risiedere nell'ordine in cui sono disposte le quattro basi azotate[10].

---

batterica (1863-1864); malattie del baco da seta (1865-70); colera dei polli (1880); carbonchio di bovini, ovini, equini (1881); rabbia silvestre. [3]

[6] Joseph Lister (1827-1912) collega le idee di Pasteur al problema della sepsi. Utilizza una soluzione diluita di acido carbolico (fenolo) per disinfettare le ferite. Nel 1867 annuncia alla British Medical Association che nel suo reparto di chirurgia alla Glasgow Royal Infirmary non vi è stato nessun caso di sepsi per nove mesi. [3]

[7] Come afferma Alexander Fleming (1881-1955) "*Un certo tipo di penicillium produce una potente sostanza antibatterica in una coltura […] L'agente attivo si può filtrare senza difficoltà e si è dato il nome di penicillina a filtrati del brodo di coltura delle muffe […] La penicillina risulta non tossica agli animali anche in dosi elevate e non è irritante […] Si ritiene che possa svolgere una funzione di antisettico efficace se applicato a, oppure iniettato in, aree infette da microrganismi sensibili alla penicillina*." [4]. Per la scoperta della penicillina Fleming ricevette il premio Nobel con Florey e Chain nel 1945.

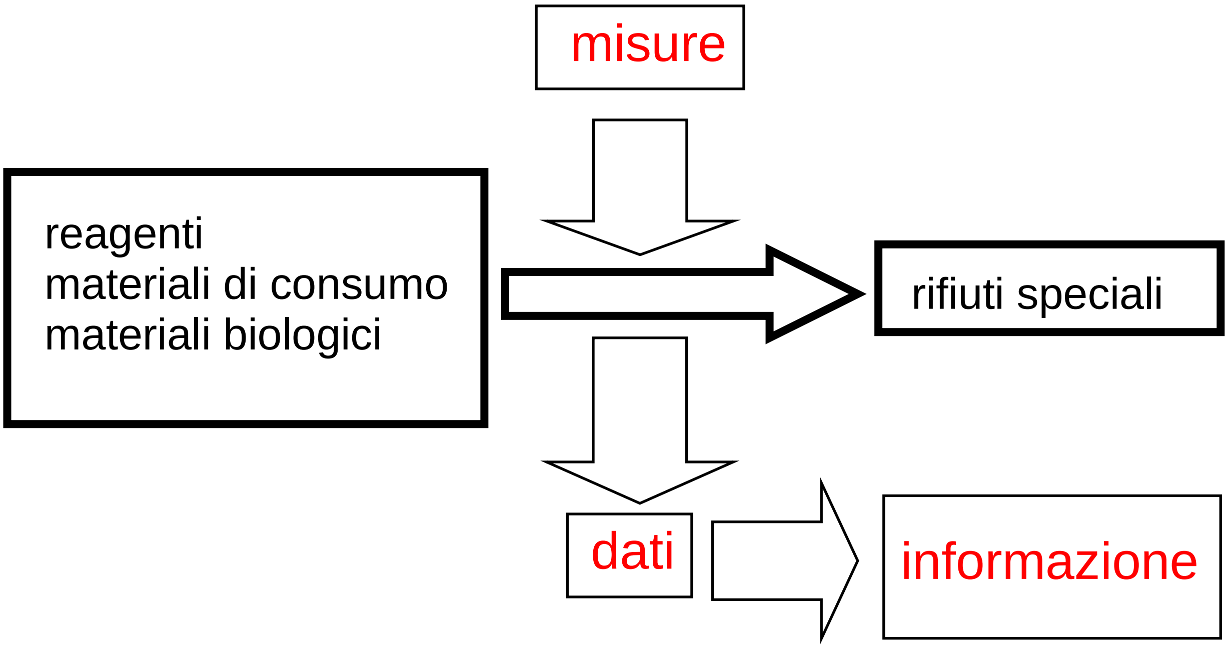
[8] Alfred D. Lother Wegener (1880-1930) è stato un geologo e meteorologo tedesco. Iniziò a lavorare alla teoria della deriva dei continenti nel 1910. Nel libro *The Origin of Continents and Oceans*, Wegener ricorda che la teoria si sviluppò a partire dall'osservazione della straordinaria concordanza delle coste dei continenti affacciati attorno all'Oceano Atlantico. Nel 1911 Wegener venne a conoscenza delle nuove teorie che stavano emergendo dallo studio dei fossili, in particolare quella di un antico collegamento fra Brasile e Africa. L'anno successivo Wegener annunciò la teoria della deriva dei continenti in una conferenza della Società Geologica di Francoforte sul Meno dal titolo *La formazione dei continenti e degli oceani in base alla geofisica*, a cui seguì una seconda conferenza dal titolo *Gli spostamenti orizzontali dei continenti*, tenuta presso la Società per il Progresso delle Scienze naturali di Magdeburgo.

[9] Nel sistema GPS (Global Positioning System) gli orologi satellitari sono affetti dalle conseguenze della teoria della relatività. Infatti, a causa degli effetti combinati della velocità relativa, che rallenta il tempo sul satellite di circa 7 microsecondi al giorno, e della minore curvatura dello spaziotempo a livello dell'orbita del satellite, che lo accelera di 45 microsecondi, il tempo sul satellite scorre ad un ritmo leggermente più veloce che a terra, causando un anticipo di circa 38 microsecondi al giorno, e rendendo necessaria una correzione automatica da parte dell'elettronica di bordo. Questa osservazione è un'ulteriore prova dell'esattezza della teoria in un'applicazione del mondo reale. L'effetto relativistico rilevato è infatti esattamente corrispondente a quello calcolabile teoricamente, almeno nei limiti di accuratezza forniti dagli strumenti di misura attualmente disponibili.

[10] James Watson (1928) e Francis Crick (1926-2004) il 25 aprile del 1953 pubblicano su Nature la struttura del DNA [5] servendosi di dati dei lavori non ancora pubblicati di Rosalind Franklin e di Maurice Wilkins. Per giungere alla scoperta i due scienziati si basano su ciò che gli studiosi che li avevano preceduti avevano scritto. Per costruire il modello del DNA si armarono di cartoncino, filo di ferro e molta pazienza... Le analisi di cristallografia ai raggi X di Rosalind Franklin furono determinanti nel suggerire a Watson e Crick che l'informazione doveva risiedere nell'ordine in cui erano disposte le quattro basi azotate. Watson, Crick e Wilkins ricevettero il Premio Nobel per la Medicina nel 1962 per la scoperta della struttura del DNA. L'esclusione della Franklin dal premio è dovuta alla sua morte di cancro nel 1958, probabilmente a causa delle radiazioni a cui i suoi studi la avevano lungamente sottoposta. Nell'immaginario comune, la scoperta della struttura del DNA è legata ai soli Watson e Crick. Nella comunità scientifica è opinione diffusa che i motivi di ciò siano legati alla scarsa reputazione di Wilkins (che secondo molti era solo "alla ricerca di pubblicità") e, soprattutto, alla morte della Franklin.

Un modo particolarmente interessante di vedere anche il laboratorio di analisi cliniche “con nuovi occhi” è quello di vederlo come “officina”, quindi come un processo produttivo. Quale è il valore aggiunto della trasformazione, che avviene nel laboratorio/officina, di reagenti, materiali di consumo e materiali biologici in rifiuti? In realtà un valore aggiunto esiste (ovviamente).

Ma è costituito da un prodotto intangibile: l’informazione. Attraverso le misure[11] si arriva ai dati, e da questi si genera l’informazione.



E sarà proprio l’informazione al centro della nostra attenzione: con l’ambizione di vedere con nuovi occhi il lavoro quotidianamente svolto nel laboratorio clinico.

Ma se tutto ruota attorno all’informazione, non possono non sorgere alcune domande:
- *che valore dobbiamo/possiamo assegnare all’informazione prodotta con i dati di laboratorio?*
- *possiamo misurare il contributo che l’informazione di laboratorio fornisce alla conoscenza medica?*
- *esistono delle regole in grado di garantire scelte razionali nelle condizioni di incertezza tipiche della diagnosi medica*[12]*?*

L’obiettivo è di dimostrare come, nell’ambito della matematica, la teoria della probabilità, e in particolare il teorema di Bayes, fornisca concetti cruciali e strumenti chiave per rispondere a queste domande. Cosa che cercheremo di fare sviluppando i seguenti argomenti:
- errori cognitivi;
- il problema gnoseologico;
- dati, informazione e conoscenza;
- complessità, probabilità e teorema di Bayes;
- teorema di Bayes e informazione diagnostica;

[11] Poiché i dati derivano dalle misure, è necessario considerare tutte le relative implicazioni a livello metrologico. Nonostante ciò vada di là degli obiettivi di queste note, una introduzione all’argomento metrologico viene riportata nell’appendice A, e una introduzione al problema dell’incertezza delle misure viene riportato nell’appendice B.

[12] Come vedremo, e come è necessario, il discorso si allargherà immediatamente, in quanto il problema della diagnosi medica è un problema specifico nell’ambito di un problema più generale: come garantire scelte razionali nelle condizioni di incertezza che per definizione caratterizzano qualsiasi decisione umana.

- teorema di Bayes e strategie diagnostiche;
- teorema di Bayes e decisioni mediche.

Bibliografia e riferimenti

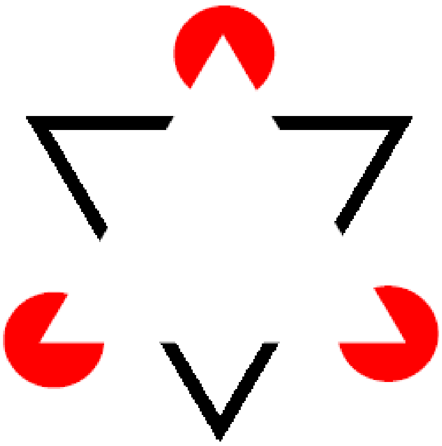
[1] http://astrolink.mclink.it/messier/m/m001.html
[2] http://it.wikipedia.org/wiki/Principio_di_inerzia
[3] Guthrie D. Storia della medicina. Feltrinelli, 1967.
[4] Gillies D, Giorello G. La filosofia della scienza nel XX secolo. Laterza, 1995.
[5] Watson J D, Crick FHC. Molecular structure of nucleic acids. Nature. No. 4356, 25 aprile 1953, p. 737 (http://www.nature.com/nature/dna50/watsoncrick.pdf)

2. Errori cognitivi

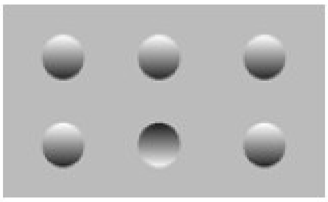
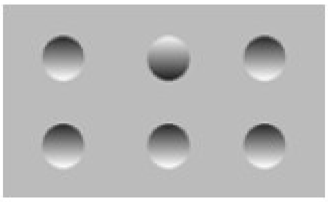
“Considerate la vostra semenza:
fatti non foste a viver come bruti,
ma per seguir virtute e canoscenza".
(Dante Alighieri, Inferno, canto XXVI)

"To the small part of ignorance that we arrange and classify we give the name knowledge". (Ambrose Bierce)

Alcune immagini dimostrano come i nostri sensi possono essere ingannati. Un’illusione percettiva molto nota è il *triangolo di Kanizsa* (dal nome di Gaetano Kanizsa, lo psicologo triestino che la ha individuata). Nella figura i soggetti dichiarano di percepire un triangolo bianco parzialmente sovrapposto a tre cerchi rossi e a un triangolo con i bordi neri. Ma, a ben vedere, il triangolo non c’è. La sua presenza è soltanto suggerita da tre segni neri a V e da tre cerchi rossi privi di una “fetta”. E’ il nostro sistema percettivo a completare l’immagine interpretando segni a V e cerchi come parti di figure parzialmente nascoste. (In certi casi l’illusione è così forte che alcuni soggetti dichiarano di percepire il triangolo bianco come leggermente più chiaro dello sfondo) [1].

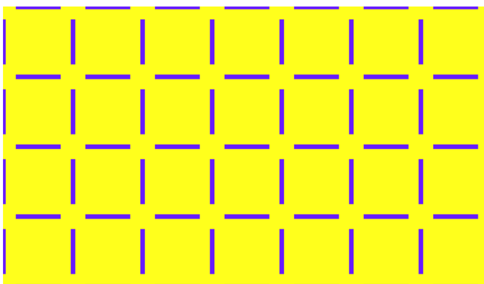


O ancora, si considerino le due immagini seguenti:

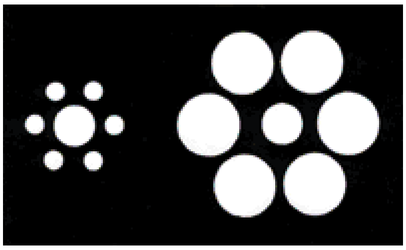
a)  b) 

Tutti diremmo che a) raffigura cinque cerchi convessi e uno concavo, e che b) raffigura cinque cerchi concavi e uno convesso. Tuttavia b) non è altro che a) capovolta. Per sincerarsene basta girare il foglio e constatare che ora in a) ci sono cinque cerchi concavi e uno convesso, e in b) cinque cerchi convessi e uno concavo [1].

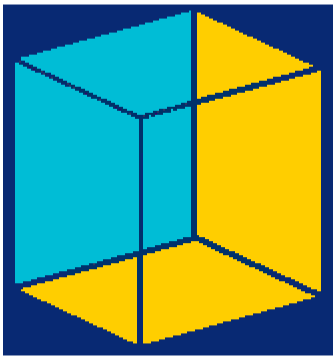
Nell’immagine accanto appaiono dei cerchi nelle aree vuote, dove dovrebbero idealmente trovarsi le intersezioni tra le righe orizzontali e le righe verticali. Si tratta di cerchi “fantasma”, che in realtà non esistono, ma anche se non sono mai stati tracciati fisicamente, risultano “indubitabilmente” visibili.



Un’altra ”illusione visiva” abbastanza nota è rappresentata dall’immagine nella quale un cerchio bianco centrale, circondato dai sei cerchi bianchi periferici, è rappresentato due volte, la prima con i sei cerchi periferici “piccoli”, la seconda con i sei cerchi periferici “grandi”. Nelle due rappresentazioni il cerchio centrale appare di diverse dimensioni, indicando una sensibilità del sistema visivo al contesto: in un contesto di cerchi più piccoli il cerchio centrale appare più grande, e viceversa.



Il famoso cubo di Necker è un esempio di immagine che presenta una intrinseca ambiguità. Si può percepirlo come un cubo visto dall'alto: la faccia blu e la faccia gialla verticali sono opache, così come la parete inferiore. La faccia superiore è vista dall'esterno ed è trasparente, così come sono trasparenti le due facce anteriori, che si proiettano verso l'osservatore. Ma si può anche percepirlo come un cubo visto dal basso: a questo punto sono le facce colorate ad essere trasparenti. La faccia inferiore viene vista dall'esterno, così come le due facce anteriori, la sinistra blu, e la destra gialla. Mentre la faccia superiore e le due facce laterali che si allontanano dall'osservatore sono viste in trasparenza. Alcuni soggetti lo percepiscono la prima volta "istintivamente" in un modo, altri lo percepiscono nel modo alternativo: ma, anche se con un po' di fatica, si più imparare a passare da un modo di vederlo all'altro e ad alternarli a piacimento abbastanza rapidamente.



La conclusione di queste brevi considerazioni? Attenti alla frase "credo solo quello che vedo!"…

Le illusioni visive/ottiche sono in realtà solo la punta dell'iceberg: mentre le illusioni della ragione, molto più subdole e il più delle volte inapparenti anche alle menti più allenate (vedremo poco più avanti l'esempio del problema di Monty Hall), sono la parte sommersa che solamente da poco si incomincia ad esplorare.

Parafrasando Shakespeare e il suo famoso "essere o non essere: questo è il problema", diciamo oggi che "decidere in condizioni di incertezza: questo è il problema"[13]. Così come avviene per la nostra percezione visiva, anche la nostra ragione può essere ingannata, e conseguentemente le nostre scelte e le nostre decisioni possono essere falsate da questo inganno. Sull'argomento "decidere in condizioni di incertezza" la ricerca si muove in due direzioni. La prima direzione è quella della teoria dei giochi. L'idea di base della teoria è che razionalità e utilità siano tra loro direttamente collegate: più razionalità ci metto, più utilità mi aspetto di ricavarne.

Un eccellente esempio si trova nella "Tosca" di Giacomo Puccini (1858-1924). Scarpia, capo della polizia papalina, deciso a fare fucilare il pittore Mario Cavaradossi, bonapartista, promette a Tosca di non farlo se Tosca gli concederà le sue grazie. Tosca vuole salvare Mario, ma non vuole cedere a Scarpia. In ordine di razionalità decrescente le scelte, per Tosca, sono:
- salvare Mario senza cedere a Scarpia;
- salvare Mario cedendo a Scarpia;
- non salvare Mario senza cedere a Scarpia;
- non salvare Mario cedendo a Scarpia.

Simmetricamente e in ordine di razionalità decrescente le scelte, per Scarpia, sono:
- fare fucilare Mario e ottenere le grazie di Tosca;
- fare fucilare Mario e non ottenere le grazie di Tosca;
- non fare fucilare Mario e ottenere le grazie di Tosca;

[13] Prendere decisioni in condizioni di incertezza comporta un rischio: il concetto di rischio è molto importante in quanto in medicina ci collega agli odds e a una forma particolare per esprimere il teorema di Bayes, il rapporto di verosimiglianza (LR o Likelihood Ratio), che vedremo nell'ultima parte, dedicata a teorema di Bayes e decisioni mediche.

- non fare fucilare Mario e non ottenere le grazie di Tosca;

Al momento cruciale Tosca tradisce la sua promessa di cedere a Scarpia, pugnalandolo (e uccidendolo) prima di concederglisi, mentre Scarpia, che ha fatto credere a Tosca che la fucilazione sarebbe stata simulata con i fucili caricati a salve, tradisce anche da morto la sua promessa, avendo predisposto le cose in modo che Mario sia effettivamente fucilato.

L'interesse della vicenda risiede nel fatto che si tratta di un bell'esempio del cosiddetto dilemma del prigioniero[14], e quindi di una applicazione della teoria dei giochi[15]. La situazione è paradossale in quanto presenta un evidente conflitto fra razionalità e utilità. Perseguendo fino in fondo la strategia della massima razionalità dell'esito, ciascuno per sé stesso, i due personaggi finiscono per andare contro il loro migliore interesse/utilità, che sarebbe per Tosca salvare l'amante, e per Scarpia avere i favori di Tosca.

La teoria dei giochi ci dice che decidere in condizioni di incertezza è difficile anche assumendo di avere a che fare con (i) individui completamente razionali, (ii) il cui unico interesse è di perseguire il proprio utile, e che (iii) abbiano una completa informazione (iii_a) sia della situazione in cui si trovano, (iii_b) sia degli effetti delle azioni proprie e altrui. In altre parole ci dice che anche in condizioni ideali la nostra ragione può essere ingannata, e che le nostre scelte e le nostre decisioni ne possono essere conseguentemente falsate.

Considerare che nella vita reale gli individui non sono mai perfettamente razionali e in genere posseggono soltanto informazioni parziali del problema che devono risolvere, ci porta alla seconda direzione, quella delle scienze cognitive, che introducono il concetto di errore cognitivo, cioè di errore legato alla mera fallibilità del ragionamento umano. Il modello classico della razionalità è basato su tre pilastri [2]:
- la logica, intesa come lo studio delle inferenze deduttive valide[16];
- la teoria della probabilità, applicata allo studio delle inferenze induttive (di cui il teorema di Bayes, come vedremo, rappresenta il paradigma)
- la teoria della scelta razionale, che prevede di calcolare l'utilità attesa e di massimizzarla[17].

Causa degli errori cognitivi [3,4,5,6] è il fatto che trovare la soluzione "ottimale" del nostro problema è generalmente troppo difficile se rapportato alla necessità di compiere la scelta in un tempo ridotto. Per questo motivo spesso utilizziamo un insieme di informazioni limitato ed un meccanismo inferenziale subottimale per raggiungere una decisione soddisfacente piuttosto che ricercare l'ottimo. Questo procedimento semplificato, riassunto nel detto popolare *"il meglio è*

---

[14] Proposto negli anni Cinquanta da Albert Tucker come problema di teoria dei giochi.

[15] L'importanza della teoria dei giochi come strumento teorico è stato testimoniato dall'assegnazione nel 1994 del premio Nobel per l'economia a John Nash (ricordato nel film biografico "A beautiful mind"), John Harsanyi e Reinhard Selten.

[16] Ecco un esempio di inferenza deduttiva valida:
- gli eritrociti sono presenti nel sangue di tutti gli uomini (premessa 1)
- Mario è un uomo (premessa 2)
- gli eritrociti sono presenti nel sangue di Mario (deduzione)

[17] Si supponga di dovere scegliere fra due scommesse:
(a) 20% di probabilità di vincere € 25, oppure niente;
(b) 40% di probabiilità di vincere € 10, oppure niente.
Per confrontare le due scommesse, bisogna calcolare l'utilità attesa moltiplicando la vincita per la probabilità di ottenerla. La scommessa (a) ha una utilità attesa di € 5 (0,2 per € 25), la scommessa (b) ha una utilità attesa di € 4 (0,4 per € 10).

nemico del bene", e finalizzato ad evitare di fare la fine dell'asino di Buridano[18], è assolutamente fondamentale nella vita di tutti i giorni ed in molti casi è estremamente più efficiente di un processo di ottimizzazione[19]. Occorre tuttavia essere consapevoli delle sue debolezze. Vediamo quindi alcuni esempi.

PROBLEMA 1
Si supponga che una moneta sia stata lanciata cinque volte, mostrando sempre testa.
Quale è la probabilità che al sesto lancio la moneta segni croce?[20]

PROBLEMA 2
Ecco quattro possibili esiti (T = testa, C = croce) di un lancio di monete, TCCTCTTC, TCTCTCTC, TTTTTTTT, CCTTCCTC.
Quale è la sequenza meno probabile?[21]

Il problema di Monty Hall è un eccellente esempio del fatto che l'intuizione non sempre ci offre la

---

[18] Sufficientemente conosciuto dagli specialisti di filosofia medievale come uno dei più eminenti sorbonagri del secolo XIV, il francese Giovanni Buridano (1295? - 1358?) è poi noto lippis et tonsoribus [ai cisposi e ai barbieri, ovvero a tutti, (Orazio, Satire, 1, 7)] per il famoso esempio dell'asino affamato, che, posto ad eguale distanza da due eguali mucchi di fieno, morirà di fame non riuscendo a decidere a quale dei due rivolgersi per primo (o dell'asino egualmente famelico e sitibondo, che, posto ad eguale distanza da un mucchio di fieno e da un paiolo d'acqua, morirà di fame e di sete non riuscendo a risolvere se prima mangiare o prima bere). La storia dell'asino di Buridano – che però negli scritti del Buridano non si trova; nel commento del filosofo al De Caelo di Aristotele si parla in realtà di un cane; e nel testo aristotelico, che è probabilmente la fonte originale di questa storia, ripresa poi anche da Dante (Par. IV, 1-3), si cita genericamente un affamato e assetato – la storia dell'asino di Buridano, dicevamo, dovrebbe servire a dimostrare l'impotenza della volontà, se essa non sia illuminata dall'intelletto o mossa da sollecitazioni esterne; ma in tutta la cultura occidentale degli ultimi sei secoli "l'Asino di Buridano" è, più genericamente, il tipo dell'irresoluto, che perde una quantità di vantaggi per non saper determinarsi a sceglierne uno.

[19] Quante volte, nel decidere un investimento economico o il luogo dove trascorrere le prossime vacanze, nello stipulare una assicurazione o persino nel fornire un parere professionale, nella scelta di un lavoro o della scuola a cui mandare i nostri figli, ci siamo affidati all'intuizione o, più semplicemente, al buon senso? Niente è più naturale del resto: ogni giorno dobbiamo risolvere decine di problemi piccoli e grandi, nei campi più disparati, con informazioni spesso insufficienti, e con poco tempo a disposizione. Per trarci d'impaccio abbiamo imparato a servirci di una specie di "colpo d'occhio" mentale o, se vogliamo, a prendere delle scorciatoie per riuscire a tagliare il traguardo della soluzione nel tempo massimo che ci è concesso. Quale cacciatore, per esempio esperto di balistica, si metterebbe mai a calcolare con carta e penna la traiettoria del suo proiettile avendo di fronte un leone inferocito? Il cacciatore, e noi con lui, preferisce azzardare una risposta al problema nel più breve tempo possibile, piuttosto che essere sbranato dal problema prima di aver trovato al soluzione giusta. Ma quel "colpo d'occhio"che ci rende quotidianamente tali servigi è anche la fonte principale dei nostri errori.

[20] Poiché nel lancio di una moneta ogni singolo evento è indipendente, e casuale, la risposta è ovviamente: "*la probabilità che al sesto lancio la moneta segni croce è del 50%*". Tuttavia numerose persone ragionano con uno schema mentale secondo cui se si lancia una moneta un certo numero di volte, come risultato si determina una equa distribuzione tra testa e croce. In altre parole, la rappresentatività fa ritenere che il lancio della moneta determini lo stesso numero di teste e croci. Al sesto lancio dunque è comune la previsione che il risultato debba essere croce. La rappresentatività in questo caso spinge all'errore, poiché come nel caso del comportamento dei giocatori d'azzardo si tende a interpretare in modo non corretto la legge delle medie, o "legge dei grandi numeri". In questo caso l'errore cognitivo è quello di ritenere che la legge dei grandi numeri trovi realizzazione anche quando il campione di riferimento è ridotto.

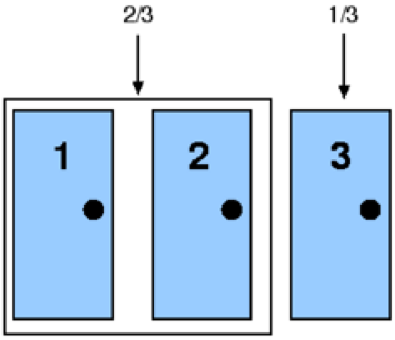
[21] Di nuovo, nel lancio di una moneta ogni singolo evento è indipendente, e casuale, e la sua probabilità è del 50%. La probabilità di una specifica sequenza (si noti che tutte le sequenze indicate sono sequenze di 8 eventi) è pari al prodotto delle probabilità dei singoli eventi, quindi per 8 eventi è sempre identica e, volendola calcolare, è pari a 0,5x0,5x0,5x0,5x0,5x0,5x0,5x0,5 = 0,003906.

soluzione corretta, e quindi del fatto che gli errori cognitivi sono sempre in agguato. Il problema è tratto dal gioco a premi americano *Let's Make a Deal*, e il nome gli deriva da quello del conduttore dello show, Monty Hall appunto. La formulazione del problema è contenuta in una lettera del 1990 di Craig F. Whitaker, indirizzata alla rubrica di Marilyn vos Savant *Ask Marylin* nel *Paradise Magazine*. La vos Savant risolse il problema correttamente; ma alla rivista arrivarono almeno 10.000 lettere, il 92% delle quali indicavano la risposta fornita dalla vos Savant come sbagliata. Notevole scalpore fu determinato dal fatto che alcuni matematici accademici non riconobbero la correttezza della soluzione proposta dalla vos Savant finché questa non la spiegò nel dettaglio in un successivo articolo.

PROBLEMA 3
In un gioco a premi, il giocatore può scegliere tra tre porte, e vince il premio nascosto dietro alla porta che ha scelto.
Le regole del gioco sono le seguenti:
- dietro ciascuna di tre porte c'è o un'automobile o una capra (due capre e un'automobile in tutto); la probabilità che l'automobile si trovi dietro una data porta è identica per tutte le porte;
- il giocatore sceglie una delle porte; il suo contenuto non è rivelato;
- il conduttore sa ciò che si nasconde dietro ciascuna porta;
- il conduttore deve aprire una delle porte non selezionate, e deve offrire al giocatore la possibilità di cambiare la sua scelta;
- il conduttore aprirà sempre una porta che nasconde una capra; ovvero, se il giocatore ha scelto una porta che nasconde una capra, il conduttore aprirà la porta che nasconde l'altra capra; se invece il giocatore ha scelto la porta che nasconde l'automobile, il conduttore sceglie a caso una delle due porte rimanenti;
- il conduttore offre al giocatore la possibilità di mantenere la scelta che ha fatto inizialmente, o di cambiare, scegliendo l'altra porta rimasta chiusa.
Le possibilità di vittoria cambiano per il giocatore se cambia la propria scelta?[22]

La soluzione può essere illustrata come segue [7]. Ci sono tre scenari possibili, ciascuno avente probabilità 1/3:
- il giocatore sceglie la capra numero 1. Il conduttore sceglie l'altra capra. Cambiando, il giocatore vince l'auto;
- il giocatore sceglie la capra numero 2. Il conduttore sceglie l'altra capra. Cambiando, il giocatore vince l'auto;
- il giocatore sceglie l'auto. Il conduttore sceglie una capra, non importa quale. Cambiando, il giocatore trova l'altra capra.

Nei primi due scenari, cambiando il giocatore vince l'auto; nel terzo scenario il giocatore che cambia non vince. Dal momento che la strategia "cambiare" porta alla vittoria in due casi su tre, le *chance* di vittoria cambiando la scelta sono pari a 2/3.

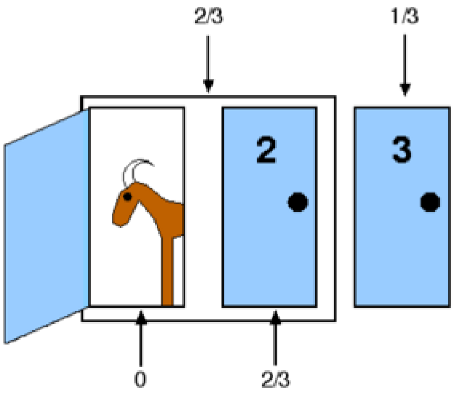


In alternativa la probabilità che l'auto sia dietro la porta restante può essere calcolata con l'ausilio dello schema qui illustrato. Dopo aver (per esempio) scelto la porta 3, il giocatore ha probabilità 1/3 di aver selezionato la porta con l'auto, il che assegna una probabilità pari a 2/3 alle due porte restanti. Si osservi che c'è una probabilità pari a 1 di trovare una capra dietro almeno una delle due porte non selezionate dal giocatore,

[22] La risposta è sì; le probabilità di trovare l'automobile raddoppiano.

dal momento che c'è una sola auto in palio.

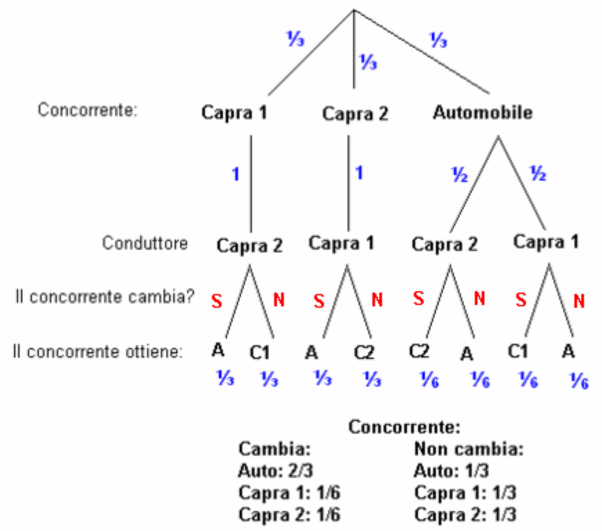
Si supponga che il conduttore apra la porta 1. Dal momento che può solo aprire una porta che nasconde una capra, e non apre una porta a caso, questa informazione non ha effetto sulla probabilità che l'auto sia dietro la porta originariamente selezionata, che resta pari a 1/3. Ma l'auto



non è dietro la porta 1, dunque l'intera probabilità di 2/3 delle due porte non selezionate dal giocatore è ora assegnata alla sola porta 2, come mostrato a fianco. Un modo alternativo per arrivare a questa conclusione è osservare che se l'auto si trova dietro la porta 1 o dietro la porta 2, aprire la porta 1 implica che l'auto si trova dietro la 2.

In maniera più formale, è possibile visualizzare il problema con l'ausilio del diagramma ad albero riportato qui sotto[23].

L'obiezione più comune alla soluzione è fornita dall'idea che, per varie ragioni, il passato possa essere ignorato quando si valutano delle probabilità. Dunque, la scelta della prima porta e il ragionamento del conduttore circa quale porta aprire si possono trascurare; dal momento che si può scegliere tra due porte, la probabilità di scegliere quella giusta dovrebbe essere pari al 50%.



Sebbene ignorare il passato funzioni in certi giochi, come ad esempio nel lancio di una moneta, non funziona necessariamente in *tutti* i giochi. Un rilevante controesempio è fornito dal conteggio delle carte uscite in certi giochi di carte, che consente ai giocatori di sfruttare a proprio vantaggio l'informazione riguardante eventi passati. Questo tipo di informazione è utile nella soluzione del problema di Monty Hall. Inoltre è alla base dei concetti di probabilità a priori (cioè prima di avere acquisito una informazione specifica) e di probabilità a posteriori (cioè dopo avere acquisito una informazione specifica) di un evento, che vedremo nel trattare il teorema di Bayes (che ovviamente può essere applicato alla soluzione del problema di Monty Hall).

Un analogo di questo problema, presentato da Roberto Vacca [8], è il seguente.

[23] Il matematico ungherese Andrea Vàzsonyi ha rianalizzato il problema utilizzando il metodo Montecarlo, simulando gli esiti del gioco con un programma che sceglieva a caso se cambiare scelta o non, e che teneva ovviamente conto di quando si vinceva e quando no. Il risultato ha dimostrato che, effettivamente, cambiando scelta si vinceva in media due volte su tre.

PROBLEMA 4
Abbiamo tre scatole. Nella scatola 1 ci sono due palle di oro. Nella scatola 2 ci sono due palle di argento. Nella scatola 3 ci sono una palla di oro e una palla di argento. Esternamente le scatole sono indistinguibili una dall'altra. Viene scelta una scatola a caso e, senza vederne il contenuto, se ne estrae una palla: è di argento. Ovviamente la scatola non può essere la 1. Non abbiamo altri dati. *Quale è la probabilità che si tratti della scatola 2?*[24]

Errori cognitivi molto più comuni di quello che si può pensare derivano dalla mancata conoscenza dei principi della logica e della matematica [si ignorano/sono sbagliate le regole o non sono applicate correttamente], e più in generale da quella che Douglas R. Hofstadter ha chiamato "innumeracy" (incapacità di lavorare con i numeri). Il termine è stato ripreso in un libro di successo del matematico americano John Allen Paulos. Commette un errore di questo tipo, chi pensa ad esempio, che nel lungo periodo sia possibile giocare alla roulette e vincere. Infine un altro errore cognitivo in cui spesso si incorre è detto "ottimismo irrealistico" ed è l'equivalente del classico "non può capitare a me"[25].

In definitiva le cause fondamentali dell'errore cognitivo appartengono o al dominio delle informazioni o al dominio delle regole, e sono pertanto riconducibili a due fattori:
- *mancano/sono sbagliate o non sono utilizzate correttamente le informazioni*;
-*si ignorano/sono sbagliate o non sono applicate correttamente le regole*[26].

[24] La probabilità è di 2/3, dato che due delle tre palle di argento si trovano nella scatola 2.

[25] L'ottimismo irrealistico produce una sottostima del rischio che si corre personalmente rispetto a una generica persona media. Si tratta di un errore di giudizio che si manifesta in coloro che, coinvolti nella medesima situazione, si collocano in una stato di basso rischio come se si sentissero superiori alla media, meno vulnerabili rispetto agli altri.

[26] A proposito delle basi delle procedure inferenziali che gli individui della nostra specie utilizzano è particolarmente interessante la recensione di Gilberto Corbellini che viene riportata per intero qui di seguito:

"Evoluzione e ragione
Addio al Pleistocene
Commettiamo gli stessi «errori cognitivi» che erano assai utili ai nostri antenati. Una forma di razionalità intuitiva che fa parte della nostra natura - Una raccolta di studi sulle basi biologiche delle credenze e sui modelli normativi per migliorarle

di Gilberto Corbellini (dall'inserto culturale del Sole 24 Ore - domenica 17 dicembre 2006)

Negli ultimi tre decenni le scienze cognitive, in particolare psicologia e antropologia, hanno prodotto una serie di dati empirici piuttosto rilevanti sui condizionamenti naturali e culturali delle modalità di ragionamento umano. Quasi tutti ormai conoscono le scoperte di Kahneman e Tversky, la cui teoria degli errori sistematici e delle euristiche naturali è presa a riferimento sia dagli studiosi dei processi decisionali nei più svariati settori dell'operare umano - ma soprattutto in ambito economico - sia da filosofi ed epistemologi che hanno preferito sviluppare in un senso per così dire strettamente funzionale, piuttosto che evolutivo, le considerazioni di Herbert Simon sulla razionalità limitata. Anche perché, e questo non andrebbe nascosto, mentre con qualche ritardo si celebrano a livello popolare anche in Italia le scoperte di Kahneman e Tversky, se queste ricerche sono esaminate dal punto di vista della storia della metodologia statistica della logica evoluzionistica, come ha fatto Gigerenzer, la loro portata ne risulta alquanto ridimensionata. Rimane comunque fondamentale lo stimolo di queste scoperte a studiare empiricamente le basi biologiche e culturali delle procedure inferenziali che gli individui della nostra specie utilizzano nei contesti più quotidiani, come quando devono decidere se qualcosa è commestibile o meno ovvero valutare se un comportamento è morale o immorale, e in quelli più straordinari, come quando devono giudicare la validità di una teoria scientifica.
Per avere un quadro più composito e articolato del dibattito intorno a questi temi, è particolarmente utile una raccolta di saggi [9] che cerca anche di trarre delle lezioni utili per l'epistemologia dalle scoperte che sono scaturite dallo studio empirico degli stili di ragionamento umano. Si tratta del prodotto finale di una serie di incontri promossi in varie sedi internazionali dalla Fondazione Rosselli e sponsorizzati dalla European Science Foundation intorno al tema «Human reasoning and decision making». L'obiettivo del programma era di trasferire nel campo delle scienze sociali,

È a queste specifiche cause che rivolgeremo la nostra attenzione, non dopo avere svolto, nella prossima parte, una indispensabile premessa in merito al problema gnoseologico.

<u>Bibliografia e riferimenti</u>

[1] www.dif.unige.it/epi/hp/frixione/RAGIONAMENTO2.pdf
[2] Motterlini M, Crupi V. Decisioni mediche. Raffaello Cortina, 2005.
[3] Piattelli Palmarini M. L'illusione di sapere. Mondadori, 1995.
[4] Gigerenzer G. Quando i numeri ingannano. Imparare a vivere con l'incertezza. Raffaello Cortina, 2003.
[5] Crupi V, Genuini GF, Motterlini M. La dimensione cognitiva dell'errore in medicina. Fondazione Smith Kline, Franco Angeli, 2006.
[6] Pravettoni G, Vago G, (a cura di). La scelta imperfetta. Caratteristiche e limiti della decisione umana. McGraw-Hill, 2007.
[7] http://it.wikipedia.org/wiki/Problema_di_Monty_Hall
[8] Vacca R. Anche tu matematico. Garzanti, 1989.
[9] Viale R, Andler D, Hirschfeld L, (a cura di). Biological and Cultural Bases of Human Inference, Lawrance Erlbaum Associates, 2006.

---

in particolare dell'economia e della politica, le conoscenze acquisite dalle ricerche empiriche sui condizionamenti neuropsicologici del ragionamento umano.
Le questioni che vengono affrontate nei saggi da alcuni dei massimi esperti, tra cui Dan Sperber, Scott Atran, Richard Nisbett, Ara Norenzayan, Lawrence Hirschfed, Stephen Stich, Laura Macchi e Riccardo Viale, ruotano, alla fine, intorno alla questione se esistono degli universali cognitivi e se sì come riconoscerli sin dall'infanzia e poi nel loro preservarsi durante lo sviluppo cognitivo, ovvero se gli stili di ragionamento sono culturalmente e geograficamente condizionati, e dobbiamo semplicemente prender atto del relativismo epistemologico oltre che culturale. Problema antico quanto la filosofia occidentale! Che implica quello ancor più complesso di come si può giustificare, validamente, all'interno di un particolare quadro epistemologico, innatista o relativista, una scelta o il cambiamento di una credenza o di una spiegazione.
Il fatto è che, come osserva Viale nell'introduzione, vi sono dati sperimentali che vanno a sostegno di entrambe le posizioni. E per una serie di fraintendimenti, che sono probabilmente il riflesso di un'euristica di comodo, è la posizione relativista quella che risulta più appetibile per gli scienziati sociali. E che oggi rappresenta una comoda, ma purtroppo reale, caricatura contro cui possono scagliarsi tutti gli integralismi che cercano di soddisfare le istanze normative ricorrendo ai dogmi religiosi. Come sfuggire quindi a una deriva relativista e irrazionalista dell'epistemologia, che nega la plausibilità empirica di qualsiasi criterio normativo, senza cadere nel fanatismo ridicolo degli atei devoti? Viale ritiene che comunque sia possibile riconoscere l'esistenza di un principio normativo che viene utilizzato universalmente per decidere, quando è il caso di cambiare una propria credenza o teoria. E che fa uso in vari modi del contenuto empirico della teoria, ovvero di nuove conoscenze acquisite per decidere tra teorie alternative. Si tratta di un principio che ha una sua logica evolutiva. Ma sempre per ragioni evolutive, l'esperienza e il contesto giocano sempre un ruolo importante perché, come dice Wolpert, le caratteristiche del nostro cervello si sono selezionate nella savana del Pleistocene, quando era più adattativo costruirsi credenze assurde e usare l'intuizione. Un approccio diverso, che comincia ad affacciarsi nella ricerca empirica e che nei prossimi anni porterà importanti novità, è quello di mettere queste scoperte in relazione con l'organizzazione funzionale del cervello umano e i processi di integrazione che mette a confronto nei processi decisionali le dimensioni cognitive e quelle emotive dell'esperienza."

3. Il problema gnoseologico

"*La scienza si fa con i fatti come una casa si fa con i mattoni, ma l'accumulazione dei fatti non è scienza più di quanto un mucchio di mattoni non sia una casa.*"
(Henri Poincaré).

"L'immaginazione è più importante della conoscenza".
(Albert Einstein)

Il riconoscimento dell'esistenza di errori cognitivi implica, per motivi meramente logici, che si sappia cosa è la conoscenza. In altre parole è necessario fondare una teoria della conoscenza: ed ecco quindi il problema gnoseologico[27]. Nel quale un ruolo di rilievo è stato assunto dall'analisi logica, scientifica e razionale portata avanti dalla cosiddetta filosofia analitica.[28].

Magia, filosofia e scienza sono le tre tappe storiche sulla via della conoscenza.

La magia non si preoccupa di avere dati misurabili, né di dare loro una forma. Il suo approccio alla comprensione/interpretazione del mondo è basato sulla sola intuizione, e ignora qualsiasi esigenza di oggettività delle fonti della conoscenza. Ma le conseguenze di questo approccio sono rappresentate da un mondo assolutamente irrazionale: l'uomo è in balia di dei capricciosi, di ciclopi, draghi e streghe, dell'influsso degli astri. Per di più non si riesce a trovare la pietra filosofale, e l'alchimia non porta ad alcun risultato. La magia fallisce quindi nel suo obiettivo più ambizioso, quello di dominare, fino a poter cambiare, il mondo reale.

Con la filosofia l'uomo incomincia a ragionare in termini di modelli, attraverso i quali dare forma alla realtà. Modelli che prevedono anche intuizioni geniali, come quello di Democrito, che arriva all'idea di indivisibilità dei costituenti ultimi della materia[29]. La capacità di modellizzare rappresenta un salto in avanti di importanza storica nello sviluppo della cultura umana. E alcuni risultati, come l'incommensurabilità tra misura del raggio e misura della circonferenza, la geometria

---

[27] È bene precisare che nell'ambito della cultura anglosassone la teoria della conoscenza viene chiamata anche epistemologia, laddove in Italia per epistemologia si intende essenzialmente una branca specifica della gnoseologia, ossia quella che si occupa dello studio conoscenza scientifica, conosciuta anche come filosofia della scienza.

[28] Con l'espressione filosofia continentale ci si riferisce generalmente ad una moltitudine di correnti filosofiche del XX secolo, quali la fenomenologia, l'esistenzialismo (in particolare Martin Heidegger), il post-strutturalismo e post-modernismo, decostruzionismo, la teoria critica come quella della Scuola di Francoforte, la psicoanalisi (in particolare Sigmund Freud), ed il Marxismo e la filosofia Marxista. Ci sono differenze talmente grandi tra queste scuole, che sembra impossibile individuare una linea generale condivisa da tutte, che sarebbe la caratteristica principale della filosofia continentale. Le correnti continentali sono così chiamate perché si sono sviluppate soprattutto sul continente europeo, specialmente in Germania e Francia, mentre in Inghilterra e negli Stati Uniti si è sviluppata la cosiddetta filosofia analitica. La differenza principale tra le due correnti sarebbe che la filosofia analitica è piuttosto basata su un'analisi logica, scientifica e razionale che si concentra sui dettagli, mentre la filosofia continentale si occuperebbe di più dei grandi concetti nella loro totalità (ad es. *il senso della vita*) e dei rapporti interpersonali (il rapporto con *l'Altro*, il ruolo dell'Uomo nella società) e sarebbe piú scettica riguardo alle capacità conoscitive della scienza. Recentemente è stato messo in dubbio che vi siano effettivamente grandi differenza tra le due correnti. Si tratterebbe piuttosto di una esagerazione che si concentra sulle posizioni estreme nei due campi, mentre la verità sta nel mezzo. Visto che l'origine della filosofia analitica è situata proprio sul continente, con Frege, Ludwig Wittgenstein, Rudolf Carnap, il positivismo logico del circolo di Vienna, l'empirismo logico di Berlino e la logica polacca, la distinzione sembra non essere molto rilevante.

[29] La sua teoria atomistica prende forma da un immediato ragionamento: se la materia si dividesse all'infinito essa non sarebbe più. Deve esistere una particella indivisibile, dunque, che stia alla base della materia: l'atomo (ἀ-τέμνω, alfa Privativo + voce del verbo greco "tagliare" = indivisibile).

di Euclide [1], le scoperte della scuola filosofica pitagorica riguardanti i rapporti “naturali” tra le note, sono degni del più grande rispetto e della più grande considerazione. In particolare la scoperta di Pitagora che i rapporti “armonici” tra le note sono quelli descritti da rapporti tra interi (1:2 l’ottava, 3:2 la quinta, 4:3 la quarta), fu la prima formulazione matematica di una legge fisica. Ma, per capire in termini storici tali eventi, va ricordato che Pitagora [2], uno dei più grandi “scienziati” dell’era filosofica, non solo rimane una eccezione in un mondo che continuava ad ignorare il concetto di misurabilità come strumento per rendere oggettivi i dati, ma si appassionò di orfismo e misticismo e, in virtù di questi interessi, non esitò ad attribuire ai numeri poteri magici.

Aristotele afferma che "*i corpi leggeri* [come la fiamma, (nota dell'autore)] *si muovono verso l'altro, mentre quelli pesanti si muovono verso il basso*" [3]. Sono necessari 2000 anni prima che Galileo misuri la velocità di caduta dei gravi, con una serie di esperimenti storici. Galileo non solo modellizza, ma sottopone il modello a verifica. Esiste ovviamente il rischio di scoprire che il modello è sbagliato. Ma Galileo accetta il rischio. Non solo, ma svolge lui stesso il lavoro manuale necessario per effettuare le misure richieste. E scopre che i gravi cadono tutti alla stessa velocità. L'insieme di dati misurabili e di una legge matematica quantitativa porta al primo modello scientifico di conoscenza della realtà. Galileo passa meritatamente alla storia come il fondatore del metodo sperimentale. Il metodo che consentirà a Newton, con la sua monumentale opera “*Philosophiae naturalis principia matematica”*, di strutturare definitivamente l'evoluzione in scienza di una branca della filosofia [4].

La possibilità di basarsi su misure anziché sull’intuizione, e il dare forma matematica anziché intuitiva ai dati sono pertanto i due motori del passaggio dall’interpretazione/comprensione magica (e filosofica) all’interpretazione/comprensione scientifica della realtà.

Questi concetti possono essere così riassunti:

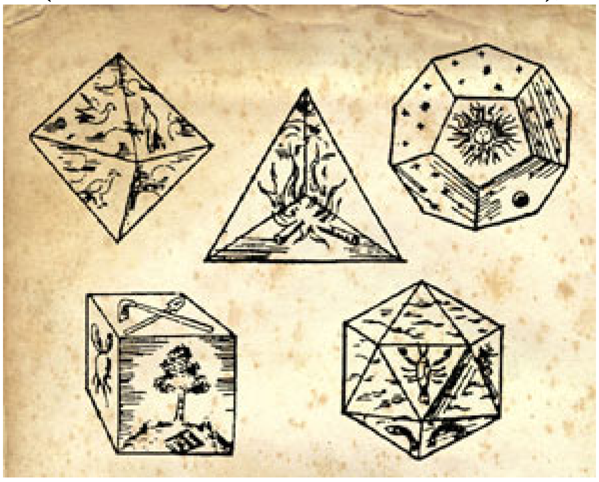
Dati	***Informazione***	***Conoscenza***
Non misurabili	*In forma “intuitiva”, non modellizzata, ignora il criterio di oggettività*	*Magica*
Non misurabili	*In forma di "modello filosofico", basato su modelli ovvero leggi filosofiche qualitative, con al più deboli elementi di oggettività (riscontro nelle “osservazioni”)*	*Filosofica*
Misurabili	*In forma di "modello scientifico", basato su modelli ovvero leggi matematiche quantitative che governano il comportamento della natura, con forti elementi di oggettività (riscontro nelle “misure” effettuate)*	*Scientifica*

Magia e filosofia hanno quindi in comune la caratteristica di essere stadi pre-scientifici della conoscenza. Affidando chi volesse approfondire la descrizione magica del mondo alla lettura dell’opera di James G. Fraser [5], vediamo ora quale è stato il percorso seguito dalla filosofia nel rispondere alle prime due delle tre classiche (e fondamentali) domande:
- “che cosa esiste” (ontologia);
- “cosa possiamo conoscere” (epistemologia);
- “cosa dobbiamo fare” (etica).

Il problema di base, il primo ad essere affrontato, è stato il problema ontologico: “che cosa esiste”. E le prime documentazioni in occidente risalgono ai filosofi presocratici [6]. A Democrito, Pitagora ed Euclide abbiamo già accennato. Dopo che Talete aveva indicato l’*archè* (il principio primo) nell’acqua, Anassimene nell’aria ed Eraclito nel fuoco, Empedocle, aggiungendo la terra, pone come princìpi quattro corpi semplici: l’acqua, l’aria, la terra e il fuoco. Oggi sappiamo che l’idea che la natura sia spiegabile come composizione di elementi base è presente in culture che si sono sviluppate in modo indipendente [3] e che l’idea che la natura in tutte le sue manifestazioni sia spiegabile attraverso una particolare mescolanza di quattro irriducibili elementi ha dominato in occidente il sapere fino alla fine del ‘700, quando con la scoperta dell’ossigeno da parete di Lavoisier nel 1778 è nata la chimica moderna.

Nonostante non abbia lasciato alcuno scritto, e il suo pensiero lo si debba ricavare dalle opere dei discepoli, Socrate [7] è stato uno dei più importanti esponenti della tradizione filosofica occidentale. Il contributo più importante che egli ha dato alla storia del pensiero filosofico consiste nel suo metodo d'indagine: il dialogo che utilizzava lo strumento critico dell'*elenchos* (confutazione). Molti studiosi di storia della filosofia concordano nell'attribuire a Socrate la nascita di quel peculiare modo di pensare che ha consentito l'origine e lo sviluppo della riflessione astratta e razionale, che sarà il fulcro portante di tutta la filosofia greca successiva[30]. Il primo a sviluppare questa interpretazione della dottrina socratica fu Aristotele che attribuì a Socrate la scoperta del metodo della definizione e induzione, che egli considerava uno degli assi portanti del metodo scientifico.

Il merito di avere formulato la prima teoria organica della conoscenza spetta a Platone [8], che afferma che ogni atto conoscitivo è sempre una reminiscenza: la conoscenza deriva non dall’esperienza, bensì da un sapere preesistente, prenatale, e connaturato all’uomo. La conoscenza è quindi un processo di anamnesi, con cui si ricostruisce la verità attraverso il ricordo. Dall’ambiente, ossia dalla percezione sensoriale del mondo, può venire al massimo uno spunto, uno stimolo alla rievocazione. Esiste pertanto un *dualismo metafisico* (il mondo sensibile e il mondo ideale), cui corrisponde un *dualismo gnoseologico* (la percezione delle cose, cioè il vedere e il sentire, che producono solo un’opinione più o meno falsa, e la scienza, cioè il sapere delle idee del filosofo) [3]. Un altro contributo di Platone è aver cercato, nel Timeo (il dialogo dedicato alla scienza), di riassumere in una sintesi unitaria tutte le intuizioni scientifiche a lui precedenti. L’assunto da cui parte Platone è la perfezione dei cinque poliedri regolari convessi [9], che godono di due proprietà: (i) hanno lati, facce ed angoli uguali, e (ii) sono perfettamente inscrivibili in una sfera e perfettamente circoscritti da una sfera[31]. Ipotizzando che questi solidi rappresentino le forma di base della materia, Platone spiegò le proprietà fisiche degli elementi: la terra è stabile e compatta perché formata da cubi, strutture



---

[30] Al contrario dei sofisti che usavano il monologo e che praticamente parlavano da soli, il suo discorrere era un *dià logos*, una parola che *attraversava* i due interlocutori . Mentre i sofisti infatti miravano ad abbindolare l'interlocutore usando il *macròs logos* , il grande e lungo discorso che non dava spazio alle obiezioni, Socrate invece dialogava con brevi domande e risposte - la cosiddetta *brachilogia* (letteralmente breve dialogare) socratica - proprio per dare la possibilità di intervenire e obiettare ad un interlocutore che egli rispettava per le sue opinioni. L'altra caratteristica del dialogo socratico, che lo distingueva dal discorso torrentizio dei sofisti, era il continuo domandare di Socrate su quello che stava affermando l'interlocutore, alla ricerca di una precisa definizione dell'oggetto del dialogo. *Ti estì*. che cos'è quello di cui tu parli? Chi dialoga con Socrate tenterà varie volte di dare una risposta precisa ma alla fine si arrenderà e sarà costretto a confessare la sua ignoranza [7].

[31] Si può dimostrare che esistono solamente 9 poliedri regolari. I cinque poliedri regolari convessi, chiamati anche solidi platonici [10], sono il tetraedro regolare, il cubo (o esaedro regolare), l'ottaedro regolare, il dodecaedro regolare e l'icosaedro regolare. E’ ovviamente notevole il fatto che fossero già conosciuti ai greci. I poliedri regolari non convessi sono quattro. Due di essi, i così detti poliedri di Keplero (1571-1630) hanno come facce poligoni regolari stellati; altri due, i così detti poliedri di Poinsot, dal nome del matematico francese Louis Poinsot, (1777-1859) sono costruiti in modo che le facce possano interpenetrarsi.

massicce che tendono a compattarsi in blicchi solidi e immobili; il fuoco è costituito da tetraedri mobili e instabili, mentre l'acqua l'aria e l'etere sono formati rispettivamente icosaedri, ottaedri e dodecaedri, forme progressivamente tendenti alla sfericità e quindi proprie di sostanze via via sempre più liquide, aeree, volatili.

Alla svalutazione del mondo materiale operata da Platone, Aristotele [11] sostituisce una forte attenzione per la realtà naturale e per la scienza. Di fatto nel Liceo, la scuola filosofica fondata da Aristotele appena fuori dalle mura di Atene, tutte le scienze furono sviluppate attraverso un lavoro di ricerca collettivo e coordinato, il cui obiettivo era quello di costruire una enciclopedia del sapere: riordinando in un quadro strutturato la massa di conoscenze specifiche fino ad allora disperse fra un grande numero di specialisti [3]. Dal punto di vista metodologico Aristotele vuole procurarsi basi solide, e le individua (i) nel *principio di non contraddizione* ("è impossibile che la stessa cosa sia e insieme non sia") e (ii) nel *sillogismo*. Con la teoria del sillogismo di Aristotele nasce la *logica*, la scienza della dimostrazione capace di indicare con esattezza quando e perché un ragionamento è valido (ben costruito, e coerente rispetto alle premesse), o falso[32]. Sebbene sia possibile costruire sillogismi di quattro tipi diversi, quello relativo alla mortalità di Socrate è un esempio di *sillogismo perfetto*, in quanto dotato di evidenza palese, tale da non richiedere ulteriori riflessioni. Il sillogismo perfetto è il fondamento intuitivo dell'intera logica; i ragionamenti esprimibili in questa forma sono tanto semplici e certi da non richiedere ulteriori conferme. Il lavoro del logico aristotelico consisteva quindi nel verificare la trasformabilità di una proposizione qualsiasi (incerta, complicata) in un sillogismo perfetto (o in altri equivalenti), Ciò assumeva il carattere di prova: se la riduzione era possibile, il ragionamento doveva essere accettato [3].

L'utilizzo del sillogismo nella pratica della ricerca scientifica poneva dei problemi. Infatti passando dalla logica (che bada solamente alla coerenza interna del discorso) alla ricerca scientifica diventa essenziale *il problema della verità delle premesse*. Un sillogismo scientifico, oltre a usare correttamente le regole deduttive, deve partire da presupposti giusti. L'idea di dimostrare questi ultimi tramite un nuovo sillogismo non fa altro che spostare a monte il problema, creando una catena che peraltro non può diventare infinita. La conoscenza delle premesse (per definizione non dimostrabili) è la questione più delicata della dottrina aristotelica della scienza. Secondo Aristotele è possibile cogliere le verità fondamentali su cui fondare ogni scienza attraverso due vie:

[32] Aristotele osservò innanzitutto che non tutto il pensiero è ragionamento: un giudizio preso isolatamente (per esempio "oggi piove") può essere vero o sbagliato rispetto alla realtà, ma non possiede alcun valore logico, in quanto è semplicemente la constatazione di un dato di fatto. Il ragionamento nasce quando il pensiero passa di giudizio in giudizio, colleganto fra loro le proposizioni con nessi necessari, in modo che le conclusioni seguano obbligatoriamente le premesse. In questa consequenzialità, per cui la proposizione antecedente è causa delle conseguenti, consiste la coerenza logica. Un ragionamento minimo (sillogismo) deve quindi consistere in almeno tre proposizioni: (i) la premessa maggiore, (ii) la premessa minore, (iii) la conclusione. Per fare un esempio pratico:
- Tutti gli gli uomini sono mortali (premessa maggiore)
- Tutti i greci sono uomini (premessa minore)
- Quindi tutti i greci sono mortali (conclusione)

Quando il termine medio (nell'esempio "uomini") è soggetto della premessa maggiore e predicato della premessa minore, come qui sopra, il sillogismo è nella sua forma perfetta.
Un sillogismo è considerato valido se un qualsiasi ragionamento di quella forma è sempre valido. Quindi il silllogismo:
- Alcuni uomini sono italiani
- Alcuni uomini sono biondi
- Quindi alcuni italiani sono biondi

non è valido anche se tutte le sue proposizioni sono vere, perché il corrispondente sillogismo, diverso ma della stessa forma:
- Alcuni esseri viventi sono uomini
- Alcuni esseri viventi sono elefanti
- Quindi alcuni uomini sono elefanti

non conclude correttamente.

l'induzione e l'intuizione. L'induzione è il procedimento per cui si generalizza una verità generale partendo dall'analisi di casi particolari. E' una forma di razionalità non sillogistica (e neppure logica in senso stretto); è la constatazione (universalizzata) di innumerevoli verità particolari. L'intuizione è la capacità dell'intelletto di cogliere una verità immediata, tanto ovvia ed evidente da non richiedere alcuna dimostrazione. In questo modo la matematica parte dalla fondazione per via intuitiva dei suoi oggetti (il numero, il pari e il dispari, le nozioni di somma, sottrazione e così via). La geometria parte dai cinque postulati di Euclide, e le varie scienze da quei principi generalissimi che sono universalmente ammessi dagli studiosi [3]. Per comprendere le conseguenze di questo approccio, ecco alcune delle teorie di Aristotele, ovvero le conseguenze di alcune di esse:
- esiste solo la materia, il vuoto non esiste;
- l'universo è delimitato, compatto, chiuso all'interno di un ultimo luogo, l'Empireo;
- la terra è al centro dell'universo;
- le specie sono eterne, immutabili e si tramandano immutate di generazione in generazione;
- la biologia è interessata solamente alle specie e non agli individui;
- la forma degli organi è sempre finalizzata alle funzioni;
- per quanto riguarda la procreazione, non si deve parlare di maternità, ma di paternità soltanto; il ruolo della donna si riduce a offrire in modo passivo un ambiente adatto (l'utero) allo sviluppo dello sperma in feto[33].

Anche se la dottrina aristotelica rapidamente si impone, alcuni dubbi rimangono, e alla base stessa della logica. Un esempio solo apparentemente banale di contraddizione logica, è fornito da uno dei tanti paradossi[34] individuati da Bertrand Russel[35]. Si supponga che per gli abitanti dell'isola di Creta esistano solo due alternative: o affermano tutti sempre e solamente il vero, o affermano tutti sempre e solamente il falso. Si consideri ora la domanda "Un cretese dice che tutti i cretesi mentono: egli mente o non mente?". Se gli abitanti dell'isola di Creta affermano tutti sempre e solamente il vero, il cretese, che per definizione dice il vero, avrà mentito dicendo che tutti i cretesi mentono. Se gli abitanti dell'isola di Creta affermano tutti sempre e solamente il falso, il cretese, che per definizione dice il falso, avrà detto la verità. Anche il paradosso del sorite, attribuito a Zenone, crea non pochi problemi logici: quale è il granello che fa passare un mucchio di sabbia in un non-mucchio[36]? Problemi di una modernità incredibile, se si pensa ad un equivalente moderno:

---

[33] Le teorie aristoteliche furono in grado di influenzare persino il primo pensiero scientifico moderno. Nel 1677 l'olandese A. van Leeuwenhoek, esaminando per la prima volta il liquido seminale umano al microscopio, individuò gli spermatozoi, e li descrisse (li vide e li disegnò) come "piccoli uomini in miniatura [homuncoli] che si muovono molto graziosamente".

[34] Un paradosso, dal greco parà (contro) e doxa (opinione), è qualcosa che sfida l'opinione comune: si tratta, infatti di "*una conclusione apparentemente inaccettabile, che deriva da premesse apparentemente accettabili per mezzo di un ragionamento apparentemente accettabile*".

[35] Come riportato da J. D. Barrows nella sua opera "Il mondo dentro il mondo" Russel dice "*Sulle prime pensai che sarei riuscito a venire a capo delle contraddizioni senza troppe difficoltà… Ma gradualmente divenne chiaro che le cose non stavano così… Per tutta l'ultima parte del 1901 continuai a credere che la soluzione sarebbe stata semplice ma verso la fine di quell'anno avevo ormai concluso che si trattava di un'impresa ardua… Passammo le estati del 1903 e del 1904 a Churt e a Tilford… passammo gli inverni a Londra, e in quei periodi non tentai di lavorare; ma le due estati del 1903 e del 1904 mi sono rimaste in mente come fasi di completa paralisi intellettuale. Mi era chiaro che non sarei potuto procedere senza prima avere risolto le contraddizioni, ed ero deciso a non permettere ad alcuna difficoltà di distogliermi dal completamento dei Principia Mathematica; ma sembrava molto probabile che avrei passato il resto della mia vita a fissare il foglio bianco, A rendere la situazione ancor più irritante era il fatto che le contraddizioni apparivano banali, e che passavo il mio tempo a pensare su questioni che non apparivano degne di seria attenzione…*".

[36] L'assioma universale aristotelico, sottostante a ogni forma di pensiero razionale, è il "principio di non contraddizione". Esso afferma che "è impossibile che la stessa cosa sia e insieme non sia". Se è dato un A, allora ogni B sarà diverso da A. A o non-A. Tertius non datur. O bianco o nero. Eppure già Zenone mette in crisi l'assioma. Raccoglie un granello di sabbia da un mucchio e chiede se il mucchio è ancora un mucchio. Zenone non trova il

quale è il risultato di una specifica analisi di laboratorio che fa passare il soggetto dallo stato di "malato" a quello di "sano"?

Indiscutibilmente tutto il pensiero occidentale è figlio del pensiero greco. Per l'intero medioevo i solidi platonici e la "numerologia" pitagorica saranno tra le basi per la sopravvivenza del pensiero magico. Mentre il dualismo platonico e il metodo e il dubbio socratici anticipano addirittura il pensiero di Cartesio. Inoltre il metodo socratico della definizione e dell'induzione aprono problemi modernissimi, che troveranno una soluzione solamente nel XX secolo. Ma in realtà è il progetto enciclopedico di Aristotele a segnare una svolta che di fatto cambierà la storia: perché per i successivi 2000 anni la dottrina aristotelica condizionerà, e di fatto ingesserà, tutto il pensiero del mondo occidentale, fino alla svolta impressa da Galileo Galilei [12].

Galileo propone di abbandonare la teoria aristotelica delle forme sostanziali[37], e propone di affrontare il problema della conoscenza partendo dallo studio delle "affezioni" delle "sustanze naturali"[38]. Che non sono i colori, i sapori, i suoni, gli odori: caratteristiche non del corpo ma del suo rapporto con la nostra facoltà sensitiva, tolta la quale cesserebbero di esistere, o di esistere così[39]. Bensì affezioni primarie, necessità che attengono al corpo molto più dell'aristotelica forma sostanziale: che sono "il luogo, il moto, la figura, la grandezza, l'opacità, la mutabilità". Qualità primarie, oggettive, cioè pertinenti al corpo in modo indipendente dall'osservazione[40]. Perché, ecco

---

granello di sabbia che cambia il mucchio in un non-mucchio. Né il granello di sabbia che cambia il non mucchio, formato dal primo granello di sabbia sottratto al mucchio di partenza, in un mucchio. Il mondo reale sembra essere dominato, più che dal bianco e dal nero, dall'infinita gamma di grigi che li collegano.

[37] Per la parte da Galileo a Kant è stato seguito da vicino il lavoro "Lo sviluppo della gnoseologia moderna" curato da Giovanno Boniolo [13] e pubblicato nella biblioteca di SWIF (Sito Web Italiano per la Filosofia).

[38] "*O noi vogliamo specolando tentar di penetrare l'essenza vera ed intrinseca delle sustanze naturali; o noi vogliamo Contentarci di venir in notizia di alcune loro affezioni. Il tentar l'essenza, l'ho per impresa noon meno impossibile e per fativa non men vana nelle prossime sustanze elementari che nelle remorissime e celesti: e a me par essere ugualmente ignarodella sustanza della Terra che della Luna, delle nubi elementari che delle macchie del Sole; né veggo che nell'intender queste sostanze vicine aviamo altro vantaggio che la copia de' particolari, ma tutti egualmente ignoti, per i quali andiamo vagando, trapassando con pochissimo o niuno acquisto dall'uno all'altro. E se, domandando io qual sia la sustanza delle nugole, mi sarà detto che è un vapore umido, io di nuovo desidererò sapere che cosa sia il vapore; mi sarà per avventura insegnato, esser acqua, per virtù del caldo attenuata, ed in quello resoluta; ma io, egualemente dubbioso di ciò che sia l'acqua, ricercandolo, intenderò finalmente, esser quel corpo fluido che scorre per i fiumi e che noi continuamente maneggiamo e trattiamo: ma tal notizia dell'acqua è solamente più vicina e dipendente da più sensi, ma non più intrinseca di quella che io avevo per avanti delle nugole. E nell'istesso modo non più intendo della vera essenza della terra o del fuoco, che della Luna o del Sole; e questa è quella cognizione che ci vien riservata da intendersi nello stato di beatitudine, e non prima.*" (G. Galilei, Istoria e dimostrazioni intorno alle macchie solari, 1613, in Opere, vol. V., pp. 187-8).

[39] "*Per lo che vo io pensando che questi sapori, odori, colori, etc., per la parte del suggetto nel quale ci par che risegganno, non sieno altro che puri nomi, ma tengano solamente lor residenza nel corpo sensitivo, sì che rimosso l'animale, sieno levate ed annichilate tutte queste qualità.*" (G. Galilei, Il Saggiatore, 1623, in Opere, vol. VI., p. 327).

[40] "*Ma se vorremmo fermarci nell'appressione di alcune affezioni, non mi par che sia da desperar di poter conseguirle anco ne i corpi lontanissimi da noi, non meno che nei prossimi, anzi tal una per avventura più esattamente in quelli che in questi. E chi non intende meglio i periodi de i movimenti de i pianeti, che quelli dell'acque di diversi mari? chi non sa che molto prima e più speditamente fu compresa la figura sferica nel corpo lunare che nel terrestre? e non è egli ancora controverso se l'istessa Terra resti immobile o pur vadia vagado, mentre che noi siamo certissimi de i movimenti di non poche stelle? Voglio per tanto inferire, che se bene indarno si tenterebbe l'investigazione della sustanza delle macchie solari, non resta però che alcune loro affezioni, come il luogo, il moto, la figura, la grandezza, l'opacità, la mutabilità, la produzione ed il dissolvimento, non possino da noi esser apprese, ed esserci poi mezi a poter meglio filosofare intorno ad altre più controverse condizioni delle sustanze naturali; le quali poi finalmente sollevandoci all'ultimo scopo delle nostre fatiche, cioè all'amore del divino Artefice, ci conservino la speranza di poter apprender in Lui, fonte di luce e di verità, ogn'altro vero.*" (G. Galilei, Istoria e dimostrazioni intorno alle macchie solari, 1613, in Opere, vol. V., pp. 187-8).

il punto cruciale, solo questo genere di affezioni può essere descritto in forma matematica: non le sostanze aristoteliche, non i sapori o gli odori, ma “il luogo, il moto, la figura, la grandezza, l'opacità, la mutabilità”[41].

Ecco quindi l’idea di Galileo: collegare qualità primarie e descrizione matematica del mondo, esperienza e teoria, “sensate esperienze” e “certe dimostrazioni” in una correlazione dinamica, costantemente sottoposte le une al vaglio delle altre. E in questo equilibrio geniale Galileo costruisce le condizioni per la genesi di un nuovo modello di conoscenza scientifica.

La combinazione di teoria ed esperienza è possibile perché si sono preventivamente adattate le affezioni scientificamente rilevanti e le modellizzazioni teoriche, entrambe saldate da una comune struttura matematica. Tutto il sistema galileiano si regge su una fiducia assoluta sulla matematica come strumento razionale. Ma su che cosa si basa tale fiducia? Da chi, come e perché il mondo è scritto in lingua matematica? Perché, anche ammesso che sia vero, la matematica fornisce una conoscenza qualitativamente superiore a quella offerta dalla scienza aristotelica? Non avviene questo a prezzo di una riduzione e forse anche di una semplificazione di ciò che può essere oggetto di conoscenza scientifica? Domande di questo tipo mostrano che l'affermazione della nuova scienza richiede una revisione ancora più profonda dei presupposti e delle categorie del pensiero moderno. Accettare questa scienza significa rifondare dalle basi l'edificio della conoscenza umana, e per far questo occorre un laborioso impegno filosofico. Sarà Cartesio [14] il primo a porvi mano.

Nel *Discorso sul metodo*, Cartesio traccia le linee del “vero metodo con cui pervenire alla conoscenza di tutte le cose di cui il mio spirito era capace”, che riassume in quattro regole[42]. Poi, come suggerito dalla prima regola del metodo, Cartesio inizia a dubitare dalla conoscenza sensibile per passare a dubitare dell'esistenza di qualcosa di esterno che stimola questa conoscenza, e infine della stessa conoscenza intellettuale. Della prima si deve dubitare perché se i sensi hanno sbagliato una volta, come è stato, non è detto che non sbaglino ancora. Della realtà esterna dubitiamo per l'illusione vivissima, talvolta, che il sogno porta con sé, facendoci apparire presenti e vive realtà del tutto immaginarie. Ma dal sogno, con un'accelerazione iperbolica, Cartesio passa a dubitare anche delle certezze che nessuno potrebbe discutere: anche l'evidenza che due più tre fanno cinque, o che il quadrato ha quattro lati può essere messa in discussione ipotizzando l'esistenza di un genio maligno e potente che ci inganni ogniqualvolta sommiamo due a tre o contiamo i lati delle figure geometriche. Iperbolico quanto possibile, tale dubbio inficia la certezza della conoscenza intellettuale e matematica, ma non arriva a produrre l'esito scettico per cui di nulla siamo certi. Infatti, pur se esiste un tale genio maligno che mi inganna sempre “*se dubito penso, se penso esisto*”[43]. In questo

---

[41] “*La filosofia è scritta in questo grandissimo libro che continuamente ci sta aperto innanzi a gli occhi (io dico lo universo), ma non si può intendere se prima non s'impara a intender la lingua, a conoscer i caratteri ne' quali è scritto. Egli è scritto in lingua matematica, e i caratteri son triangoli, cerchi, ed altre figure geometriche, senza i quali mezi è impossibile a intenderne umanamente parola; senza questi è un aggirarsi vanamente per un oscuro labirinto.*” (G. Galilei, Il Saggiatore, 1623, in Opere, vol. VI, p. 295

[42] “*La prima era di non accettare mai per vera nessuna cosa che non riconoscessi tale con evidenza, cioè di evitare diligentemente la precipitazione e la prevenzione e di non comprendere nei miei giudizi nulla più di quanto si presentasse così chiaramente e distintamente al mio spirito, da non lasciarvi alcuna occasione di dubbio. La seconda era di suddividere ogni difficoltà che esaminavo nel maggior numero di parti possibili e necessarie per meglio risolverla. La terza di condurre per ordine i miei pensieri, cominciando dagli oggetti più semplici e più facili da conoscere, per salire a poco a poco, come per gradi, fino alla conoscenza dei più complessi, presupponendo un ordine anche tra gli oggetti che non si precedono naturalmente l'un l'altro. E l'ultima era di fare ovunque enumerazioni così complete e revisioni così generali da essere sicuro di non omettere nulla.*” (Cartesio, 1637, pp. 144-5).

[43] “Non c'è dubbio che io esista, se egli mi inganna; e mi inganni fin tanto che vorrà, non potrà mai fare che io sia nulla, nel momento in cui penserò di essere qualcosa. In tal modo, dopo aver ben pensato ed esaminato tutte le cose, devo infine

passaggio essenziale prende corpo lo spostamento moderno dall'essere al pensiero, la dipendenza dell'ontologia dalla teoria della conoscenza e dai suoi risultati. Solo ciò che è certo è. La prima regola del metodo ha dato il suo primo risultato. Nella concezione di Cartesio la *res cogitans* accorpa funzioni non solo cognitive (dubitare, concepire) ma anche morali (affermare, negare, volere), sensibili e immaginative. Se poi consideriamo "che cosa sia" ciò che viene attestato dalle idee avventizie e dai sensi sorprende la semplicità della *res extensa* rispetto alla varietà della *res cogitans*. Essa è pura estensione, spazialità[44]. Cartesio, proseguendo il cammino aperto da Galileo, rafforza il valore di matematica e geometria, che diventano modelli metodologici espliciti, generalizzati e giustificati nell'impiego che ne fa. Tuttavia quando spostandosi sulla *res extensa* Cartesio tenta una fondazione della fisica, essa gli riesce a metà, cioè solo per quanto attiene all'estensione. La posizione cartesiana diventa estremamente importante per cogliere la svolta che il problema della conoscenza compie nell'epoca moderna: *la fondazione dell'ontologia sulla gnoseologia è infatti la cifra costitutiva della epistemologia moderna.*

Che cosa accade infatti se, pur riconoscendo il valore della scienza moderna come modello per la conoscenza, non si accetta il dualismo cartesiano? Come si spiega la conoscenza se questa non è il rapporto tra due *res*? E' questo lo scenario in cui si colloca il pensiero di Thomas Hobbes [15]. Per Hobbes la filosofia consiste nella ricerca delle cause, e ciò significa, per l'orizzonte rigorosamente materialista in cui si colloca, ricerca della generazione[45]. Tuttavia il duplice percorso che presenta, discensivo dalle cause agli effetti, ascensivo dagli effetti alle cause, rimanda in realtà a due ben diverse accezioni del sapere. La prima via è a priori, la seconda a posteriori: solo la prima porta a conoscenze certe, mentre la seconda porta a congetture, essendo possibili più cause per uno stesso effetto[46]. Questa fondamentale assunzione sembra dividere in modo netto le sorti epistemologiche della geometria e della matematica da quelle della fisica e di tutte le altre scienze della natura, deduttive le prime perché costruite da noi, induttive e congetturali le seconde perché la loro genesi ci è estranea.

Da questo momento il problema gnoseologico imbocca due vie: l'empirismo, la corrente filosofica per la quale solo l'esperienza offre una base solida di conoscenza, e il razionalismo, per il quale l'evidenza dei principi che rendono intelligibile la realtà non sarebbe di tipo empirico. Testimonia la difficoltà di abbracciare in modo radicale l'una contro l'altra la figura di John Locke [16]. In Locke l'esperienza è il fondamento della conoscenza, ciò che permette di scrivere su quel foglio bianco che è il nostro intelletto. Ma partito da posizioni di empirismo moderato Locke, circa vent'anni dopo, sostiene che la scienza non può limitarsi alla raccolta empirica di dati e alla generalizzazione delle idee semplici, ma deve mirare a ottenere quella certezza che posseggono le proposizioni matematiche, legate a una conoscenza intuitiva e dimostrativa. Il rigore di un sapere costruito deduttivamente a partire da verità evidenti e universali secondo Locke si ha solo nella matematica - e nella morale. Esso rappresenta

---

concludere e tener per certo che questa proposizione: *io sono, io esisto*, è necessariamente vera ogni volta che la pronuncio o la concepisco nel mio spirito" (Cartesio, 1641, p. 204).

[44] "*La natura della materia, cioè del corpo, in generale non consiste nell'essere una cosa dura o colorata o che tocca in qualche modo i sensi, ma soltanto nell'essere una cosa estesa in lunghezza, larghezza e profondità.*" (Cartesio, 1644, II, IV, p. 639).

[45] "*La filosofia è conoscenza, acquisita con retto ragionamento, degli effetti o fenomeni partendo dai concetti delle loro cause o generazioni e ancora delle generazioni, che possono aver avuto luogo, partendo dalla conoscenza degli effetti.*" (Hobbes, 1655, I, 2).

[46] "*Due sono i metodi scientifici, uno dalla generazione ai possibili effetti, l'altro dagli effetti che si manifestano alla loro possibile generazione. Nel primo caso, siamo noi stessi a rendere veri i principi stessi del ragionamento, vale a dire le definizioni, attraverso il consenso circa i nomi delle cose. E questa prima parte l'ho svolta nelle pagine precedenti […] Affronto ora la seconda parte, che va dai fenomeni o effetti naturali a noi noti mediante il senso, alla ricerca di un qualche modo secondo il quale, non dico furono generati, ma avrebbero potuto esserlo. Pertanto i principi da cui dipendono le cose che seguono non li abbiamo fatti noi, né li diamo per universali, come le definizioni, ma li osserviamo in quanto posti nelle cose stesse dal creatore della natura.*"

quindi un ideale epistemologico che sposta l'asse del pensiero lockiano da un dichiarato empirismo di partenza a una sempre più articolata combinazione di esperienza e ragione, di percezione sensibile e conoscenza intellettuale.

Ma non accetta il compromesso David Hume [17]. Per Hume solo la conoscenza che ha relazione con l'esperienza è valida, anche se si tratta di un sapere che manca di necessità e universalità. La principale associazione di idee su cui costruiamo il nostro sapere, cioè il principio di causa[47], è il cuore della critica humeana e il ponte per negare che vi sia la possibilità di una conoscenza necessaria che parta dalle impressioni sensibili. Noi abbiamo a che fare solo con fasci di impressioni distinte: il resto è solo costruzione arbitraria e probabile. Una posizione allora accusata di indurre l'uomo al totale scetticismo, ma che di fatto anticipa la rivoluzione che sarà determinata dagli sviluppi della scienza nel XIX e nel XX secolo.

Le conclusioni di Hume sono rafforzate da Stuart Mill [18]. Gli assiomi della logica, ovvero le verità indiscutibili ed evidenti come il fatto che gli alberi sono fatti di legno, si fondano non su una relazione necessaria con qualche essenza metafisica (al di là del mondo dell'esperienza) ma sul processo di generalizzazione delle osservazioni empiriche proprio dell'induzione, la quale da un numero ripetuto di osservazioni identiche, fonda la legge che accomuna queste osservazioni. Il processo di induzione permette di far derivare teorie valide generalmente dall'osservazione ripetuta di un certo fenomeno. Ad esempio, tutte le fragole sono rosse perché viene osservato in natura il fatto ripetuto per cui le fragole mature hanno una colorazione rossa. Stuart Mill critica la possibilità propria dell'induzione di essere portatrice di verità assolute. La certezza propria dell'induzione di essere un processo che va a formare verità generali, viene meno in quanto l'induzione non può escludere che un singolo caso che va a formare la legge generale invalidi da un momento all'altro la teoria con il suo comportamento contrario a quello degli altri casi. In sostanza, la possibilità che un'esperienza si ripeta in modo assoluto date certe premesse non può essere garantita da alcuna conoscenza certa, poiché le certezze della scienza si basano sul meccanismo incompleto dell'induzione, che non potrà mai affermare la validità assoluta di un evento e di un fatto, poiché non potrà mai controllare la totalità dei casi, dovendo escludere necessariamente dal computo quelli che ancora non sono accaduti.

Abbiamo visto come in definitiva le principali discussioni nella comunità filosofica tra XVI e XVIII secolo vertevano sul tema della conoscenza. Vederne gli aspetti è stato un modo per riflettere sul ruolo strategico che assume la definizione di conoscenza nel progetto filosofico e nella stessa giustificazione del sapere scientifico.

La rivoluzione "copernicana" di Immanuel Kant [19] pone al centro "cosa possiamo conoscere". Secondo Kant il mondo non si dà a noi, pronto con le sue leggi, ma siamo noi che lo *costituiamo* grazie a concetti e principi a priori (e questo è cosa del tutto diversa da una sua produzione totale che è del tutto esclusa all'uomo in quanto l'*intellectus archetypus* necessario è posseduto solo da Dio). La possibilità di seguire la via appena indicata comporta che vi siano "due tronchi dell'umana conoscenza": la sensibilità e l'intelletto. Dove la conoscenza sensibile è immediata, recettiva, avviene per intuizioni ed è basata su affezioni; mentre la conoscenza intellettuale è spontanea, avviene per concetti ed è basata su funzioni. Questo significa che benché i concetti intellettuali, da

[47] La critica di Hume al principio di causalità comincia con una ormai nota distinzione: "*Due sono le questioni che mi accingo a esaminare: 1. Per quale ragione diciamo necessario che tutto ciò che ha un cominciamento debba avere una causa? 2. Perché affermiamo che certe cause particolari debbono necessariamente avere certi particolari effetti? Qual è la natura di quest'inferenza, per cui passiamo dalle une agli altri, e della credenza che riponiamo in essa?*". Il primo è un problema che riguarda la fondazione del principio di causalità, mentre il secondo riguarda la fondazione dell'inferenza induttiva, ma entrambi i problemi sono affrontati da Hume con il chiaro scopo di mostrare che non vi è alcuna possibilità di fondare in maniera indiscutibile il principio di causalità e il procedimento induttivo, ma che anche questi, come tutto il sapere, sono basati sull'esperienza e su una sua generalizzazione arbitraria.

una parte, e le intuizioni empiriche, dall'altra, siano egualmente necessari e complementari per il fine conoscitivo, sono i concetti che permettono di trasformare le intuizioni empiriche in conoscenze vere e proprie[48]. Come afferma Mauro Dorato "*Da Kant in poi, le classiche domande filosofiche su "che cosa esiste" (ontologia), "cosa possiamo conoscere" (epistemologia), o "cosa dobbiamo fare" (etica), non possono più essere affrontate seriamente senza tener conto dei risultati dell'indagine scientifica.*" [20].

Dopo Kant, il problema gnoseologico è stato riproposto non tanto "a tavolino" dai filosofi, quanto dalle riflessioni imposte dalle scoperte della scienza del XIX e del XX secolo: dalla scoperta di Riemann delle geometrie non euclidee al secondo principio della termodinamica, dalla teoria della relatività di Einstein al teorema di Gödel.

La critica all'induttivismo, riproposta da Karl Popper, viene qui utilizzata come emblema della filosofia della scienza del XX secolo [21]. Esiste una storiella relativa alla critica all'induzione, la quale, pur originariamente pensata da Bertrand Russell (la storiella è citata nel *The Problems of Philosophy*), viene comunemente attribuita, nella sua forma più elaborata, a Karl Popper: è la storiella del "tacchino induttivista" (*inductivist turkey*). Popper descrive questa situazione: esisteva un tacchino in un allevamento al quale veniva portato il cibo sempre alle 9 di mattina. Il tacchino osservava dunque che qualsiasi giorno della settimana, che vi fosse stato il sole o il cattivo tempo, il cibo gli veniva portato sempre alla stessa ora. Da queste osservazioni ripetute e identiche in qualsiasi condizione meteorologica e che erano comuni per tutti i giorni della settimana, il tacchino applicò il metodo induttivo quando formulò la teoria seguente: "mi danno il cibo sempre alle 9 di mattina". Tuttavia, alla vigilia di Natale, il tacchino constatò a sue spese il venire meno di questa regola: il tacchino venne ucciso e servito a tavola. Questa storiella dimostra quindi il meccanismo erroneo che porta l'osservazione ripetuta di un fenomeno a fondare una legge universale: ancora una volta è chiaro che nessun numero di osservazioni identiche può affermare nulla circa l'universalità della legge che sembra esprimere, mentre anche una sola osservazione contraria la contraddice.

Vi è quindi un'asimmetria tra verificazione e falsificazione di un evento: se per verificare l'universalità di una legge non basta far riferimento alla frequenza degli eventi che costituisce il fondamento di questa legge, dall'altro lato basta solo un evento contrario che interrompa tale frequenza a falsificare la legge universale che vuole considerare tale frequenza come eterna. Tale constatazione è nota come "principio di falsificazione". "*Una proposizione universale può essere falsificata da un solo caso contrario, mentre nessun numero di casi non contrari, per quanto elevato, può verificarla.*" Da ciò deriva l'evidente corollario per cui, secondo Popper, perché una teoria sia veramente scientifica, occorre che possa essere falsificata. E in generale questo accade quando una teoria fa delle previsioni verificabili con l'esperimento.

Alla concezione di induzione come passaggio dal particolare all'universale si è quindi progressivamente sostituita una concezione diversa, che definisce l'induzione come "inferenza ampliativa ma solo probabile", laddove la deduzione è definita come una "inferenza non ampliativa ma necessaria". Sono state le riflessione di David Hume e poi di John Stuart Mill, via via fino a Karl Popper, a mettere a punto progressivamente questa concezione. La forma canonica dell'induzione diventa la seguente:

[48] "*…la ragione vede solo ciò che lei stessa produce secondo il proprio disegno, e che, con principi de' suoi giudizi secondo leggi immutabili, deve essa entrare innanzi e costringere la natura a rispondere alle sue stesse domande; e non lasciarsi guidare da lei, per dir così, colle redini; perché altrimenti le nostre osservazioni, fatte a caso e senza disegno prestabilito, non metterebbero capo a una legge necessaria, che pure la ragione cerca e di cui ha bisogno. E' necessario dunque che la ragione si presenti alla natura avendo in mano i principi, secondo i quali soltanto è possibile che fenomeni concordanti abbian valore di legge, e nell'altra l'esperimento, che essa ha immaginato secondo questi principi: per venire, bensì, istruita da lei, ma non in qualità di scolaro che stia a sentire tutto ciò che piaccia al maestro, sibbene di giudice, che costringa i testimoni a rispondere alle domande che egli loro rivolge*".

1. Ho visto un corvo ed era nero; 2. Ho visto un secondo corvo ed era nero; 3. Ho visto un terzo corvo ed era nero; …...

Conclusione 1. *Il prossimo corvo che vedrò sarà probabilmente nero.*

oppure, in forma generalizzante,

Conclusione 2. *Tutti i corvi sono probabilmente neri.*

Se oltre alla probabilità si tiene conto dell'ampliatività del contenuto della conclusione rispetto a quanto è contenuto nelle premesse, possiamo dire che nell'induzione, diversamente dalla deduzione, il contenuto informativo della conclusione non è interamente incluso nelle premesse. Ad esempio se le premesse affermano che in n casi gli x osservati hanno mostrato di possedere la proprietà A, allora si inferisce che il prossimo x che verrà osservato nel caso $n+1$ probabilmente avrà la proprietà A, oppure che tutti gli x che verranno osservati mostreranno probabilmente la proprietà A. L'avverbio 'probabilmente' è cruciale: mentre è impossibile che la conclusione di un ragionamento deduttivo sia falsa se le sue premesse sono vere (e se il ragionamento è effettuato in modo corretto), in un argomento induttivo questa certezza si riduce a un grado di probabilità maggiore di 0 e inferiore a 1. Si potrebbe dire che questo è il prezzo che si deve pagare per il vantaggio che gli argomenti induttivi offrono rispetto a quelli deduttivi, cioè la possibilità di scoprire e prevedere fatti nuovi in base a quelli vecchi.

Come afferma Dario Antiseri [22] "*I fatti, cioè le asserzioni che, per quel che ci è possibile saperne, descrivono fatti - le basi empiriche della scienza, insomma, sono artefatti che vengono continuamente rifatti attraverso costruzioni e demolizioni teoriche. Essi non sono dati immutabili, ma "costrutti" che hanno una storia: una genesi, uno sviluppo, mutazioni, e talvolta anche una morte. Ciò che oggi chiamiamo un fatto, ieri era una teoria. Ed è gran parte dell'epistemologia del nostro [XX] secolo da Henri Poincaré per giungere alle proposte di P.K. Feyerabend, H.R. Hanson e N. Goodman che ha frantumato il mito della sacralità dei fatti. Certo, le teorie scientifiche poggiano sui fatti, ma questi non sono una roccia indistruttibile. In altri termini, la scienza ha sì una base, ma questa base non è un fondamento certo. Per dirla con Popper: "[…] la base empirica delle scienze oggettive non ha in sé nulla di "assoluto". La scienza non poggia su un solido strato di roccia. L'ardita struttura delle sue teorie si eleva, per così dire, sopra una palude. E' come un edificio costruito su palafitte. Le palafitte vengono conficcate dall'alto, giù nella palude: ma non in una base naturale "data"; e il fatto che desistiamo dai nostri tentativi di conficcare più a fondo le palafitte non significa che abbiamo trovato un terreno solido. Semplicemente, ci fermiamo quando siamo soddisfatti e riteniamo che almeno per il momento i sostegni siano abbastanza stabili da sorreggere la struttura". In breve: "la nostra conoscenza ha fonti d'ogni genere, ma nessuna ha autorità".*

L'epitaffio al problema sembra essere stato scritto dal matematico e logico tedesco Kurt Gödel, che negli anni '30 ha dimostrato che qualsiasi sistema assiomatico, che parta cioè da affermazioni "certe", ma che siano anche moderatamente complicate, arriva prima o poi a delle proposizioni "indecidibili"[23]. In altre parole, anche se le assunzioni di base sono "certe", alla fine si arriva a dei passaggi per i quali la certezza non vale più, e quindi la "verità" o la "falsità" della proposizione non può più essere dimostrata. L'unica alternativa pare essere quella di accettare che la proposizione sia "vera" o "falsa" con un certo grado di probabilità.

Come si vedrà il teorema di Bayes svolge un ruolo centrale nell'induzione, consentendo di aggiornare (ovviamente in termini probabilistici) l'informazione sulla base dell'esperienza, aumentando l'informazione disponibile, e superando l'ostacolo rappresentato da un falsificazionismo ingenuo: perché anche se all'aumentare delle osservazioni a suo favore una teoria non diventa sempre più vera (vedansi i problemi incontrati dal povero tacchino induttivista), diventa peraltro più plausibile, cioè più probabile (o almeno questa è l'interpretazione "naturale" che anche alcuni fisici forniscono).

Obiettivo delle due parti che seguono è di analizzare i due fattori chiave della conoscenza:
- la correttezza dell'informazione (ci devono essere/devono essere giuste e devono essere utilizzate correttamente le informazioni);
- la correttezza delle regole (si devono sapere/devono esser giuste e devono essere applicate correttamente le regole).

Bibliografia e riferimenti

[1] http://it.wikipedia.org/wiki/Euclide
[2] http://it.wikipedia.org/wiki/Pitagora
[3] Nicola U. Atlante illustrato di filosofia. Demetra, 1999.
[4] Guillen M. Le cinque equazioni che hanno cambiato il mondo. Longanesi & C, 1999.
[5] Fraser JC. Il ramo d'oro. Boringhieri, 1973.
[6] http://it.wikipedia.org/wiki/Presocratici
[7] http://it.wikipedia.org/wiki/Socrate
[8] http://it.wikipedia.org/wiki/Platone
[9] http://it.wikipedia.org/wiki/Poliedro
[10] http://it.wikipedia.org/wiki/Solido_platonico
[11] http://it.wikipedia.org/wiki/Aristotele
[12] http://it.wikipedia.org/wiki/Galileo_Galilei
[13] http://lgxserve.ciseca.uniba.it/lei/biblioteca/cxc/public/b/boniolo2.pdf
[14] http://it.wikipedia.org/wiki/Cartesio
[15] http://it.wikipedia.org/wiki/Thomas_Hobbes
[16] http://it.wikipedia.org/wiki/John_Locke
[17] http://it.wikipedia.org/wiki/David_Hume
[18] http://it.wikipedia.org/wiki/John_Stuart_Mill
[19] http://it.wikipedia.org/wiki/Immanuel_Kant
[20] http://lgxserve.ciseca.uniba.it/lei/biblioteca/lr/public/dorato-1.0.pdf
[21] Gillies D, Giorello G. La filosofia della scienza nel XX secolo. Laterza, 1995.
[22] Antiseri D. Le "evidenze" della EBM: "fatti" o artefatti? Keiron, 2002 (http://www.farmindustria.it/pubblico/004/etica.pdf).
[23] Barrow JD. La luna nel pozzo cosmico. Milano, Adelphi, 1994.

4. Dati, informazione e conoscenza

“Dove è la sapienza
che abbiamo smarrita nella conoscenza
dove è la conoscenza
che abbiamo smarrita nell’informazione
dove è l’informazione
che abbiamo smarrita nei dati.”
(Mark Porat)

Come abbiamo detto chiudendo la breve trattazione del problema gnoseologico, uno dei due fattori chiave della conoscenza è rappresentato della correttezza dell’informazione: ci devono essere/devono essere giuste e devono essere utilizzate correttamente le informazioni. I problemi sorgono quando mancano/sono sbagliate o non sono utilizzate correttamente le informazioni.

La frase di Mark Porat, incisa nell’ala delle Comunicazioni del Museo Scientifico dei Bambini, a Washington, ha un significato profondo. Letta dall’alto verso il basso può essere intesa come la metafora della società moderna. I ritmi di vita sempre più serrati, hanno prima rotto (per sempre?) quell’armonico senso di unità tra corpo, spirito e natura che gli antichi definivano *sapienza*. Quindi hanno imposto (stanno sempre più imponendo) la necessità di suddividere l’enorme quantità di *conoscenza* globalmente disponibile, in sottoinsiemi (ecco che nasce la figura dello specialista) di dimensioni compatibili con la limitatezza della capacità di gestione della nostra mente (la corteccia cerebrale umana, ancorché sia una delle macchine più incredibilmente complesse e sofisticate presenti in natura, possiede un limite fisico finito, invalicabile). Ancora, i ritmi di vita sempre più serrati hanno imposto (stanno sempre più imponendo) la necessità di comprimere i processi di comunicazione. Il paradigma è rappresentato dalla comunicazione mediante immagini. Il processo di comunicazione aumenta in efficienza[49], e aumenta in efficacia[50]. Ma la comunicazione mediante immagini, a fronte di un aumento dell’efficienza e dell’efficacia del processo di comunicazione, comporta, ahimè, una perdita del dettaglio[51]. Inoltre il messaggio “visivo” ricevuto viene tendenzialmente percepito come compiuto: la società dell’immagine spinge quindi anche a perdere il valore aggiunto (ricchezza inestimabile) determinato dalla riflessione/analisi critica dell’individuo sul suo contenuto. La conoscenza ha ormai lasciato il posto all’*informazione*. L’informazione fine a sé stessa, l’informazione-business, è il grande fratello orwelliano già presente tra noi (ma quanti se ne sono accorti?). Ma c’è di peggio nell’aria. Il flusso informativo è talmente imponente, rispetto al tempo dedicato/dedicabile alla riflessione critica su di esso da parte del singolo individuo, da rischiare di diventare semplicemente un flusso di *dati*.

Letta dal basso verso l’alto la frase di Porat può essere intesa come la metafora della crescita e dello sviluppo dell’individuo. Già nell’utero materno il feto riceve in continuazione *dati* provenienti dalla madre. Fino a un certo punto solo di dati si tratta. Il battito del cuore materno che batte con regolarità, il rumore dell’aria che entra ed esce dagli alveoli polmonari materni, la voce della madre che risuona attraverso i tessuti e i liquidi che lo circondano. Appena nato però il bambino impara subito ad integrare (operazione matematica) i dati provenienti dal mondo esterno. Lo fa perché solo

[49] Per trasmettere il messaggio è necessario meno tempo, e meno tempo è necessario per memorizzarlo. In altre parole è necessario meno lavoro per inviare e per ricevere il messaggio.
[50] La comunicazione mediante immagini aumenta anche i risultati ottenuti (l’outcome, efficacia del messaggio). Basta pensare all’efficacia di uno spot, in grado, se ben concepito, di indirizzare milioni di consumatori verso un prodotto, spesso solo voluttuario.
[51] Come conseguenza della compressione dell’informazione, si ha la perdita di parte del contenuto informativo del messaggio.

deve sopravvivere. Deve imparare immediatamente ad integrare i dati, ancora per lui confusi, provenienti dal mondo esterno (per ora solo la madre) al fine di avere una prima *informazione* essenziale: dove trovare il cibo. E impara subito a cercare il latte materno e a succhiare. Altre informazioni egli dovrà presto imparare a generare, integrandoli, dai dati che gli provengono dal mondo in cui vive. La scuola lo aiuterà poi ad integrare le informazioni al fine di generare la *conoscenza* che gli servirà per inserirsi nella vita adulta e nella attività lavorativa. Un percorso non facile. Pochi privilegiati (forse) riescono ad effettuare l'integrazione/salto logico che separa il livello della conoscenza da quello della *sapienza*.

In senso generale la frase di Porat è la descrizione degli elementi che entrano nella generazione della conoscenza e della loro posizione reciproca: e ci consente di riflettere sulle implicazioni dei processi che consentono di effettuare il passaggio

dati ⇨ *informazione* ⇨ *conoscenza*

trasformando prima i *dati* in *informazione*, e successivamente questa in *conoscenza*. L'obiettivo, pur con tutte le difficoltà che vedremo, è di indurre a utilizzare un approccio scientifico, non arbitrario, e quantitativo, nella gestione dell'informazione. Che ovviamente vede al centro il teorema di Bayes.

Prima di continuare vale la pena di ricordare le definizioni di *dato*, *informazione* e *conoscenza* fornite da un noto vocabolario [1].

Dato ⇨ *elemento o serie di elementi accertati e verificati che possono formare oggetto di indagini, ricerche, elaborazioni o che comunque consentono di giungere a determinate conclusioni*.

Informazione ⇨ *atto, effetto dell'informare o dell'informarsi* [essendo a sua volta] ***informare*** ⇨ *modellare secondo una forma*.

Conoscenza ⇨ *facoltà, atto, modo, effetto del conoscere* [essendo a sua volta] ⇨ ***conoscere*** *prendere possesso intellettualmente o psicologicamente, specialmente con un'attività sistematica, di qualunque aspetto di quella che è considerata realtà*.

Si sottolinea la rinuncia deliberata, in quanto assolutamente impervia, all'analisi dell'elemento al vertice della frase di Porat, la ***sapienza*** [definita come] ⇨ *il più alto grado di conoscenza delle cose* [ovvero definita come] ⇨ *sapere vasto e profondo uniti a doti morali e spirituali*.

È necessario infine, come ultimo passo, dare una definizione del termine ***intelligenza*** ⇨ *capacità generale che consente di adattarsi all'ambiente e che nell'essere umano si manifesta nei comportamenti e nel grado di elaborazione dei processi mentali* [1]. Ricordando che il temine intelligenza deriva dal latino *intelligere*, per il quale si rimanda alla voce ***intelletto*** ⇨ *facoltà di intuire le idee, le rappresentazioni e i loro rapporti* [1].

La corrispondenza tra gli stati attraversati dalle rappresentazioni mentali sulla via dell'*intelligere* (*dati* ⇨ *informazione* ⇨ *conoscenza*) e le azioni intraprese dall'individuo sono le seguenti:

Rappresentazione mentale	***Deriva dall'azione di…***
Dato	*…ricerca [astrarre]*
Informazione	*…modellare secondo una forma [modellizzare]*
Conoscenza	*…prendere possesso intellettualmente … della realtà [connettere, rappresentare]*

Iniziamo dal dato, che è la prima rappresentazione mentale, quella più elementare. Ma è già di per sé una grande conquista intellettuale. Il dato è il risultato di un processo di astrazione. Fornire una tabella che riporta l'altezza in centimetri di 1000 soggetti appartenenti alla popolazione italiana, corrisponde a fornire un elenco di dati. Dove "l'altezza in centimetri" è il risultato di un procedimento di misura. Per definizione "*la misura (dell'entità) di una grandezza fisica*" consiste nell'esprimere la grandezza in modo quantitativo, dando ad essa un "*valore numerico*" che è un numero puro, ottenuto per confronto della *"(entità della) grandezza in esame*" con la *"(entità di una) grandezza di riferimento ad essa omogenea, definita unità di misura*" [2]. Un concetto astratto, che consente peraltro di dare ai dati una connotazione oggettiva[52] (data la sua importanza, una breve introduzione all'argomento metrologico è riportata nell'appendice A).

Di esso interessa ora sottolineare come dalla definizione stessa di misura

$$\frac{\textit{(entità della) grandezza}}{\textit{unità (di misura)}} = \textit{valore numerico}$$

si ottiene

$$\textit{(entità della) grandezza} = \textit{valore numerico} \cdot \textit{unità (di misura)}$$

il che indica che un dato deve per definizione essere espresso con un valore numerico e con le unità di misura. Abbiamo ottenuto la prima conseguenza razionale di una impostazione rigorosa delle definizioni: *ci stiamo preoccupando del fatto che i dati, come prerequisito dell'informazione, poggino su solide fondamenta*. E per allargare queste fondamenta dobbiamo inevitabilmente occuparci dei numeri, i personaggi che le popolano.

Il primo aspetto che consideriamo riguarda l'acquisizione dei dati, e quindi è connesso con il procedimento di misura. Ci ricorda l'importanza straordinaria della misura Robert Laughlin, premio Nobel per la fisica nel 1998 con Daniel C. Tsui e Horst Störmer per la scoperta dell'effetto Hall quantistico [4]. Dice Laughlin [3]: "…*misurando con una precisione di una parte su centomila, scopriamo che la lunghezza di un mattone varia da un giorno all'altro. Analizzando i fattori ambientali, scopriamo che ciò è dovuto alle variazioni di temperatura, che determinano infinitesimali contrazioni ed espansioni del mattone. Il mattone si comporta quindi come un termometro. Non si tratta affatto di un'osservazione sciocca, dal momento che la dilatazione termica è il principio alla base di qualsiasi comune termometro. Misurando il peso con analoga precisione, non si registra alcuna variazione, ed è proprio questa una delle osservazioni che portano al principio di conservazione della massa. Se peraltro lavoriamo con una precisione di una parte su cento milioni, il peso del mattone potrà variare leggermente da un laboratorio all'altro. A questo punto, il nostro mattone è diventato un misuratore di gravità, trattandosi dell'effetto di minuscole variazioni della froza di gravità, dovute alla diversa densità delle rocce che giacciono immediatamente sotto la superficie terrestre…*"

L'esperienza ha mostrato che nessuna misura, per quanto effettuata con cura, può essere completamente libera da errori [5]. E' a Karl Friedrich Gauss, "il principe dei matematici" [6] che dobbiamo la formalizzazione matematica dell'errore di misura, un prerequisito essenziale per calcolare l'incertezza globale delle nostre conclusioni al termine del procedimento di misura (per i concetti minimi sull'errore di misura e sulla sua espressione si rimanda all'appendice B). Quello

[52] Non importa che la misura sia basata su una unità di misura "arbitraria" (come ad esempio il metro e il kilogrammo). Un dato misurabile può finalmente essere confrontato con un altro dato, essendo a questo punto il confronto relativo di due dati reso "oggettivo" dal concetto soggiacente di misurabilità, che consente di eliminare l'arbitrarietà del giudizio.

che qui preme sottolineare è che non solo le misure iniziali sono affette da un errore, ma che l'errore tende poi a propagarsi (mai ad elidersi!)[53]. Possiamo dire che globalmente è la legge del GIGO che domina la fase di acquisizione di dati. GIGO come acronimo di Garbage In-Garbage Out: se ci metti dentro spazzatura ne tiri fuori solo spazzatura. Se il dato (il segnale fisico, ovvero il numero ricavato da una misura) contiene molto rumore, l'informazione ne risulterà proporzionalmente deteriorata. Se, diciamo, l'altezza di un individuo viene misurata con una approssimazione (errore di misura, ovvero incertezza) di 5 centimetri, che senso ha dire che il tal dei tali, alto 172 centimetri, è più basso del tal altro, che è alto 174 centimetri? Nessuna, ovviamente. Mentre se l'altezza viene misurata con un'approssimazione (errore di misura, ovvero incertezza) di 1 centimetro, l'affermazione ha (intuitivamente) più senso[54].

Camminando per strada a Milano, una domenica mattina di blocco totale del traffico a causa dell'inquinamento, per la prima volta sono riuscito a riconoscere, provenienti dal marciapiedi opposto al mio, in una della vie più congestionate della città, le voci alcune persone che chiacchieravano davanti ad un bar, e la musica che proveniva da una finestra aperta (erano "Le quattro stagioni" di Vivaldi"). In condizioni "normali" (notare il concetto di abnormità che può essere insito nell'espressione "normale"[55]) il rumore del traffico mi aveva fino ad allora impedito di sentirle. Da un punto di vista fisico, in generale, qualsiasi segnale può essere riconosciuto solamente se riesce a sopravanzare il "rumore di fondo" dell'ambiente[56]. Se si ascoltano "Le quattro stagioni", dal "Cimento dell'armonia e dell'invenzione" di Antonio Vivaldi, in una incisione su CD, si possono distinguere agevolmente i suoni dei violini e dei violoncelli. Se si ascolta la stessa incisione su una musicassetta, si può riconoscere l'insieme degli archi, ma i suoni dei violini e quelli dei violoncelli non possono più essere distinti. La causa è insita nel fatto ben noto che il CD ha un rapporto segnale/rumore migliore della musicassetta. E questo consente al CD di avere un potere di risoluzione maggiore tra due suoni "vicini" quali quelli dei violini e dei violoncelli. Il concetto di potere di risoluzione si applica al microscopio, così come all'ibridazione genomica con segmenti di 44 kilobasi piuttosto che di 244 kilobasi, così come alla capacità di immagazzinazione dei dati su un hard.-disk, così come alla mappa della radiazione di fondo a microonde rilevata da COBE.

Il secondo aspetto riguarda la pre-elaborazione dei dati. Si rammenta innanzitutto che per riportare i risultati con un adeguato <u>numero di cifre significative</u> gli zeri dopo i numeri diversi da zero fanno parte delle cifre significative stesse mentre, nel caso di valori numerici minori di uno, gli zeri prima dei numeri diversi da zero non fanno parte delle cifre significative. Così per esempio 0,25 e 0,00025 hanno entrambi due sole cifre significative. Mentre scrivendo 0,250 si intende significare che le cifre significative sono tre [7].

---

[53] La maggior parte delle grandezze fiche non possono essere di solito misurate direttamente. Di solito, infatti, in primo luogo si misurano direttamente una o più grandezze fisiche, dalle quali la grandezza che ci interessa può essere calcolata. In secondo luogo, utilizzando tale o tali grandezze fisiche, si calcola la grandezza in questione. Per le regole di propagazione dell'errore di misura vedere [5].

[54] Non ha nessun senso ricorrere a tecniche statistiche sofisticate se i dati di partenza sono viziati da un rumore di fondo (errore) elevato, che nessuna tecnica statistica è in grado di eliminare. Meglio quindi un disegno sperimentale e procedimenti di misura rigorosi, che consentano di ottenere dati con il minor rumore di fondo (errore) possibile.

[55] A causa di queste e di altre ambiguità e assurdità derivanti dall'uso del termine "normale", nella medicina di laboratorio, per esempio, si è sostituita l'espressione "valori normali" e "intervalli di normalità", con riferimento ai valori "più probabili" in un soggetto sano, con l'espressione "valori di riferimento" e "intervalli di riferimento". Nel caso della concentrazione di un farmaco nel siero si parla di "intervallo terapeutico". Nel caso delle varie classi di lipidi (colesterolo, trigliceridi, colesterolo HDL, colesterolo LDL, trigliceridi) si parla di "valore desiderabile". Nel caso di sostanze tossiche (metalli pesanti, metaboliti dello xilolo e del toluolo, eccetera) si parla di "intervalli di riferimento in soggetti esposti" e di "intervalli di riferimento in soggetti non esposti" al tossico.

[56] Questo sta alla base delle curve ROC, che come vedremo si possono derivare da una analisi bayesiana dei dati.

Al fine di non introdurre distorsioni nei dati/risultati, esistono delle regole di arrotondamento: un numero deve essere arrotondato per difetto se termina per 1, 2, 3 o 4, e deve essere arrotondato per eccesso se termina per 6, 7, 8, o 9. Se termina per 5 deve essere arrotondato per eccesso se la cifra che precede il 5 è dispari, ovvero per difetto se la cifra che lo precede è pari[57].

Come regola pratica, nel caso di un numero a precisione elevata, il numero delle cifre significative dovrebbe essere tale che l'ordine di grandezza del campo di variazione (differenza fra il valore massimo e quello minimo) calcolato ignorando l'eventuale presenza della virgola sia compreso tra 30 e 300. Si supponga che la misura della concentrazione dello ione calcio (Ca^{++}) nel sangue in un campione di 50 soggetti abbia fornito risultati compresi tra 2,06442 (valore minimo) e 3,05620 (valore massimo) mmol/L. La differenza tra il valore massimo e quello minimo (senza tenere conto della virgola) è pari a 99178 (305620 - 206422) mmol/L e quindi il numero delle cifre utilizzate per rappresentare i risultati è eccessivo. Arrotondando i risultati a due cifre decimali il valore minimo diventa 2,06 mmol/L, e il valore massimo diventa 3,06 mmol/L. La differenza (senza tenere conto della virgola) è pari a 100 mmol/L. Pertanto tre cifre significative sono sufficienti per rappresentare i risultati [7].

Come regola teorica, *la precisione finale con cui registrare i risultati non può essere maggiore di quanto permesso dall'incertezza (imprecisione)* dalla quale sono affetti i risultati. Il numero di cifre significative si deduce allora dalla deviazione standard della singola misura. Ed è un errore esprimere il valore numerico della misura con un numero di cifre superiore o inferiore a quello delle cifre significative soprattutto se non viene indicata anche la suddetta deviazione standard. Si supponga che la misura del calcio nel siero abbia fornito (media ± deviazione standard) un valore pari a 9,376 ± 0,10835 mg/dL. La deviazione standard indica 2 cifre significative, in quanto l'incertezza compare nel risultato già sulla seconda cifra da sinistra (la prima cifra decimale). Il modo corretto di scrivere il risultato è 9,4 ± 0,1 mg/dL [7].

Discorso analogo deve essere fatto per il numero di cifre significative da usare nel caso di indici statistici (statistiche) come la media e la deviazione standard. Sinteticamente si può dire che la precisione con cui si deve riportare qualsiasi statistica non può essere maggiore dell'errore che si commette nel processo di stima. Quindi, nel caso più semplice, quello della media, non può essere maggiore del suo errore standard. Si supponga che, sempre nel caso precedente della determinazione dello calcio nel siero, alla media di 9,376 e alla deviazione standard di 0,10835 corrispondesse un errore standard della media pari 0,0320489 (l'errore standard è uguale alla deviazione standard divisa per la radice quadrata del numero delle osservazioni che compongono il campione). L'errore standard conferma che già la seconda cifra decimale è affetta da errore. Non avrebbe quindi senso riportare la media e la deviazione standard con più di due cifre decimali. La media deve essere scritta come uguale a 9,38 e la deviazione standard deve essere scritta come uguale a 0,11 [7].

Ci siamo preoccupati finora del fatto che i dati, come prerequisito dell'informazione, poggino su solide fondamenta. E abbiamo visto come sia nella fase di acquisizione dei dati sia nella fase di pre-elaborazione dei dati gli strumenti che abbiamo a disposizione sono di tipo quantitativo. Il che, dopo avere dimostrato che i nostri dati sono inevitabilmente incerti, ci consente almeno di quantificare in modo oggettivo tale incertezza, di misurarla.

---

[57] Esempio: 8,25 e 8,35 (hanno come media 8,30) sono arrotondati secondo la regola indicata rispettivamente a 8,2 e a 8,4 (che conservano, ed è questo il fatto fondamentale, la stessa media).

Il problema diventa assai più complesso quando passiamo dai dati all'informazione: perché informare, come abbiamo detto, significa dare forma, e questo aspetto mal si presta ad una analisi quantitativa. Quindi già partiamo male: e i guai che possono derivare dal "dare forma" ai dati sono sotto gli occhi di tutti, basta esaminare la grande metafora del quotidiano[58]. Dove troneggia sempre il più classico degli aspetti, cioè il dare forma differente agli stessi dati del "visto da destra" e del "visto da sinistra". Ma dove anche l'informazione ridondante e urlata, a fronte di un minimo senso critico da parte del destinatario dell'informazione stessa, fornisce da sola tutto quanto necessario per dimostrare la sua mancanza di neutralità rispetto ai dati, come dai due emblematici sberleffi di Roberto Fedi de "La morsa del caldo"[59] e de "La morsa del gelo"[60] che appare doveroso riportare

---

[58] Non inteso come giornale, ma inteso come "della giornata". Ma il doppio senso vale comunque.

[59] "*La morsa del caldo - di Roberto Fedi – data di pubblicazione sul web 27/06/2005*
Siamo pronti a scommettere che non c'è nessuno di voi che non conosca l'aurea massima del giornalismo: che la notizia è l'uomo che morde il cane, e non viceversa. Allo stesso modo, secondo noi la notizia vera, quella da strombazzare su tutti i tiggì, sarebbe che d'estate fa freddo, o che d'inverno fa caldo. E non che alla fine di giugno c'è il sole che picchia, e che si suda, e che chi suda ha sete; e che, in questi casi, fa più piacere trincarsi una bella bibita fresca piuttosto che una tazza di latte bollente. O no? Perché, invece, i tiggì di casa nostra (tutti) sembrano pensarla diversamente. Vabbè che d'estate le notizie scarseggiano (o almeno sembra: in realtà mica tanto, anzi qualche volta abbondano – in questi giorni, ad esempio: sciopero dei magistrati, gravissimo; scioperi dei mezzi pubblici; omicidi vari; kamikaze in Iraq; Fiat a picco; problemi economici come d'inverno se non peggio), ma insomma non è proprio il massimo che la prima, lunghissima notizia di ogni telegiornale sia questa: la morsa del caldo. La frase, inevitabile (che fa il paio con quella invernale: La morsa del gelo), è stata anche detta sul serio, ad esempio nel Tg delle 12.30 de La7 di domenica 26 giugno (a Studio aperto, Italia uno, dello stesso giorno hanno voluto distinguersi: "la morsa del grande caldo"): e ci vuole anche un bel coraggio, diciamo la verità, a pronunciare senza vergognarsi questo vertice del luogo comune. Ma il bello viene con quello che segue, e cioè con i consigli degli esperti. È una gara a chi dice più scempiaggini. Per esempio: sapete come si fa a star meglio la notte? Dormite al fresco, in una stanza con un condizionatore. Ma davvero? E chi l'avrebbe detto? È invece un prezioso consiglio nel Tg de La7 di cui sopra. Oppure: voi poveri scemi pensavate che per sentire meno caldo fosse bene stare al sole nel primo pomeriggio? Che sbadati. Dovete stare in casa, a persiane chiuse (se avete le persiane): ecco perché finora vi siete bolliti come la pastasciutta. O anche: siamo sicuri che col solleone voi, poveri trogloditi carnivori, vi pappate tutti i giorni bistecche di maiale. Ahi ahi ahi… Ecco perché vi fa male il pancino. Dovete invece magna' verdura e frutta fresca. E perché vi scolate quei mezzi litri di cioccolata calda quando fuori sono 35 gradi all'ombra? Bevete acqua del frigo, oppure succhi di frutta (freddi). E scommettiamo che il pomeriggio accendete il termosifone. Ma siete proprio incorreggibili! Accendete l'aria condizionata, se ce l'avete; altrimenti un bel ventilatore, che diamine. E via, levatevi quei cappottoni color antracite con cui uscite tutti i santi giorni, e ricordatevi, scemi siberiani che non siete altro, di indossare vestiti leggeri. Ma se non ci fosse la televisione che vi insegna anche a mettervi le mani in tasca, ma voi poveracci che fareste? Questo, accaldati amici, il tono e i preziosi consigli dei tiggì di fine giugno. Emilio Fede, addirittura, ha ormai trasformato il suo in un lungo spettacolino di previsioni del tempo e di servizi sul caldo (ne siamo contenti, francamente). Ma anche tutti gli altri sono di questo tono. E poi dicono che la televisione non è un pubblico servizio. PS. A chi obiettasse che un articolo più o meno come questo l'avevamo già scritto un paio d'anni fa, di questi tempi (Afa), risponderemmo che non è colpa nostra se l'estate fa caldo, e se la Tv che abbiamo è questa. E che, comunque, fate bene ad accendere il condizionatore, se ce l'avete."

[60] "*La morsa del gelo - di Roberto Fedi – data di pubblicazione sul web 26/01/2004*
Ce lo aspettavamo. Eravamo sicuri che non c'era scampo, e che prima o poi doveva succedere, anche nelle migliori famiglie. E infatti è successo: e lo registriamo quasi con dolore. Un paio di giorni fa, al Tg5 in prima serata, Lamberto Sposini l'ha detto. Mentre stavano per scorrere le solite immagini del freddo (neve che cade, macchine coperte di bianco, strade con gli spazzaneve, gente imbacuccata, bambini che fanno a pallate), con aria seria s'è lasciato un po' andare: "L'Italia è nella morsa del gelo". Forse gli è scappato, o forse è stato un momento di disattenzione. Ma l'ha detto: siamo pronti a giurarlo. Naturalmente ci sarebbero tanti modi per dire che l'Italia è investita da un'ondata di freddo. Per esempio questo: "l'Italia è investita da un'ondata di freddo". Oppure: in molte regioni d'Italia sta nevicando. O ancora: accidenti che freddo (questo per chi preferisce un tono amichevole e disinvolto). Ma volete mettere la metafora della "morsa"? Sa di Siberia, di ghiacci eterni, di Michele Strogoff. Ti fa sentire non infreddolito e basta, ma quasi il protagonista di un film sull'Antartide. Con gli orsi polari e qualche pinguino in trasferta. La "morsa" fa il paio con le metafore del caldo. Vi ricorderete che la scorsa estate, con quell'afa micidiale, i luoghi comuni si sprecavano - quella volta, naturalmente, all'opposto (ne parlammo a suo tempo: Afa). Per esempio verso la fine dell'estate, quando la televisione aveva dato fondo a tutte le metafore possibili, tutti cominciarono a parlare della "bolla di caldo". Era un caldo africano (altro luogo comune), ma il fatto di essere dentro una "bolla" ci faceva sentire, se non altro, terminologicamente più autorizzati al sudore. La faccenda è meno superficiale di quello che sembrerebbe.

per intero come sommo richiamo alla serietà dell'informare (giornalisticamente parlando). Mentre il richiamo di Armando Massarenti ai discorsi che si svolgevano nel 1683 in tema di cambiamento climatico, appare come un richiamo alla serietà dell'informare scientificamente parlando[61].

La causa del disastro sembra risiedere come dice Ugo Volli [8] in "*una inflazione comunicativa della semiosfera*"[62] così definita: "*Sembra difficile negare che nella semiosfera sia in atto una crisi strutturale: in particolare una crisi del modello secondo cui la nostra semiosfera è stata sviluppata negli ultimi decenni. È ragionevole pensare che questa crisi abbia un carattere inflazionario. In termini molto banali, l'inflazione economica si può descrivere come una condizione di instabilità che consiste nella circolazione di una quantità di moneta sproporzionata alla quantità di beni che*

Perché rientra nel linguaggio dei Tg, che sono o dovrebbero essere strumenti primari di informazione. E quindi anche di diffusione linguistica, sperabilmente di un genere non proprio banale. E invece questo linguaggio dell'informazione assomiglia sempre di più a quello di plastica - e quindi uguale a se stesso, senza sfumature, impersonale - dei programmi di intrattenimento. Dove si usano per esempio solo aggettivi al grado più alto: fantastico, eccezionale, straordinario, bellissimo, meraviglioso. Niente mezzi toni, niente chiaroscuri. Per fare qualche esempio: è possibile che il freddo sia sempre "polare" o "siberiano"? che il termometro non scenda, ma "precipiti"? Se così fosse, avremmo temperature da meno venti o trenta. E invece, al massimo abbiamo robetta tipo "Firenze, meno due". Capirai. La lingua, e i luoghi comuni, sono spie di una gestione dell'informazione un po' casuale, un po' facile, che preferisce strizzare l'occhio all'ascoltatore piuttosto che dargli notizie meno scontate, e soprattutto nuove. Il quale ascoltatore, se sente il luogo comune, subito per riflesso condizionato ricostruisce l'ipotetica scena standard, quella che ha visto al cinema e che niente ha a che vedere con la realtà; quella falsa e non comunicativa del Grande Fratello, per intendersi. L'esempio della "morsa" è naturalmente ironico, e ben poco colpevole: ma che dire (domenica 25 gennaio, ore 12.40 circa, Tg di Italia 1) della pensionata trovata "in un lago di sangue", "massacrata" (detto ben tre volte nel rapido servizio) con "ferocia" da quello che potrebbe essere un "serial killer"? (e che naturalmente non lo sarà, visto che ha ucciso una volta sola). Fateci caso. E accettate un consiglio: cercate di 'decrittare' la lingua di plastica, traducetela nel linguaggio comune, quello che serve per identificare le cose. Quello in cui, se siamo a meno due, si dice che è freddo, magari aggiungendoci un'esclamazione a scelta, e non che siamo in una "morsa", in Siberia, al Polo o sotto le nevi dell'Alaska. Almeno fino a quando, uscendo di casa, invece che col postino non vi troverete a tu per tu con un orso bianco."

[61] "*Non si sa come vestirsi- di Armando Massarenti - (dall'inserto culturale del Sole 24 Ore - domenica 14 gennaio 2007)*
Il titolo di questo articolo è una piccola scommessa. Un amico abilissimo nell'attaccare bottone con dei perfetti sconosciuti mi ha spiegato che ci sono frasi che funzionano sempre: «Per esempio, se pronunci ad alta voce a una fermata dell'autobus le parole "Non si sa come vestirsi" vedrai che qualcuno certamente ti darà retta». Ed è così che anch'io spero di aver agganciato qualche lettore. Se l'abbordaggio funziona è perché parla del tempo, e in più lo fa in maniera un po' indiretta. Ne scaturisce immediatamente un ragionamento in prima persona, basato sulla propria esperienza, e su cui in genere ci si sente in grado di poter dare dei giudizi ben ponderati. Anche perché questi riguardano precise linee d'azione: vestirsi leggeri, pesanti o a strati? Tuttavia, poiché il dubbio resta, ci interessa molto anche confrontarci con il giudizio degli altri. Finché ci ritroviamo a discutere nientemeno che di scienza: Sta cambiando il clima globale? È colpa del buco nell'ozono? ecc. ecc. Giacomo Leopardi nello Zibaldone ci ricorda quanto scriveva già nel 1683 il presidente dell'Accademia del Cimento, Lorenzo Magalotti: «Egli è pur vero che l'ordine antico delle stagioni par che vada pervertendosi. Qui in Italia è voce e querela comune, che i mezzi tempi non vi son più e in questo smarrimento di confini, non vi è dubbio che il freddo acquista terreno. Io ho udito dire a mio padre, che in sua gioventù, a Roma, la mattina di pasqua di resurrezione, ognuno si rivestiva da state. Adesso chi non ha bisogno d'impegnar la camiciuola, vi so dire che si guarda molto bene di non alleggerirsi della minima cosa di quelle ch'ei portava nel cuor dell'inverno». Leopardi dissente (e noi con lui: il Magalotti poteva andar bene citarlo nel rigido inverno dell'anno scorso, non certo nel mite gennaio 2007): «Quello che tutti sappiamo, e che io mi ricordo bene è che nella mia fanciullezza il mezzogiorno d'Italia non aveva anno senza grosse nevi, e che ora non ha quasi anno con nevi che durino più di poche ore. Così dei ghiacci, e insomma del rigore dell'invernata». È questa sicurezza di giudizio che dovrebbe farci riflettere. È il bello delle chiacchiere tra profani, dove soggettività e convinta oggettività si mescolano in maniera sublime, al punto da diventare indistinguibili, inverando il famoso incipit del Discorso sul metodo di Cartesio: «Il buonsenso è la cosa meglio distribuita tra gli uomini. Nessuno pensa che gliene occorra una quantità maggiore di quella che possiede». Alla fine, insomma, sappiamo tutti come vestirci. Ma per il riscaldamento globale forse è meglio sentire cos'hanno da dirci gli esperti. Dopo il Magalotti, un po' di strada l'avranno pur fatta!"

[62] La semiosfera è definita dallo stesso Volli [8] come "*il testo di tutte le testualità, di tutte le comunicazioni che si depositano e si producono in una società*".

si rendono disponibili. Si può azzardare l'ipotesi che l'inflazione comunicativa della semiosfera consista nel fatto che una società emetta una quantità di messaggi (e di assiologie al loro interno) incompatibile con la quantità di valore semiotico in essa prodotto."
In altre parole la nostra società sta dunque emettendo troppi messaggi rispetto ai valori semiotici che vi circolano, e questo sta producendo una perdita di qualità, una forte svalutazione del senso, dei significati, dei contenuti comunicati.

Eccoci allora usciti dal maluso dell'informazione nel quotidiano, per affrontare quello del "dare forma" ai dati secondo regole scientifiche: ove abbiamo l'esigenza di avere strumenti di modellazione del dato assolutamente neutrali. Abbiamo detto che fornire una tabella che riporta l'altezza in centimetri di 1000 soggetti appartenenti alla popolazione italiana, corrisponde a fornire un elenco di dati. Dire che "l'altezza media, in un campione rappresentativo di 1000 italiani, è di 168,4 centimetri" corrisponde a dare un'informazione. *Ma chi ha dato forma ai dati? E in che modo ha dato forma ai dati?* Ovviamente questo apre una questione assai spinosa.

PROBLEMA 5
Supponiamo di avere determinato la concentrazione nel siero dei trigliceridi (espressi in mg/dL) in 1000 soggetti, scelti a caso tra i pazienti che effettuano ambulatoriamente analisi di laboratorio. I risultati sono riportati nella tabella dell'appendice C. Dopo avere analizzato statisticamente i dati, riassumiamo le conclusioni che abbiamo attenuto a un nostro amico, comunicandogli questa informazione: "abbiamo verificato che in 1000 soggetti, scelti a caso tra i pazienti che effettuano ambulatoriamente analisi di laboratorio, la concentrazione media dei trigliceridi era 136.509 mg/dL, con una deviazione standard (d.s.) di 92.76759 mg/dL".
Quale conclusione il nostro amico può dedurre dai dati che gli abbbiamo fornito?

Il nostro amico, dall'informazione che gli abbiamo fornito, e rammentando che in base alle proprietà della distribuzione gaussiana tra la media - 1,96 d.s. e la media + 1,96 d.s. si deve trova il 95% dei dati campionari, rispettivamente sottrae e aggiunge alla media 181,8144764 mg/dL (cioè 1,96 · 92.76759 mg/dL) e ottiene rispettivamente, dopo arrotondamento alla prima cifra decimale, -45,3 mg/dL e 318,3 mg/dL. Dal che il nostro amico deduce che una certa quota dei pazienti aveva valori negativi della concentrazione nel siero dei trigliceridi! Tuttavia, prima di pubblicare questa rivoluzionaria conclusione, la prudenza impone un riesame dei risultati effettivamente ottenuti (tabella dell'appendice C): e il riesame dimostra che nessuno dei valori era inferiore a zero. Un evidente paradosso biologico (una concentrazione negativa) non porta ad una rivoluzionaria conclusione, ma più semplicemente consente di individuare un grave errore nel dare forma ai dati: calcolando media e deviazione standard è stata (implicitamente) data forma gaussiana a dati che non sono distribuiti in modo gaussiano.

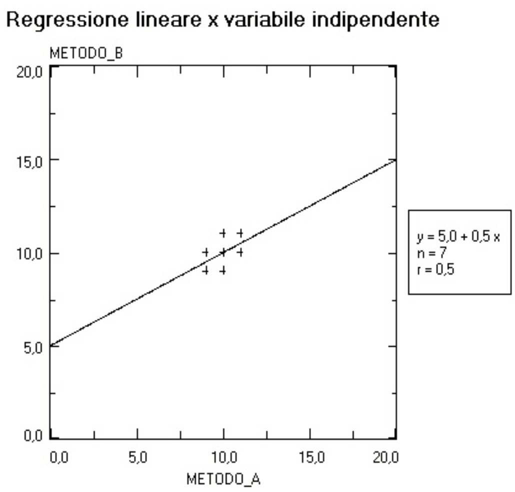
Per illustrare ulteriormente le implicazioni del "dare forma ai dati" vediamo un esempio un po' più tecnico, ma molto istruttivo. Esistono numerose occasioni nelle quali quello che interessa è ricostruire la relazione di funzione che lega due variabili, la variabile y (variabile dipendente, in ordinate) alla variabile indipendente (variabile x, in ascisse); se si ritiene che la relazione esistente fra le due variabili possa essere convenientemente descritta mediante una retta, l'equazione di tale retta può essere calcolata mediante la tecnica statistica nota come regressione lineare. Tale denominazione deriva dagli studi sull'ereditarietà condotti da F. Galton sul finire dell'800. Nel corso di questi studi vennero, fra l'altro, registrate le altezze dei componenti di più di 1000 gruppi familiari. Ponendo su un sistema di assi cartesiani in ascisse le altezze dei padri e in ordinate le altezze dei figli, si notò un fatto: sebbene in genere padri più alti avessero figli più alti (come del resto era atteso), padri che erano di 16 centimetri circa più alti della media dei padri, avevano figli che erano solamente 8 centimetri circa più alti della media dei figli. In altre parole sembrava che vi fosse un "tornare indietro", una regressione delle altezze dei figli rispetto a quelle dei padri : e il termine che descriveva il risultato di questa iniziale applicazione, finì con l'essere impiegato per indicare la tecnica statistica, ed è rimasto ancora oggi nell'uso, anche se l'attributo di "regressione" non avrebbe più alcun significato di essere.

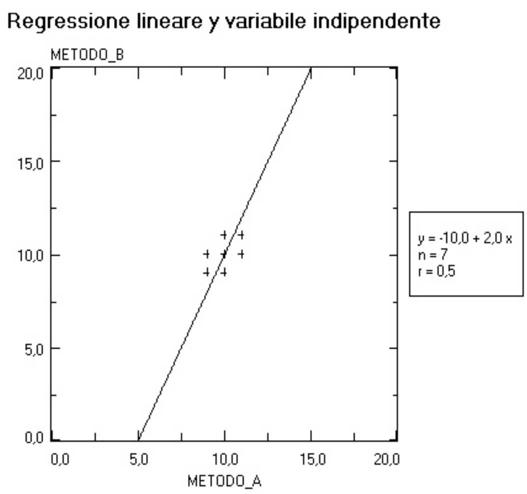
PROBLEMA 6
Si consideri il confronto tra due metodi analitici per la determinazione del calcio nel siero. Il confronto dà i seguenti risultati (concentrazione in milligrammi per decilitro di siero, mg/dL):

Metodo A	Metodo B
9,0	9,0
10,0	10,0
11,0	11,0
9,0	10,0
10,0	9,0
10,0	11,0
11,0	10,0

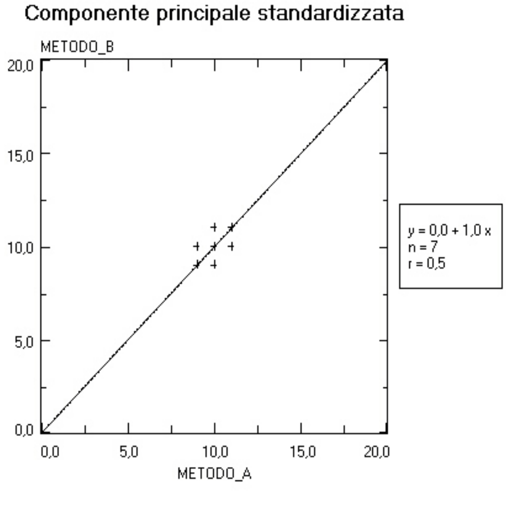
Quali conclusioni possono essere tratte dei risultati ottenuti?

Utilizzando tre diversi modelli di regressione lineare (appendice D), si ottengono i seguenti risultati:





Come è possibile che differenti modelli di regressione forniscano conclusioni così diverse? Quale modello fornisce i risultati più attendibili? Ebbene, gli assunti alla base dei modelli di regressione lineare utilizzati (x variabile indipendente, y variabile indipendente e componente principale standardizzata) sono identici, tranne che per il fatto che la regressione x variabile indipendente assume che la variabile in ascisse (x) sia la variabile misurata senza errore, la regressione y variabile indipendente assume che la variabile in ordinate (y) sia la variabile misurata senza errore, mentre la componente principale standardizzata assume che sia la variabile in ascisse sia la variabile in ordinate siano misurate con un errore dello stesso ordine di grandezza.



La risposta alle due domande è che, nel caso specifico, il rapporto tra il segnale contenuto nei dati e in rumore introdotto dagli assunti intrinseci a ciascuno dei diversi modelli di regressione è troppo basso. In effetti il fatto che r sia uguale solamente a 0,5 indica che l'informazione contenuta nei dati è troppo scarsa per essere attendibile. In questo caso la "forma" che viene dati ai dati, e quindi la conclusione che da essi viene tratta, risulta determinata più dagli assunti del modello matematico (di regressione) che dal contenuto informativo dei dati stessi!

Un terzo esempio è quello relativo all'espressione dei risultati di uno studio clinico controllato in termini di rischio o di rischio relativo (RR)[63]. Dal 19 al 23 marzo 2002 a Barcellona si è tenuto il 3rd European Breast Cancer Conference (EBCC-3). Dove sono stati presentati in maniera preliminare i dati di uno studio svolto su 103.000 donne norvegesi e svedesi, che ha dimostrato un lieve aumento del rischio di sviluppare un cancro della mammella con l'uso di contraccettivi orali (CO). I dati sono stati ripresi e pubblicati con grande clamore da un settimanale molto diffuso. Il messaggio è in sintesi il seguente: dopo cinque anni di assunzione di CO si ha un aumento del 26% di sviluppare un cancro della mammella. L'articolo comunica ovviamente una forte sensazione di allarme a chiunque lo legga. Ma vediamo di analizzare i dati. Innanzitutto la statistica sanitaria ci dice che sette donne su cento avranno un tumore al seno nel corso della loro esistenza. Questo è dunque il rischio basale (RR = 1,0) che ogni donna ha: sette probabilità su cento. In base i risultati dello studio queste donne passano da 7, nel gruppo che non assume CO, a 8,82 nel gruppo che assume CO (8,82/7 = 1,26). Quindi il messaggio è che, in valore assoluto, circa due donne su cento ammaleranno in più nel gruppo che assume CO[64].

Quest'ultimo esempio si colloca in un'area che va dal pre-giudizio del ricercatore, in grado peraltro di "informare" i dati, alla deformazione "propagandistica" del dato. I due precedenti esempi illustravano invece veri e propri "artefatti". Comunque tutti e tre gli esempi sono accomunati dal problema soggiacente: nel dare forma ai dati, questi possono risultarne alterati (e questo anche in assenza, che vogliamo dare per certa, di malafede).

Il terzo elemento sulla via dell'intelligere, la conoscenza, intesa come "prendere possesso intellettualmente … della realtà [connettere]" ci ricollega al metodo di cartesiana memoria e al problema gnoeologico. La conoscenza scientifica non può essere di tipo esclusivamente deduttivo. Ma, seguendo le critiche all'induzione da Hume e K. Popper, non può nemmeno essere di tipo induttivo. Cerchiamo di riassumere questi due fondamentali concetti seguendo l'impostazione data da Peirce. Secondo il quale [7] "*tutta la conoscenza assume la forma dell'inferenza, nel senso che essa è sempre mediata da un ragionamento e non è mai semplicemente intuitiva. Questo è un altro modo per dire che non abbiamo intuizioni immediate, non abbiamo alcun accesso diretto alle cose che ci circondano, ma al contrario tutto quello che possiamo conoscere circa il mondo è frutto di un complesso ragionamento. Tale ragionamento avviene per lo più a livello inconsciio e automatico: non ci rendiamo conto di compierlo. Ciò non toglie, tuttavia, che ogni volta che*

---

[63] Il rischio relativo (RR) (anche detto *relative risk* o *risk ratio*) è il rapporto tra la probabilità che si verifichi un evento in un gruppo esposto (ad un trattamento, ad un fattore di rischio, ad un fattore protettivo), e la probabilità che si verifichi lo stesso evento in un gruppo di non esposti (allo stesso trattamento, al fattore di rischio, al fattore protettivo). Il RR indica la forza dell'associazione fra l'esposizione (o trattamento) e l'esito ed è sempre un numero positivo. Se RR = 1 significa che il rischio che si verifichi l'evento nei 2 gruppi è uguale. In questo caso non c'è associazione fra l'esposizione e l'esito, cioè essere esposti ad un trattamento, o fattore ritenuto protettivo o fattore ritenuto di rischio non modifica la probabilità che un evento si verifichi. Se RR>1 significa che il rischio del verificarsi dell'evento nel gruppo dei trattati o esposti è superiore rispetto al gruppo di controllo. In questo caso l'esposizione è dannosa. Se RR<1 significa che il rischio che si verifichi un evento nel gruppo dei trattati o esposti è inferiore rispetto al gruppo di controllo. In questo caso l'esposizione è protettiva.

[64] A questo va aggiunto che, considerando non solo i potenziali effetti negativi sul carcinoma della mammella, ma anche i possibili effetti positivi su neoplasie di altre sedi, i CO riducono del 49% (RR = 0,51) il rischio relativo di tumori all'ovaio e conferirebbero una riduzione del 40-50% del rischio di sviluppo del cancro endometriale.

impariamo qualche cosa sul mondo, lo facciamo mediante una serie di operazioni concettuali complesse, le inferenze. Gli elementi che possono entrare in gioco in qualsiasi processo inferenziale sono tre: un caso (antecedente, causa), un risultato (conseguente, effetto) e una regola che propone un legame fra i due elementi precedenti. Ad esempio, se il caso è fuoco è il risultato è fumo, la regola dice: se c'è fuoco, c'è fumo. A seconda della disposizione assunta da questi tre elementi, avremo tre diversi tipi di inferenza: la deduzione, l'induzione e l'abduzione. Peirce introduce i tre tipi di inferenza attraverso un esempio ormai celebre, che chiameremo qui l'esempio dei fagioli. Immaginiamo di essere in una stanza dove si trova un tavolo. Sul tavolo c'è un sacco di tela con su scritto fagioli bianchi. Sappiamo dunque che dentro un sacco vi sono fagioli bianchi. Di conseguenza, se estraiamo a caso una manciata di fagioli dal sacco, abbiamo la certezza che essi saranno tutti bianchi (a meno che la scritta sul sacco non ci abbia mentito). Questa è la struttura della deduzione:

REGOLA: Tutti i fagioli in questo sacco sono bianchi
CASO: Questi fagioli provengono da questo sacco
RISULTATO: Questi fagioli sono bianchi (<u>sicuramente</u>)

Come si vede, il ragionamento deduttivo non comporta alcun accrescimento del sapere. Noi sapevamo fin dall'inizio che i fagioli nel sacco erano bianchi, e ci siamo limitati a calcolare le conseguenze logiche di questo assunto: se estraiamo dei fagioli dal sacco, questi saranno necessariamente bianchi.

L'induzione procede diversamente. In questa caso noi non sappiamo ancora che cosa ci sia nel sacco (manca la scritta). Infiliamo la mano nel sacco, ed estraiamo una manciata di quello che vi troviamo dentro. Sono fagioli bianchi. Ma siamo sicuri che nel sacco non ci siano altro che fagioli bianchi? Assolutamente no. Allora ripetiamo l'operazione: ancora fagioli bianchi. Ogni volta che estraiamo una nuova manciata di fagioli bianchi, aumentaimo le possibilità che il sacco contenga solo fagioli bianchi. Ma in linea di principio, non possiamo esserene sicuri fino al momento in cui non abbiamo tirato fuori l'ultimo fagiolo bianco. La struttura logica dell'induzione sarà allora la seguente:

CASO: Questi fagioli provengono da questo sacco
RISULTATO: Questi fagioli sono bianchi
REGOLA: Tutti i fagioli di questo sacco sono bianchi (<u>forse</u>)

L'induzione, dice Peirce, ci consente di allargare orizzontalmente la nostra conoscenza del mondo. La sua essenza è la generalizzazione: noi immaginiamo che ciò che è vero per un certo campione preso a caso da un insieme sia vero anche per tutti gli altri componenti dell'insieme. Non ci vuole molta inventiva per compiere questo salto logico, che comunque è sempre passibile di errore." .

Peirce introduce un terzo tipo di inferenza: l'abduzione [8].

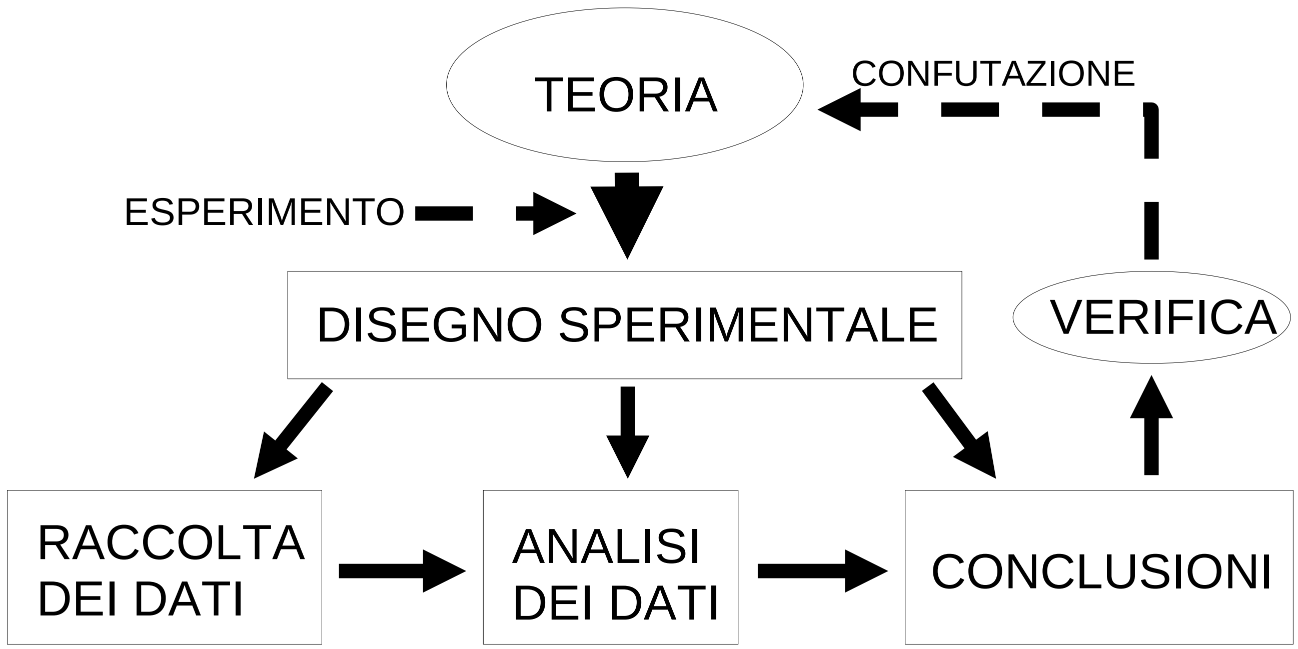
"L'unico modo per penetrare più a fondo nella comprensione delle cose e delle leggi che ne regolano il funzionamento è attraverso la formulazione di ipotesi o abduzioni. Entriamo nella stanza e vediamo il tavolo. Sul tavolo vi sono già sparsi, dei fagioli bianchi, ma noi non sappiamo ancora da dove provengono. Guardando in giro, scopriamo che uno dei sacchi della stanza contiene solo fagiolo bianchi. Cosa facciamo? Congetturiamo (ipotizziamo) che i fagioli bianchi provengano da quel sacco, ossia che costituiscano un caso di questa regola generale. Ma possiamo sbagliarci. Scomponendo il ragionamento nelle sue parti costitutive avremo:

1) RISULTATO: Questi fagioli sono bianchi

2) REGOLA: Tutti i fagioli di questo sacco sono bianchi
3) CASO: Questi fagioli provengono da questo sacco” (<u>forse</u>).

Osservando quello che per il filosofo americano è "un fatto sorprendente" (abbiamo dei fagioli bianchi) e avendo a disposizione una regola in grado di spiegarlo (sappiamo che tutti i fagioli di questo sacchetto sono bianchi) possiamo *ipotizzare* che *si dia il caso* che questi fagioli vengano da questo sacchetto. In questo modo noi abbiamo accresciuto la nostra conoscenza in quanto sappiamo qualcosa di più sui fagioli: prima sapevamo solo che erano bianchi, ora possiamo anche supporre che provengano da questo sacchetto. L'abduzione, secondo Peirce, è l'unica forma di ragionamento suscettibile di accrescere il nostro sapere, ovvero permette di ipotizzare nuove idee, di indovinare, di prevedere. In realtà tutte e tre le inferenze individuate permettono un accrescimento della conoscenza, in ordine e misura differente, ma solo l'abduzione è totalmente dedicata a questo accrescimento. È altresì vero che l'abduzione è il modo inferenziale maggiormente soggetto a rischio di errore. L'abduzione, come l'induzione, non contiene in sé la sua validità logica e deve essere confermata per via empirica.

Ecco che finalmente possiamo capire la frase di A. Einstein: *“L’immaginazione è più importante della conoscenza”* . Al vertice del pensiero scientifico sta l’ipotesi, ovvero la teoria: “*l’immaginazione*” nelle parole di Einstein, o ancora “*l’abduzione*” nella parole di Peirce. E’ questo il fatto veramente innovativo sul piano della conoscenza. L’esperimento viene concepito per verificare la teoria. Esperimenti che portano a risultati in accordo con la teoria ne aumentano la plausibilità (probabilità), anche se non ne potranno mai asserire la veridicità (verità). Esperimenti che portano ad osservazioni e conseguenze che non sono spiegabili con la teoria, sono in grado di confutarla (falsificarla).



Procede sempre nello stesso modo il “ricercatore”: che sia fisico teorico che formula un’ipotesi sulla struttura della materia, che sia il medico che formula un’ipotesi sulla malattia del paziente, che sia il detective che formula un’ipotesi sul responsabile del delitto. Formula, appunto, un’<u>ipotesi</u>, e ne <u>deduce</u> delle conseguenze che sottopone a <u>verifica sperimentale</u> (empirica). Non è vero che non abbia un pre-giudizio. Anzi paradossalmente lo deve avere. Perché è proprio l’immaginazione che innesca il ciclo del modello ipotetico-deduttivo, a sua volta motore, con l’esperimento, della conoscenza. Una conoscenza mai certa, ma sempre aperta a nuove ipotesi, a nuove deduzioni, a nuove verifiche sperimentali.

Bibliografia e riferimenti

[1] Vocabolario della lingua italiana di Nicola Zingarelli. Zanichelli, 1994.
[2] De Angelis G, Franzini C. Nomenclatura, grandezze e unità di misura nel laboratorio di analisi chimico-cliniche. Giorn. It. Chim. Clin., Vol. 13, 1988, pag. 1-50.
[3] Laughlin R. Un universo diverso. Reinventare la fisica da cima a fondo. Codice, 2005.
[4] http://it.wikipedia.org/wiki/Effetto_Hall_quantistico
[5] Taylor JR. Introduzione all'analisi degli errori. Lo studio delle incertezze nelle misure fisiche. Zanichelli, 1986.
[6] http://it.wikipedia.org/wiki/Gauss
[7] Braga M. Argomenti di biometria. Quante cifre deve avere un numero per essere statisticamente significativo? Aggiornamento del medico 1990;14:XXXI-XXXVI.
[8] Volli U. Manuale di semiotica. Laterza, 2003.

5. Complessità, probabilità e teorema di Bayes

“Il concetto di probabilità è il più importante della scienza moderna, soprattutto perché nessuno ha la più pallida idea del suo significato.”
(Bertrand Russel)

“La probabilità: chi è costei? Prima di rispondere a tale domanda è certamente opportuno chiedersi: ma davvero “esiste” la probabilità? e cosa mai sarebbe? Io risponderei di no, che non esiste.”
(Bruno de Finetti)

Il secondo dei due fattori chiave della conoscenza è rappresentato dalla correttezza delle regole: si devono sapere/devono esser giuste e devono essere applicate correttamente le regole. I problemi emergono quando si ignorano/sono sbagliate o non sono applicate correttamente le regole.

Il nostro problema cruciale è, come abbiamo visto, essere costretti a decidere in condizioni di incertezza. E poiché questa, come abbiamo capito, risulta ineliminabile, la sola cosa che possiamo ragionevolmente fare, per non subirla, è imparare a gestirla. Come? Andando alla ricerca delle regole del pre-vedere. E qui il primo elemento che incontriamo, e che domina sovrano, è il caso.

Erwin Schroedinger (1887-1961) il fisico austriaco fondatore della meccanica ondulatoria, premio Nobel nel 1933 con Paul Dirac, diceva: "*...la ricerca in fisica ha mostrato, al di là di ogni dubbio, che l'elemento comune soggiacente alla coerenza che si osserva nella stragrande maggioranza dei fenomeni, la cui regolarità e invariabilità hanno consentito la formulazione del postulato di causalità, è il caso...*".

Reciprocamente si può dire che sembra esservi un determinismo soggiacente ai sistemi caotici, che tende a emergere in seguito al comportamento "collettivo" di eventi tra loro indipendenti. L'esempio più semplice è quello della roulette. Non è possibile sapere in anticipo, al lancio della pallina, se essa si fermerà sul colore rosso o sul colore nero (a meno che la roulette sia truccata: ma qui si assume che non lo sia). Il risultato di un singolo evento è puramente casuale. Tuttavia se si lancia la pallina diciamo un miliardo di volte, possiamo avere la "certezza" che la metà delle volte essa si fermerà sul rosso e l’altra metà delle volte essa si fermerà sul nero.

In altri termini, quando ci si trova davanti a evento singolo quello che prevale è il *caso*; quando gli eventi sono molto numerosi, sembra esservi una *necessità* alla quale gli eventi finiscono con l'ubbidire. Questa è in estrema sintesi l'essenza della probabilità, ed è anche, detto per inciso, la chiave di lettura che la biologia moderna ha adottato per spiegare l’emergere spontaneo della *vita/necessità* dal *caos/caso* primordiale, come descritto da J. Monod [1].

Il sottile filo che lega la visione di Schroedinger (fisico) e di Monod (medico, anche lui premio Nobel) passa attraverso le scoperte dei fisici dei primi decenni di questo secolo, le cui conseguenze sono ancora oggi oggetto di approfondimento sia teorico sia sperimentale, e le cui conseguenze pratiche sono alla base dello sviluppo delle moderne tecnologie, che stanno aprendo la strada ad innovazioni di portata epocale (i calcolatori quantistici potrebbero essere ormai alle porte, mentre non à escluso che alcuni meccanismi, come quello che risiede alla base della memoria all’interno dei neuroni, siano realizzati a livello quantistico). Queste scoperte, che videro coinvolti Bohr, Einstein, lo stesso Schroedinger, e molti altri fisici, sono basate sull’osservazione che alcuni “strani” comportamenti della natura risultano descrivibili mediante leggi esprimibili solamente in termini probabilistici. L’argomento è di interesse fondamentale per comprendere fino in fondo come la probabilità sia uno strumento di conoscenza. L'unico scotto da pagare è quello di rinunciare alla "certezza", e di accettare una descrizione "probabilistica" degli eventi. Uno scotto pesante, che addirittura lo stesso Einstein inizialmente rifiutò, ma che con il passare dei decenni ha finito con l'essere progressivamente accettato.

I filosofi della scienza attribuiscono al termine probabilità due significati o valori: e parlano di *probabilità epistemica* e di *probabilità non epistemica* [2].

Nel caso di processi probabilistico nei quali la probabilità risulta epistemica, si ha a che fare con conclusioni che vengono espresse in modo probabilistico a causa della nostra ignoranza sullo stato reale del sistema in esame. Come ricordato da Ghirardi nell'opera sopraccitata, il concetto di probabilità epistemica rispecchia perfettamente la posizione meccanicistica del grande matematico francese Simon de Laplace, che nel 1776 scriveva "*Lo stato attuale del sistema della natura consegue evidentemente da quello che esso era nell'istante precedente, e se noi immaginassimo un'intelligenza che ad un dato istante comprendesse tutte le relazioni fra le entità di questo universo, esso potrebbe conoscere le rispettive posizioni, i moti e le disposizioni generali di tutte quelle entità in qualunque istante del passato e del futuro… Ma l'ignoranza delle diverse cause che concorrono alla formazione degli eventi come pure la loro complessità, insieme coll'imperfezione dell'analisi, ci impediscono di conseguire la stessa certezza rispetto alla grande maggioranza dei fenomeni. Vi sono quindi cose che per noi sono incerte, cose più o meno probabili, e noi cerchiamo di rimediare all'impossibilità di conoscerle determinando i loro diversi gradi di verosimiglianza. Accade così che alla debolezza della mente umana si debba una delle più fini e ingegnose fra le teorie matematiche, la scienza del caso o della probabilità*". Secondo Laplace, note le posizioni e le velocità di tutte le particelle dell'universo, e le leggi che ne governano i rapporti, sarebbe stato possibile prevederne l'evoluzione per l'eternità. La concezione di Laplace configura la probabilità nella descrizione dei processi fisici come accidentale, legata alla nostra ignoranza, ma in linea di principio eludibile.

Come ricorda Ghirardi, nel 1903 il grande matematico Jules-Henri Poincarè scriveva: "*Una causa piccolissima che sfugga alla nostra attenzione determina un effetto considerevole, che non possiamo mancare di vedere, e allora diciamo che l'effetto è dovuto al caso. Se conoscessimo esattamente le leggi della natura e la situazione dell'universo all'istante iniziale, potremmo prevedere esattamente la situazione dello stesso universo in un istante successivo. Ma se pure accadesse che le leggi naturali non avessero più alcun segreto per noi, anche in tal caso potremmo conoscere la situazione iniziale solo approssimativamente. Se questo ci permettesse di prevedere la situazione successiva con la stessa approssimazione non ci occorrerebbe di più e dovremmo dire che il fenomeno è stato previsto, che è governato da leggi. Ma non sempre è così: può accadere che piccole differenze nelle condizioni iniziali ne producano di grandissime nei fenomeni finali. Un piccolo errore nelle prime produce un errore enorme nei secondi. La previsione diviene impossibile, e si ha un fenomeno fortuito*".

L'estrema sensibilità alle condizioni iniziali descritta da Poincarè, apre la strada ai moderni concetti di "caos deterministico" e di "complessità". Laddove nella sua accezione più generale il concetto di complessità pone in crisi l'idea che in ogni caso lo studio dei sistemi complessi possa ricondursi allo studio dei loro costituenti.

Come commenta Ghirardi "*di fatto risulta relativamente facile dimostrare che esistono sistemi deterministici* [come il banale lancio di una moneta, N.d.A.] *con una tale sensibilità alle condizioni iniziali che la previsione del loro comportamento anche dopo tempi brevi richiederebbe una tale massa di informazioni (proprio perché le imprecisioni iniziali si amplificano esponenzialmente) che non potrebbero venire immagazzinate neppure in un computer che utilizzasse come chips tutte le* *particelle dell'universo e potesse immagazzinare un bit in ogni chip. La conclusione è che ci si è resi conto (e questo rappresenta indubbiamente una notevole conquista concettuale) che non sono rare situazioni in cui risulta di fatto impossibile prevedere il comportamento di un sistema per un periodo di tempo anche ragionevolmente breve… Il fatto che se anche tutto l'universo diventasse un calcolatore esso non risulterebbe abbastanza potente da permetterci di immagazzinare le informazioni necessarie a prevedere per più di qualche minuto l'evoluzione di un semplice sistema, non toglie nulla al fatto che secondo lo schema teorico che si è assunto soggiacere alla dinamica del processo, la necessità di ricorrere ad una descrizione probabilistica deriva dall'ignoranza circa le precise condizioni iniziali… Al contrario nelle schema quantomeccanico…l'aleatorietà degli esiti è incorporata nella struttura stesa del formalismo che, se assunto come completo, non consente neppure di pensare che, in generale, gli esiti siano, anche se in un modo a noi sconosciuto, predeterminati*".

La meccanica quantistica introduce il concetto di probabilità non epistemica, cioè di descrizione probabilistica degli eventi che non può essere attribuita ad ignoranza. In altre parole, la descrizione probabilistica degli eventi non può essere attribuita ad una mancanza di informazioni sul sistema che, se fosse disponibile, ci consentirebbe di trasformare le asserzioni probabilistiche in asserzioni certe. I processi fisici microscopici sono fondamentalmente stocastici, e hanno una intrinseca e irriducibile aleatorietà.

Per concludere, interessantissima la presentazione del volume curato da Casati [3] nella quale il curatore, tra le altre cose, dice: "*…è interessante osservare che la meccanica quantistica è una teoria intrinsecamente probabilistica. Una volta assegnato lo stato di un sistema al tempo t mediante una "funzione di stato" $\psi(t)$ noi siamo in grado di fare affermazioni solamente sulla probabilità che eseguendo una misura su una data grandezza si ottenga un determinato valore. Tuttavia è diverso il discorso relativo alla previsione della evoluzione futura. Infatti in meccanica quantistica lo stato $\psi(t)$ del sistema al tempo t è determinato univocamente dallo stato iniziale $\psi(0)$. Problema: siamo in grado, date le leggi del moto e dato $\psi(0)$, di determinare $\psi(t)$? Il fatto straordinariamente inaspettato è che, a differenza della meccanica classica, la risposta a questa domanda è positiva: per i sistemi quantistici è possibile, almeno in linea di principio, risolvere le equazioni del moto e predire lo stato futuro $\psi(t)$. Pertanto il quadro che si va delineando è diametralmente opposto a quello che si aveva in precedenza: la meccanica classica è sempre stata considerata come una teoria deterministica; ora abbiamo visto che, a causa dell'insorgere del moto caotico, essa porta ad un comportamento statistico. D'altro lato la meccanica quantistica è intrinsecamente probabilistica; tuttavia, grazie al suo carattere di stabilità, risulta essere più predicibile della meccanica classica… La domanda che il lettore certamente si pone è quali implicazioni può avere lo studio del caos. In altri termini, a quale utilità pratica può portare il sapere che il comportamento della gran arte dei sistemi deterministici è in realtà così complicato da apparire completamente caotico, e che quindi essi si sottraggono alla nostra capacità di previsione? Anzitutto abbiamo imparato una lezione molto importante: leggi semplici non portano necessariamente a comportamenti semplici. Sarebbe alquanto vantaggioso se questo concetto fosse tenuto presente non solo nelle discipline scientifiche, ma anche nella vita politica ed economica. Un'altra lezione è che variazioni piccole nei parametri di un qualunque sistema non portano, necessariamente, a variazioni piccole nel "risultato", cioè nella evoluzione futura: per esempio un aumento del 5% nell'inquinamento non sempre porta a un peggioramento solo del 5% nel danno ecologico. Una delle caratteristiche dello studio dei fenomeni caotici è la enorme potenzialità di una unificazione culturale in cui tutta la "filosofia naturale" e le discipline economiche, umanistiche, politiche e sociali sono coinvolte. La natura stessa sembra usare il caos nel suo programma di evoluzione: ogni schema deterministico fallirebbe se utilizzato per la sopravvivenza delle forme di vita in condizioni ambientali in continua trasformazione; la natura, pertanto, genera una quantità enorme di forme di vita attraverso mutazioni casuali e, da questa ampia possibilità di* *scelta, la selezione naturale trova candidati che si adattano alle mutate condizioni ambientali… tutte le volte che nel suo faticoso ed esaltante cammino verso la comprensione dei fenomeni naturali, l'uomo si è trovato di fronte a delle limitazioni, ciò è stato l'occasione per nuovi grandi balzi in avanti che hanno comportato un rovesciamento della filosofia precedente. Mi riferisco alla osservazione del valore limite della velocità della luce (che ha portato alla teoria della relatività), alla limitazione sulla precisione delle nostre misure (che ha portato alla meccanica quantistica); la teoria del caos ci ha messo di fronte a una limitazione ancora maggiore: l'impossibilità di prevedere il futuro…*". Quest'ultima affermazione ci collega con il problema della previsione degli eventi futuri.

Durante i primi millenni della cultura dell'uomo, e fino all'avvento della *scienza* moderna (in questo caso delle branca della matematica nota come *teoria della probabilità*), la soluzione al problema della previsione di eventi futuri è stata fornita dalla *magia*.

Si consideri un problema apparentemente banale, ma paradigmatico, come quello delle previsioni del tempo. Visto che vorrei fare un week-end al mare, ma vorrei evitare di trascorrerlo sotto la pioggia, e la domanda è: domani pioverà o no?

L'approccio magico per rispondere a questa domanda, procede utilizzando un ragionamento analogico <u>non</u> basato su dati misurabili: il mago basa la previsione sulle sue capacità di "intuire" il tempo che farà domani.

L'approccio scientifico "deterministico", quello che pretende di avere previsioni "certe", si differenzia da quello magico per il fatto che la previsione non viene effettuata sulla base di una "intuizione", bensì sulla base di dati misurabili e di leggi/modelli in forma matematica. Si misura la temperatura, la pressione atmosferica, si elaborano modelli matematici della circolazione atmosferica, sulla base dei quali si prevede il tempo che farà domani. Tuttavia si scopre che la previsione, per quanti sforzi siano fatti in termini di numero dei dati misurati dai quali si parte, e di complessità dei modelli matematici utilizzati, non porta mai alla "certezza". E la visione scientifica deterministica, quella per intenderci della fisica di Newton, per la quale la scoperta di alcune delle leggi fondamentali che governano la natura (la legge di gravitazione) fa intendere il cosmo come un gigantesco orologio rigorosamente determinato nelle sue funzioni, deve lasciare il passo ad una visione scientifica probabilistica, e ad una interpretazione probabilistica dei problemi [in realtà, come ormai dovrebbe essere chiaro da quanto finora detto, la sviluppo dell'approccio scientifico probabilistico non è dovuto alla meteorologia, bensì alla scoperte dei fisici: in particolare alla scoperta delle leggi che governano la natura a livello degli elementi fondamentali che la costituiscono (*atomi* e *quanti*)] [4].

Ecco quindi il significato profondo della frase di E. Schroedinger precedentemente citata. La "*…coerenza che si osserva nella stragrande maggioranza dei fenomeni, la cui regolarità e invariabilità hanno consentito la formulazione del postulato di causalità...*", ha portato a Newton e i primi scienziati a vedere il cosmo come un orologio perfettamente determinato. Data l'assoluta regolarità e apparente inviolabilità delle leggi che governano il moto dei corpi celesti, è "realmente" possibile prevedere con 72 anni di anticipo il ripresentarsi della cometa di Halley, e la presenza di Plutone può essere prevista dallo studio delle perturbazioni del pianeta Urano molti anni prima che Plutone venga effettivamente osservato, esattamente nella posizione prevista. A livello *macroscopico* il principio di causalità funziona, al punto da divenire un postulato: il che implica una dichiarazione di certezza "a priori" della sua validità. Tuttavia a partire dai primi del '900, la fisica scopre che le leggi che governano il comportamento degli elementi ultimi che costituiscono la materia/energia dell'universo, possono essere descritte solamente mediante formulazioni di tipo probabilistico. A livello *microscopico* (intendendo con ciò, come detto, atomi e quanti)

"*...l'elemento comune soggiacente…è il caso...*". Ad esempio l'orbitale, che descrive il moto di un elettrone attorno al nucleo di un atomo, è una funzione d'onda che fornisce la probabilità di "trovare" l'elettrone.

Il percorso che ha portato dalla spiegazione magica della realtà alla spiegazione della realtà data dalle leggi della scienza "deterministica" applicabile al mondo macroscopico, prima, e alla spiegazione della realtà data dalle leggi della scienza "probabilistica" applicabile al mondo microscopico, poi, è come detto una delle avventure più straordinarie del pensiero umano. Ed è anche stato il percorso che ha portato allo sviluppo delle tecnologie che stanno alla base del mondo in cui viviamo. La *teoria dei quanti* è rigorosamente probabilistica: ma senza di essa, ad esempio, non esisterebbe la microelettronica, e non esisterebbe il PC su cui sto scrivendo.

Il rapporto tra l'approccio fornito dalla concezione magica e l'approccio fornito dalla concezione scientifica nei confronti del problema della previsione di eventi futuri, è illustrato qui di seguito:

L'approccio	***Lo strumento di previsione***	***Utilizzo di dati misurabili***	***Il livello cui si applica***	***La previsione***
Magia	Intuizione	No	Macroscopico	Apparentemente "certa", se si accettano in modo fideistico previsioni del mago. Ma si può dimostrare che anche i migliori maghi sbagliano.
Scienza	Modello matematico deterministico	Si	Macroscopico	"Certa" in riferimento ad eventi come il moto degli astri nelle orbite determinate dalla legge di gravitazione, e a *quasi tutte* le leggi che governano il mondo macroscopico.
	Modello matematico probabilistico	Si Si	Macroscopico Microscopico	"Probabilistica" in riferimento ad eventi complessi come la circolazione dell'atmosfera (previsioni del tempo). "Probabilistica" in riferimento agli eventi che caratterizzano elementi ultimi che costituiscono la materia/energia (teoria atomica e teoria dei quanti).

Per meglio illustrare questi concetti, e le loro profonde implicazioni, bisogna ora necessariamente fare un passo indietro, riconsiderando il problema del lancio della moneta, e la domanda "al prossimo lancio uscirà testa o croce?".

In merito a questa domanda, il rapporto tra concezione magica, visione deterministica e visione probabilistica è illustrato qui di seguito:

La domanda	***Lo strumento di previsione***	***La previsione***
	La magia	Può sembrare che funzioni
	La risposta deterministica (*conclusione “certa”*)	Funziona solo se la moneta è truccata (in questo caso lanciando in un modo particolare la moneta è possibile ottenere “deterministicamente” (con certezza) un certo risultato.
Testa o croce?	La risposta probabilistica (*conclusione “probabile”*)	Ci consente di affermare che, se il risultato di un singolo lancio è imprevedibile (legato al caso), a lungo andare metà delle volte uscirà testa e l’altra metà delle volte uscirà croce (necessità), e di esprimere quindi il risultato del lancio della moneta in termini di probabilità (la probabilità che in un dato lancio esca testa è identica alla probabilità che esca croce, ed è $p = 0{,}5$)

Per approfondire ulteriormente il significato insito nella rinuncia alla visione deterministica e il passaggio ad una risposta probabilistica, si consideri ora di nuovo la domanda “domani pioverà o no?”.

Lo speaker che illustra le previsioni del tempo dice “domani generalmente soleggiato con possibilità di piovaschi”. Si tratta di un modo colloquiale di esprimere una probabilità, diciamo (i valori effettivi di p sono ovviamente qui irrilevanti) del 90% ($p = 0{,}90$) che faccia bello e del 10% ($p = 0{,}10$) che piova. Lo speaker afferma in questo modo che, date condizioni meteorologiche quali quelle odierne, e date 100 osservazioni del tempo che ha fatto l’indomani, si è osservato che 90 volte l’indomani ha fatto bello e che 10 volte l’indomani ha piovuto.

La mia domanda diventa allora “ Visto che vorrei fare un week-end al mare, ma vorrei evitare di trascorrerlo sotto la pioggia, mi piacerebbe tanto sapere: *ma domani è uno dei 90 giorni che farà bello o è uno dei 10 che giorni che pioverà?*”. Ebbene, come ormai dovrebbe essere chiaro, essendo la risposta di tipo probabilistico, la mia legittima pretesa di certezza non potrà mai essere soddisfatta. Il rapporto tra concezione magica, visione deterministica e visione probabilistica è ulteriormente sintetizzato qui di seguito:

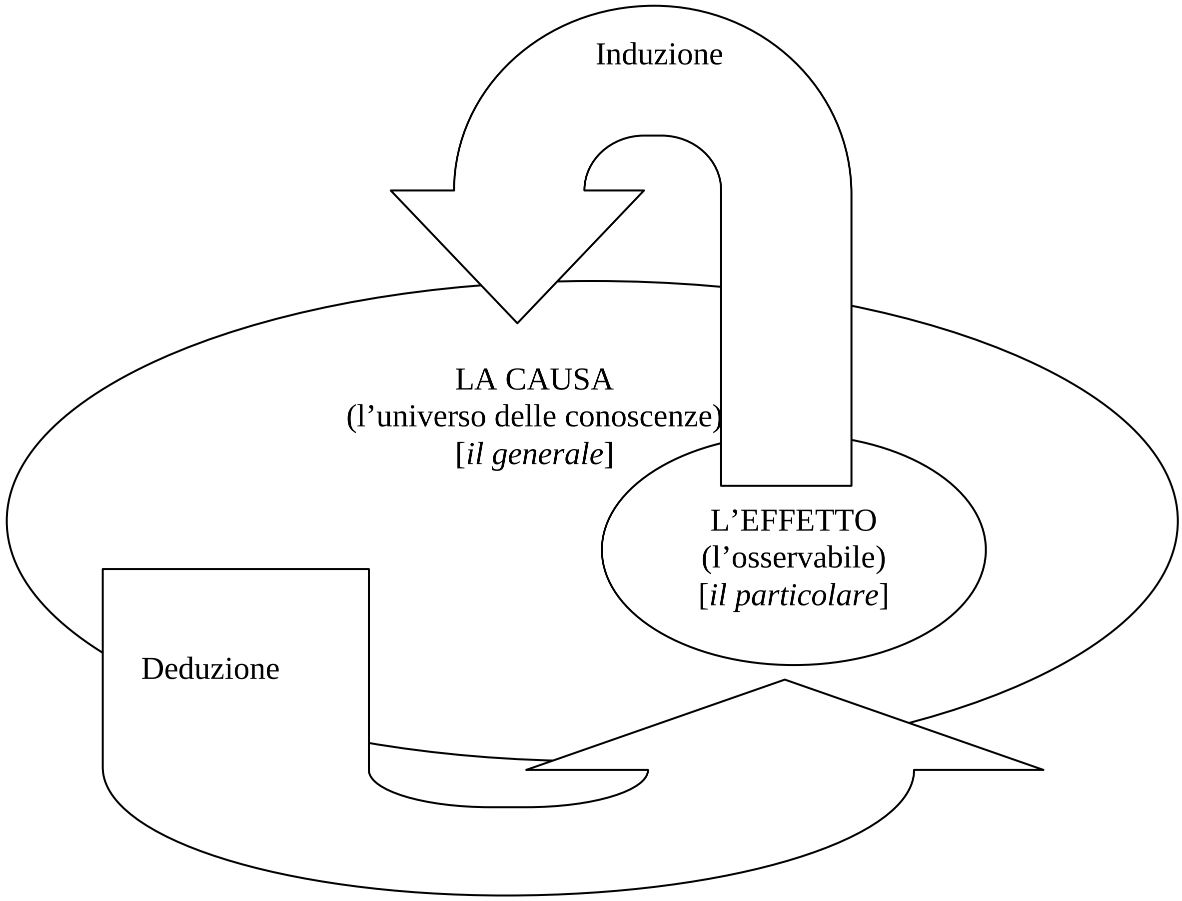
La domanda	***Lo strumento di previsione***	***La previsione***
	Intuizione (*magia*)	Può sembrare che funzioni.
	Modello matematico “deterministico” (*conclusione “certa”*)	Non è possibile.
Domani pioverà?	Modello matematico “probabilistico” (*conclusione “probabile”*)	Ci consente di affermare che (ad esempio) domani c’è il 90% di probabilità che faccia bello e il 10% di *probabilità* che piova, rinunciando peraltro alla *certezza* di sapere se domani sarà uno dei 90 giorni che fa bello o piuttosto uno dei 10 giorni che pioverà.

La famosa espressione attribuita ad A. Einstein “*…non posso credere che Dio giochi ai dadi…*” si riferisce proprio a questo. Alla difficoltà che si incontra nell’accettare che buona parte delle leggi di natura siano intrinsecamente probabilistiche. Eppure il paradosso del mentitore, il teorema di Gödel e l’epistemologia di Karl Popper ci confermano che deve essere così.

Il collegamento tra i paradossi individuati dai greci e il pensiero moderno è reso possibile da una serie di sviluppi che partono ancora una volta dai giochi, e precisamente dagli studi pionieristici “De ludo aleae” di G. Cardano (1501-1576) e “Sopra le scoperte de li dadi” di G. Galilei (1564-1642). Anche se sono le questioni poste dal cavaliere De Merè e le risposte fornite da B. Pascal (1623-1662) e P. De Fermat (1601-1665) a fondare il calcolo delle probabilità.

Fino al ‘700 è quello che viene definito il “problema classico” a dominare l’attenzione. Abbiamo un’urna contenente 500 palline di colore bianco e 500 palline di colore rosso. Cosa ci possiamo attendere dall’estrazione di una pallina? Si sa tutto sull’urna, ovvero si conosce “l’universo”, ovvero si conosce la causa. Si applica un ragionamento deduttivo. Il risultato (l’effetto, l’estrazione di una pallina) può essere calcolato.

Ma nel 1763 viene pubblicato postumo, dal suo amico Richard Price, il saggio di un reverendo inglese di nome Thomas Bayes (1702-1761), che indica la soluzione del “problema inverso” [5]. Da un’urna contenente *s* palline estraiamo *n* palline di cui *k* sono di colore rosso. Cosa possiamo concludere circa il contenuto dell’urna? Si è fatto un esperimento, si conosce l’effetto. Il problema che Bayes si pone è: esiste un qualche ragionamento induttivo che ci consenta di “calcolare” la causa (lo specifico contenuto dell’urna)?



Il matematico francese Pierre-Simon Laplace (1749-1827) replica ed estende questo risultato in un saggio del 1774, apparentemente ignaro dei risultati di Bayes. Bayes e Laplace rispondono così in linguaggio matematico al vecchio problema di Hume: sorgerà il sole domani?

Immaginiamo un alieno che arrivi sulla terra, di notte. L'indomani vede per la prima volta sorgere il sole, quindi lo vede tramontare e scomparire. Chiediamogli che grado di fiducia ripone nel verificarsi dell'evento ricomparsa del sole. Non avendo alcuna altra informazione, la sola risposta ragionevole è 1/2: la probabilità che sorga, e la probabilità che non sorga, sono per lui, a priori, uguali. Ma l’indomani vede il sole sorgere una seconda volta. Quindi il terzo giorno lo vede sorgere ancora, e così via. Chiediamogli ogni volta che grado di fiducia ripone nel verificarsi dell'evento ricomparsa del sole. La soluzione, dopo *n* giorni, è

$$P = (n+1) / (n+2) \qquad (5.1)$$

mentre più in generale per il problema dell'urna contenente *s* palline dalla quale estraiamo *n* palline di cui *k* sono di colore rosso la probabilità che la prossima pallina sia rossa è

$$P = (k+1) / (n+2) \qquad (5.2)$$

I due fatti notevoli sono che la probabilità P non dipende da s, e che il valore di P tende a 1 ma non potrà mai a raggiungere esattamente tale valore. Quindi l'induzione non potrà mai essere certa.

Fatte queste premesse di tipo epistemologico e di tipo storico, è possibile affrontare l'argomento con un po' più di sistematicità. Per scoprire subito che, in onore ai motti di Russel e De Finetti presentati all'inizio del capitolo, sono oggi ben quattro le definizioni di probabilità.

La definizione classica (Laplace) dice che:

"*La probabilità è il rapporto fra il numero di eventi favorevoli e il numero di eventi possibili, essendo questi ultimi tutti equiprobabili*" ovvero

$$P(A) = \frac{nA}{n} \qquad (5.3)$$

da notare che la definizione classica di probabilità contiene un vizio logico, in quanto la probabilità viene utilizzata per definire sé stessa.

La definizione frequentista (Von Mises) dice che:

"*La probabilità di un evento è il rapporto fra il numero di esperimenti in cui esso si è verificato e il numero totale di esperimenti eseguiti nelle stesse condizioni, essendo tale numero opportunamente grande*"

$$P(A) = \lim_{n \to \infty} \frac{nA}{n} \qquad (5.4)$$

La definizione soggettivista (De Finetti) dice che:

"...la probabilità che qualcuno attribuisce alla verità - o al verificarsi - di un certo evento (fatto singolo univocamente descritto e precisato) altro non è che la misura del grado di fiducia nel suo verificarsi"

La definizione assiomatica (Kolmogorov) dice che :

"La probabilità è un numero compreso tra 0 (evento impossibile) e 1 (evento certo) che soddisfa i tre assiomi di Kolmogorov[65]*"*

L'impostazione assiomatica della probabilità venne proposta da Andrey Nikolaevich Kolmogorov nel 1933 in *Grundbegriffe der Wahrscheinlichkeitsrechnung* (*Concetti fondamentali del calcolo delle probabilità*), sviluppando la ricerca che era ormai cristallizzata sul dibattito fra quanti consideravano la probabilità come limiti di frequenze relative (impostazione frequentista) e quanti cercavano un fondamento logico della stessa. La sua impostazione assiomatica si mostrava adeguata a prescindere dall'adesione a una o all'altra scuola di pensiero.

Un esempio dovuto a De Finetti consente di illustrare la differenza tra le prime tre definizioni. Immaginiamo una partita di calcio per la quale gli eventi possibili sono:
- la vittoria della squadra di casa;
- la vittoria della squadra ospite;
- il pareggio.

Secondo la teoria classica esiste 1 probabilità su 3 che avvenga la vittoria della squadra di casa.
Secondo la teoria frequentista ci si può dotare di un almanacco, controllare tutte le partite precedenti e calcolare la frequenza di un evento.
Secondo la teoria soggettiva, ci si può documentare sullo stato di forma dei calciatori, sul terreno di gioco e così via fino ad emettere un giudizio di probabilità (soggettiva).

Per affrontare il teorema di Bayes sono necessarie alcune definizioni. Innanzitutto la probabilità marginale che si verifichi l'evento A è

$$P(A) \qquad (5.5)$$

Dal punto di vista numerico la probabilità di un evento è un numero positivo compreso tra 0 (evento che non accade mai) e 1 (evento certo) ovvero

$$0 \leq P(A) \leq 1 \qquad (5.6)$$

Corollario: la probabilità che non si verifichi l'evento A è

$$P(nonA) = 1 - P(A) \qquad (5.7)$$

La probabilità condizionata

$$P(A|B) \qquad (5.8)$$

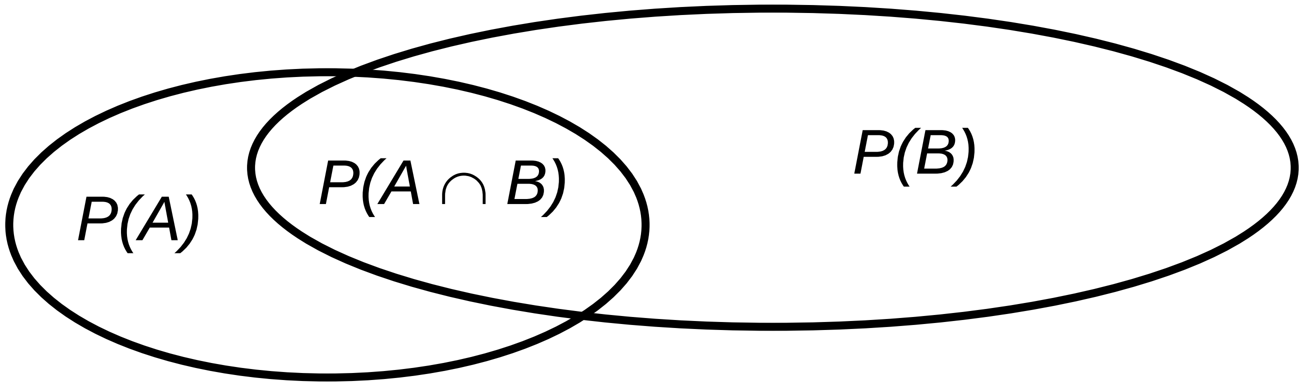
è la probabilità di un evento A condizionata ad un evento B, ovvero è la probabilità che si verifichi A a condizione che si sia verificato B e si legge "*la probabilità di A, dato B*" ovvero "*la probabilità di A condizionata a B*"

[65] "La probabilità assiomatica è una funzione d'insieme **P** definita sullo spazio degli eventi **S**, ovvero è una legge in grado di assegnare ad ogni evento **E** appartenente ad **S** un numero che soddisfa i tre assiomi di Kolmogorov:
1) la probabilità **P**(**E**) di un evento **E** è un numero reale non negativo;
2) la probabilità **P**(**U**) dell'evento certo è 1;
3) la probabilità di un evento complesso costituito dal verificarsi dell'evento elementare **A** o dell'evento elementare **B**, mutuamente incompatibili, è la somma delle probabilità di **A** e di **B**: **P**(**A** o **B**) = **P**(**A**) + **P**(**B**)"

La <u>probabilità congiunta</u>

$$P(A \cap B) \qquad (5.9)$$

è la probabilità di due eventi congiunti, ovvero <u>è la probabilità che si verifichino sia A sia B</u>, e si legge "*la probabilità congiunta di A e B*" ovvero "la *probabilità di A e B*". La probabilità congiunta può essere meglio compresa facendo riferimento al diagramma seguente.



La relazione fra probabilità condizionata e probabilità congiunta è la seguente:

$$P(A \cap B) = P(A|B) \cdot P(B) \qquad (5.10)$$

ma deve essere anche

$$P(A \cap B) = P(B|A) \cdot P(A) \qquad (5.11)$$

(perchè, intuitivamente, dobbiamo avere lo stesso risultato sia partendo da A sia partendo da B) e, combinando la (5.10) con la (5.11), si ottiene il <u>teorema delle probabilità composte</u> :

$$P(A \cap B) = P(A|B) \cdot P(B) = P(B|A) \cdot P(A) \qquad (5.12)$$

Dal precedente teorema si ricava il <u>teorema di Bayes</u>:

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)} \qquad (5.13)$$

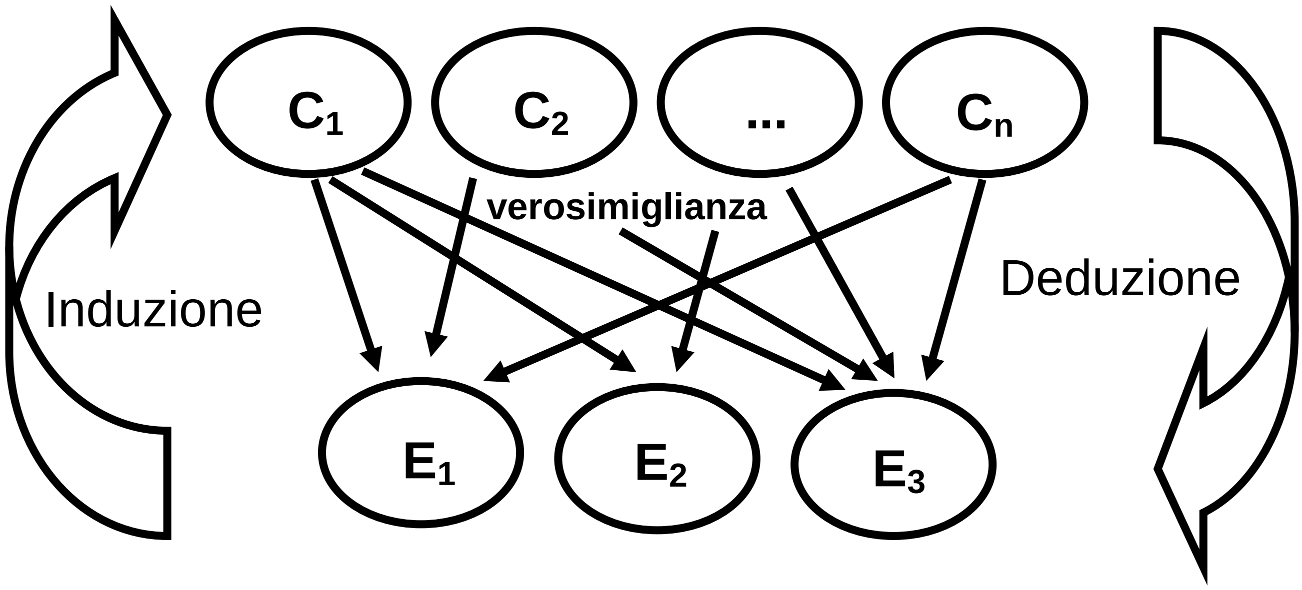
Se poniamo *A = la causa*, e *B = l'effetto*, il teorema di Bayes suona così

$$P(\text{della causa}|\text{dato l'effetto}) = \frac{P(\text{dell'effetto}|\text{data la causa}) \cdot P(\text{della causa})}{P(\text{dell'effetto})} \qquad (5.14)$$

E finalmente se consideriamo il problema specifico della diagnostica medica, e poniamo *A = la malattia* [*la causa*], e *B = il segno* [*l'effetto*], il teorema di Bayes suona così

$$P(\text{della malattia}|\text{dato il segno}) = \frac{P(\text{del segno}|\text{data la malattia}) \cdot P(\text{della malattia})}{P(\text{del segno})} \qquad (5.15)$$

Si consideri uno schema come il seguente, nel quale C rappresenta una delle tante cause osservabili, E rappresenta uno dei tanti effetti osservabili, e le frecce dirette dalla causa all'effetto rappresentano i rapporti di cause/effetto.



Il problema classico può essere risolto applicando la deduzione. Conoscendo le cause, possiamo da queste dedurre gli effetti corrispondenti. La soluzione del problema inverso, ovvero l'induzione, è oltremodo più difficile, in quanto ad un effetto possono corrispondere più cause. Il teorema di Bayes consente, a partire dagli effetti osservati, di calcolare la verosimiglianza delle cause, espressa ovviamente nell'unico modo possibile: in termini di probabilità.

Bibliografia e riferimenti

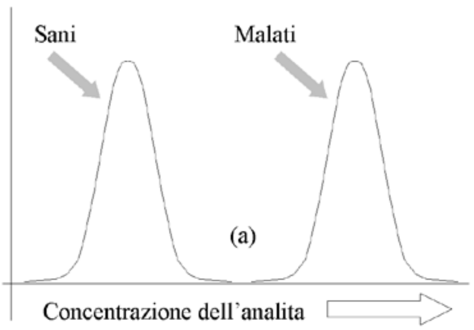
[1] Monod J. Il caso e la necessità. Saggio sulla filosofia naturale della biologia contemporanea. EST Edizioni Scientifiche e Tecniche Mondadori, 1972.
[2] Ghirardi GC. Un'occhiata alle carte di Dio. Gli interrogativi che la scienza moderna pone all'uomo. Il Saggiatore, 1997.
[3] Casati G, (a cura di). Il caos. Le leggi del disordine. Le Scienze, 1991.
[4] Barrow JD. Il mondo dentro il mondo. Adelphi, 1991.
[5] Reverend Thomas Bayes. An assay toward solving a problem in the doctrine of chance. Philo. Trans. Roy. Soc., vol. 53, 370-418, 1763 (http://www.stat.ucla.edu/history/essay.pdf).

6. Teorema di Bayes e informazione diagnostica

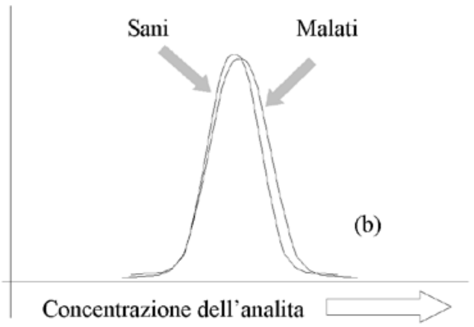
“...questi problemi sono classificati come probabilità delle cause e sono i più importanti di tutti per le loro applicazioni scientifiche ... Un effetto potrebbe essere prodotto dalla causa a o dalla causa b. L'effetto è appena stato osservato. Ci domandiamo la probabilità che sia dovuto alla causa a. Questa è una probabilità di causa a posteriori. Ma non la potrei calcolare, se una convenzione più o meno giustificata non mi dicesse in anticipo quale è la probabilità a priori che la causa a entri in gioco.“
(Henry Poincaré)

Consideriamo ora un test di laboratorio per una patologia che comporta un innalzamento della concentrazione di uno specifico analita.

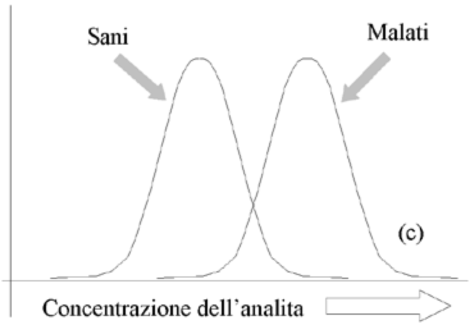
Il test ideale dovrebbe consentire di discriminare completamente tra sani e malati, come nell’esempio della figura (a). In un caso del genere, una volta eseguito il test, in base al suo risultato possiamo attribuire il paziente al gruppo dei sani o al gruppo dei malti con assoluta certezza.



La situazione diametralmente opposta è quella indicata nella figura (b). In questo caso i risultati del test nel gruppo dei sani e nel gruppo dei malati sono identici: una volta eseguito il test, in base al suo risultato non sappiamo se attribuire il paziente al gruppo dei sani o al gruppo dei malati. Il test non può essere utilizzato per la diagnosi della malattia in questione. Si tratta di un test inutile.



Nel caso di un test reale ci troviamo quasi sempre nella situazione indicata nella figura (c). La distribuzione dei risultati del test nel soggetti sani e nei soggetti malati è parzialmente sovrapposta, determinando così un certo grado di incertezza nella classificazione. Ci troviamo, come abbiamo già detto, di fronte al paradosso del sorite, attribuito a Zenone: quale è il granello che fa passare un mucchio di sabbia in un non-mucchio? La cui forma moderna è: quale è il risultato di una specifica analisi di laboratorio che fa passare il soggetto dallo stato di “malato” a quello di “sano” (o viceversa)?



Per incominciare a lavorare su questo problema possiamo innanzitutto costruirci una tabella nella quale riassumere i risultati di un test quantitativo, considerando una classificazione dicotomica tra soggetto malato (indicato con M+) e soggetto sano (indicato con M-) e tra risultato positivo del test (indicato con T+) e risultato negativo del test (indicato con T-).

	T+	T-
M+	T+M+	T-M+
M-	T+M-	T-M-

L'obiezione che in una classificazione dicotomica si perde il senso del risultato numerico viene superata dal fatto che il valore soglia tra T+ e T- può essere variato in continuo (lo vedremo successivamente nella parte riservata a teorema di Bayes e strategie diagnostiche).

Definiamo come *sensibilità* [1] la positività del test nei malati (una sensibilità del 100% significa che il test è positivo nel 100% dei malati, una sensibilità del 90% significa che il test è positivo nel 90% dei malati, e così via).

Definiamo come *specificità* [1] la negatività del test nei sani (una specificità del 100% significa che il test è negativo nel 100% dei sani, una specificità del 90% significa che il test è negativo nel 90% dei sani, e così via).

Definiamo infine come *prevalenza* della malattia il numero dei soggetti malati presenti, in un dato istante, nella popolazione (una prevalenza del 5 per mille significa che il 5 per mille delle persone è affetto dalla malattia, e così via).

Esprimendo queste grandezze in termini di probabilità abbiamo le definizioni

	T+	T-	
M+	*P(T+M+)* [sensibilità]	*P(T-M+)* [1 – sensibilità]	(6.1)
M-	*P(T+M-)* [1 – specificità]	*P(T-M-)* [specificità]	

Alle due principali grandezze, la sensibilità e la specificità, ne va aggiunta un terza, la prevalenza P(M+), cioè il numero di soggetti che hanno la specifica malattia presenti, in un dato istante, nella popolazione (in base alla (5.7) sarà P(M-) = 1 – P(M+) il numero di soggetti che <u>non</u> hanno la specifica malattia presenti, in un dato istante, nella popolazione).

Nel caso di due situazioni mutuamente esclusive (<u>affetto</u> o <u>non affetto</u> dalla malattia A) il teorema di Bayes (5.13) può essere espresso anche nella forma

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B|A) \cdot P(A) + P(B|nonA) \cdot P(nonA)} \qquad (6.2)$$

che, sostituendo *A* con *M+* e sostituendo *B* con *T+*, può essere riscritta come

$$P(M+|T+) = \frac{P(T+|M+) \cdot P(M+)}{P(T+|M+) \cdot P(M+) + P(T+|M-) \cdot P(M-)} \qquad (6.3)$$

e letta come

$$P(M+|T+) = \frac{\textit{sensibilità} \cdot \textit{prevalenza}}{\textit{sensibilità} \cdot \textit{prevalenza} + (1 - \textit{specificità}) \cdot (1 - \textit{prevalenza})} \qquad (6.4)$$

La probabilità che un test sia positivo data la malattia, *P(T+|M+)* che si trova sulla destra dell’espressione (6.3) deve essere letta come la <u>probabilità dell’effetto</u> (risultato positivo del test) <u>data la causa</u> (la malattia).

La probabilità che la malattia sia presente in un soggetto con un test positivo, *P(M+|T+)*, che si trova sulla sinistra dell’espressione (6.3) e della (6.4), deve essere letta come la <u>probabilità della causa</u> (la malattia) <u>dato l’effetto</u> (risultato positivo del test), ed è definita come il <u>valore predittivo del test positivo</u> [2].

La *patologia medica* insegna come si comportano <u>i segni data la malattia</u> (l’effetto data la causa). Ci insegna ad esempio che nell'epatite virale di tipo A è presente un aumento moderato delle transaminasi.

La *clinica medica*, insegna a diagnosticare <u>la malattia dati i segni</u> (la causa dato l’effetto). Un soggetto con aumento moderato delle transaminasi, che probabilità ha di essere affetto da una epatite A?

Il *teorema di Bayes* consente, conoscendo la prevalenza di una malattia, e la sensibilità e la specificità di un test per la sua diagnosi, di calcolare la probabilità di malattia in caso di test positivo (o la probabilità di assenza della malattia in caso di test negativo). Consente, in altre parole, il passaggio dalla patologia medica alla clinica medica, e fornisce le basi della razionalità della diagnostica di laboratorio e, a un livello superiore, della decisione medica.

Reciprocamente in questa forma del teorema di Bayes

$$P(M-|T-) = \frac{P(T-|M-) \cdot P(M-)}{P(T-|M-) \cdot P(M-) + P(T-|M+) \cdot P(M+)} \qquad (6.5)$$

che può essere letta come

$$P(M-|T-) = \frac{\textit{specificità} \cdot (1 - \textit{prevalenza})}{\textit{specificità} \cdot (1 - \textit{prevalenza}) + (1 - \textit{sensibilità}) \cdot \textit{prevalenza}} \qquad (6.6)$$

la probabilità che la malattia sia assente la malattia in un soggetto con un test negativo, *P(M-|T-)*, che si trova sulla sinistra dell’espressione, è definita come il <u>valore predittivo del test negativo</u> [2].

Rimandando ai riferimenti [3], [4] e [5] per un più completo inquadramento del teorema di Bayes applicato alla diagnostica di laboratorio, vediamo ora alcuni esempi che ci consentano di capire meglio l’enorme portata del teorema in campo diagnostico e medico.

Un esempio di come il teorema di Bayes in realtà possa essere applicato in modo intuitivo, è il seguente.

PROBLEMA 7
Si consideri un test destinato a rivelare la presenza nel siero di anticorpi anti-HIV. Si assuma che questo test abbia una sensibilità del 100% (il test, quindi, è positivo nel 100% dei malati). Si assuma che questo test abbia una specificità del 99.7% (il test, quindi, è negativo nel 99.7% dei soggetti sani). Si sa che la prevalenza dell'infezione da HIV è del 3 per mille (nella popolazione, su 1000 soggetti presi a caso, 3 sono infetti da virus).
Quale è il valore predittivo del test positivo?

Supponendo di effettuare il test su 1000 soggetti presi a caso i risultati saranno i seguenti: 3 soggetti presenteranno positività agli anticorpi anti-HIV (veri positivi), in quanto la prevalenza è del 3 per mille e la sensibilità del test del 100% ci assicura di individuare tutti i soggetti infetti; inoltre effettuando l'analisi su 1000 soggetti, a causa del fatto che la specificità del test è del 99.7% ci dovremo aspettare 3 positivi su 1000 soggetti sani (falsi positivi). Essendo in totale 6 i soggetti positivi, e 3 i veri positivi, avremo quindi un valore predittivo del test positivo pari a 3/6, cioè un valore predittivo del test positivo pari al 50% (*p* pari a 0,5).

Applicando la (6.4) avremo il calcolo espresso in dettaglio mediante il teorema di Bayes:

$$P(M+|T+) = \frac{\textit{sensibilità} \cdot \textit{prevalenza}}{\textit{sensibilità} \cdot \textit{prevalenza} + (1 - \textit{specificità}) \cdot (1 - \textit{prevalenza})} \quad \text{ovvero}$$

$$P(M+|T+) = \frac{1 \cdot 0{,}003}{(1 \cdot 0{,}003) + (0{,}003 \cdot 0{,}997)} = 0{,}501^{[66]} = \textit{valore predittivo del test positivo}$$

L'esempio riportato si presta a una prima importante considerazione.

Il teorema di Bayes rappresenta l'unico strumento che consente di fornire una misura quantitativa, e quindi oggettiva, del valore aggiunto fornito da un test diagnostico (per esempio un'analisi di laboratorio)[67].

Nel caso degli anticorpi anti-HIV, la differenza tra la probabilità di essere malati a posteriori (dopo avere effettuato il test, pari al 50%) e la probabilità di essere malati a priori (prima di avere effettuato il test, pari al 3 per mille) rappresenta appunto il valore aggiunto che il test è in grado di fornire, in termini di informazione, alla diagnosi clinica.

La seconda considerazione parte dal fatto che la determinazione degli anticorpi anti-HIV, un test di primo livello poco costoso, consente di restringere da 1000 a 6 soli individui la rosa dei candidati ad essere sottoposti a un test di secondo livello (western-blot). Che è sì risolutivo dal punto di vista diagnostico, ma che ha un costo molto superiore a quello degli anticorpi. Eseguire la western-blot a 6 individui su mille è ragionevole, mentre eseguirla a 1000 su 1000 sarebbe uno spreco di risorse.

---

[66] Il valore esatto è 0.500751126690035.
[67] Come diceva Lord Kelvin (1824-1907): "*Quando puoi misurare ciò di cui stai parlando, ed esprimerlo in numeri, puoi affermare di saperne qualcosa; se però non puoi misurarlo, se non puoi esprimerlo con numeri, la tua conoscenza sarà povera cosa e insoddisfacente: forse un inizio di conoscenza, ma non abbastanza da far progredire il tuo pensiero fino allo stadio di scienza, qualsiasi possa essere l'argomento.*"

L'analisi bayesiana è la sola che possa fornire gli indicatori necessari per effettuare una valutazione oggettiva del rapporto costi/benefici di una strategia diagnostica, e quindi di valutarne la razionalità nei confronti di strategie diagnostiche alternative.

Dal rapporto tra costi (in termini sia economici sia di diagnosi non corrette) e benefici (sia in termini economici sia di diagnosi corrette) dei diversi percorsi diagnostici si può arrivare al consenso degli operatori su una strategia diagnostica (anche se la praticabilità in sé della strategia diagnostica continuerà ovviamente a dipendere dal contesto economico e culturale, quindi sostanzialmente dai vincoli economici e dai vincoli etici che alla strategia sono attribuiti).
La terza considerazione parte dall'osservazione che un test con sensibilità del 100% e specificità del 100% è un test che ha per definizione un valore predittivo del test positivo del 100%. Nel caso esemplificato degli anticorpi anti-HIV le caratteristiche del test erano pressoché ideali (sensibilità del 100% e specificità del 99,7%). Ma nonostante questo il valore predittivo del test positivo risultava del 50% "solamente".

In condizioni di bassa prevalenza riduzioni anche minime della specificità di un test possano comportare drastiche riduzioni del valore predittivo del test positivo.

Questo non avviene per il valore predittivo del test negativo, che risulterà uguale a 994/994 (il test è negativo in 994 sani su 994) cioè uguale al 100% (o se si preferisce uguale a 1): un test negativo consente nel caso specifico di escludere la malattia.

Applicando la (6.6) avremo il calcolo espresso in dettaglio mediante il teorema di Bayes:

$$P(M\text{-}|T\text{-}) = \frac{specificità \cdot (1 - prevalenza)}{specificità \cdot (1 - prevalenza) + (1 - sensibilità) \cdot prevalenza} \quad \text{ovvero}$$

$$P(M\text{-}|T\text{-}) = \frac{0{,}997 \cdot 0{,}997}{(0{,}997 \cdot 0{,}997) + (0 \cdot 0{,}003)} = 1{,}00 = \text{valore predittivo del test negativo}$$

In condizioni di bassa prevalenza associate a bassa specificità di un test, aumenta il valore predittivo del test negativo.

Vediamo ora due altri esempi di applicazione del teorema di Bayes.

PROBLEMA 8
La probabilità che una persona con più di cinquant'anni senza sintomi abbia un cancro del colorettale è dello 0,3%. Se una persona ha un cancro colorettale, c'è una probabilità del 50% che abbia il test del sangue occulto nelle feci positivo; se non ha un cancro colorettale, c'è una probabilità del 3% che abbia comunque il test del sangue occulto nelle feci positivo. Immaginate una persona sopra i cinquant'anni, asintomatica, sottoposta a screening e con il test del sangue occulto nelle feci positivo.
Quale è la probabilità che abbia veramente un cancro colorettale?

Analizziamo innanzitutto le tre affermazioni:
- "*la probabilità che una persona con più di cinquant'anni senza sintomi abbia un cancro del colorettale è dello 0,3%*", significa che la prevalenza della malattia è del 3 per mille.
- "*se una persona ha un cancro colorettale, c'è una probabilità del 50% che abbia il test del sangue*

occulto nelle feci positivo", significa che il test ha un sensibilità del 50%;
- "*se non ha un cancro colorettale, c'è una probabilità del 3% che abbia comunque il test del sangue occulto nelle feci positivo*", significa che il test ha una specificità del 97%;

Per calcolare il valore predittivo del test positivo applichiamo la (6.4)

$$P(M+|T+) = \frac{0{,}50 \cdot 0{,}003}{(0{,}50 \cdot 0{,}003) + (0{,}03 \cdot 0{,}997)} = 0{,}048 = \text{valore predittivo del test positivo}$$

e per calcolare il valore predittivo del test negativo applichiamo la (6.6)

$$P(M-|T-) = \frac{0{,}97 \cdot 0{,}997}{(0{,}97 \cdot 0{,}997) + (0{,}5 \cdot 0{,}003)} = 0{,}998 = \text{valore predittivo del test negativo}$$

In altre parole:
- un risultato positivo del test comporta una probabilità di essere malato del 4,8%, quindi su 1000 persone con il test positivo saranno 48 i veri positivi e 952 i falsi positivi;
- un risultato negativo del test comporta una probabilità di non essere malato del 99,8%, quindi su 1000 persone con il test negativo saranno 998 i veri negativi e 2 i falsi negativi.

PROBLEMA 9
Un marcatore tumorale ha le seguenti caratteristiche: (i) è positivo in 95 su 100 pazienti con il cancro; (ii) è negativo in 95 su 100 pazienti senza in cancro; (iii) in media, 5 persone su una popolazione di 1000 hanno un cancro non ancora diagnosticato del tipo che il marcatore tumorale in questione rileva.
Se il test è prescritto a un paziente selezionato casualmente in questa popolazione e l'esito è positivo, quale è la probabilità che il paziente abbia realmente il cancro?

Analizziamo innanzitutto le tre affermazioni:
- "*(i) è positivo in 95 su 100 pazienti con il cancro*", significa che il test ha un sensibilità del 95%;
- "*(ii) è negativo in 95 su 100 pazienti senza in cancro*", significa che il test ha una specificità del 95%;
- "*(iii) in media, 5 persone su una popolazione di 1000 hanno un cancro non ancora diagnosticato del tipo che il marcatore tumorale in questione rileva*", significa che la prevalenza della malattia è del 5 per mille.

Per calcolare il valore predittivo del test positivo applichiamo la (6.4)

$$P(M+|T+) = \frac{0{,}95 \cdot 0{,}005}{(0{,}95 \cdot 0{,}005) + (0{,}05 \cdot 0{,}995)} = 0{,}087 = \text{valore predittivo del test positivo}$$

e per calcolare il valore predittivo del test negativo applichiamo la (6.6)

$$P(M-|T-) = \frac{0{,}95 \cdot 0{,}995}{(0{,}95 \cdot 0{,}995) + (0{,}05 \cdot 0{,}005)} = 0{,}9997 = \text{valore predittivo del test negativo}$$

In altre parole:
- un risultato positivo del test comporta una probabilità di essere malato del 8,7%, quindi su 1000 persone con il test positivo saranno 87 i veri positivi e 914 i falsi positivi;
- un risultato negativo del test comporta una probabilità di non essere malato del 99,97%, quindi su 1000 persone con il test negativo saranno 1000 i veri negativi e 0 i falsi negativi.

L'analisi del problema 8 e del problema 9 conferma il fatto che il valore predittivo del test positivo è fortemente penalizzato da condizioni di bassa prevalenza della malattia e/o diminuzione della specificità. Questi argomenti verranno discussi in maggior dettaglio nel capitolo che segue.

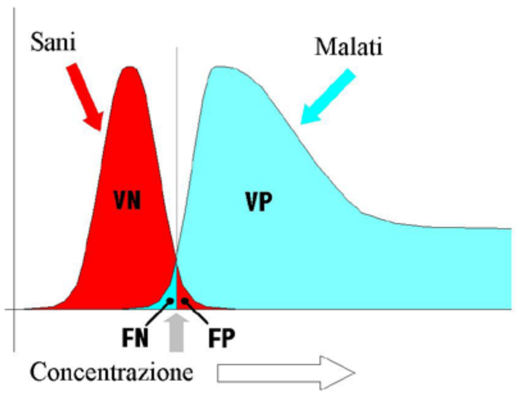
Bibliografia e riferimenti

[1] Altman DG, Bland JM. Statistics Notes: Diagnostic tests 1: sensitivity and specificità. *BMJ* 1994;308;1552 (http://www.bmj.com/content/308/6943/1552.full).
[2] Altman DG, Bland JM. Statistics Notes: Diagnostic tests 2: predictive values. *BMJ* 1994;309;102 (http://www.bmj.com//content/309/6947/102.1?variant=full-text).
[3] Galen RS, Gambino RS. Oltre il concetto di normalità: il valore predittivo e l'efficienza delle diagnosi mediche. Piccin Editore, 1980.
[4] Gerhardt W, Keller H. Evaluation of test data from clinical studies. Scand J Clin Lab Invest, 46, 1986 (supplement 181).
[5] Besozzi M. Ministat: un primer di statistica (http://www.bayes.it/html/download.html).

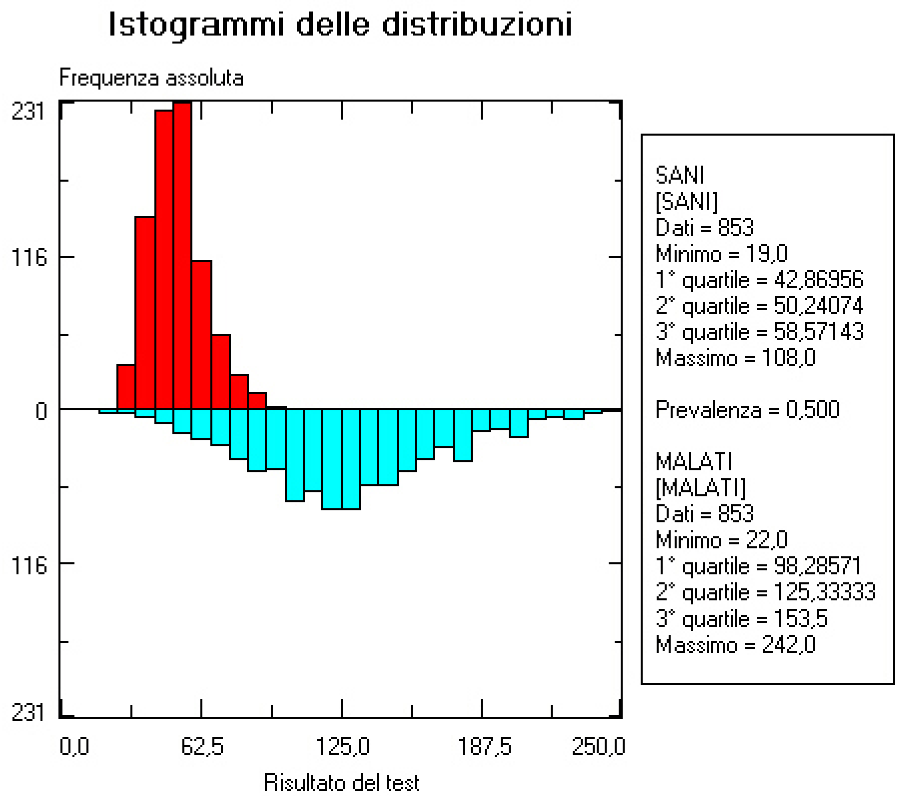
7. Teorema di Bayes e strategie diagnostiche

"Al mondo di sicuro ci sono solo la morte e le tasse."
(Benjamin Franklin)

Consideriamo ora una situazione nella quale la concentrazione dell'analita aumenti nella malattia, e nella quale la soglia tra sani e malati è fissata in un punto intermedio e tale per cui si abbiano il massimo possibile di veri negativi (VN = soggetti sani classificati come tali dal test) e il massimo possibile di veri positivi (VP = soggetti malati classificati come tali dal test), ma anche una quota di falsi positivi (FP = soggetti sani classificati come malati dal test) e una quota di falsi negativi (FN = soggetti malati classificati come sani dal test).



Per una completa analisi bayesiana dei dati, che elimina il fastidio di dovere effettuare i calcoli semplici ma lunghi che essa comporta, è possibile utilizzare software statistici biomedici, come Ministat [1], cui si riferiscono le immagini del seguente esempio simulato. Si suppongo di avere a disposizione i dati della determinazione della concentrazione nel siero di un dato analita, relativi a 853 soggetti sani e a 853 soggetti affetti da una malattia che provoca un patologia in aumento, un patologia cioè caratterizzata da un aumento della concentrazione dell'analita (nel caso di una patologia in diminuzione, cioè caratterizzata da una diminuzione della concentrazione dell'analita, bisogna semplicemente immaginare una situazione speculare a quella descritta). Le distribuzioni dei valori nei soggetti sani e nei soggetti malati sono poste a confronto nella figura seguente.



Nella tabella che segue è riportata l'analisi della casistica (VN = veri negativi, FP = falsi positivi, VP = veri positivi, FN = falsi negativi) per ciascuna delle classi in cui i dati sono stati suddivisi

			Sani		Malati	
Classe	da...	a...	VN	FP	VP	FN
1	0,0	8,33333	0	853	853	0
2	8,33333	16,66667	0	853	853	0
3	16,66667	25,0	2	851	850	3
4	25,0	33,33333	37	816	847	6
5	33,33333	41,66667	183	670	841	12
6	41,66667	50,0	408	445	831	22
7	50,0	58,33333	639	214	813	40
8	58,33333	66,66667	753	100	790	63
9	66,66667	75,0	809	44	763	90
10	75,0	83,33333	836	17	725	128
11	83,33333	91,66667	849	4	678	175
12	91,66667	100,0	852	1	632	221
13	100,0	108,33333	853	0	563	290
14	108,33333	116,66667	853	0	501	352
15	116,66667	125,0	853	0	425	428
16	125,0	133,33333	853	0	349	504
17	133,33333	141,66667	853	0	291	562
18	141,66667	150,0	853	0	233	620
19	150,0	158,33333	853	0	186	667
20	158,33333	166,66667	853	0	148	705
21	166,66667	175,0	853	0	120	733
22	175,0	183,33333	853	0	81	772
23	183,33333	191,66667	853	0	64	789
24	191,66667	200,0	853	0	48	805
25	200,0	208,33333	853	0	27	826
26	208,33333	216,66667	853	0	19	834
27	216,66667	225,0	853	0	13	840
28	225,0	233,33333	853	0	6	847
29	233,33333	241,66667	853	0	2	851
30	241,66667	250,0	853	0	0	853

Nella tabella che segue, viene riportata l'efficienza diagnostica del test (ultima colonna) in funzione dei livelli di soglia che è possibile fissare per discriminare tra sani e malati (prima colonna). Come si vede alla decima riga, la massima efficienza (0,84291) viene raggiunta se si considerano malati i soggetti con concentrazione uguale o superiore a 75, e sani i soggetti con concentrazione inferiore a 75. Si fa notare come sensibilità e specificità variano in modo opposto: fissando il valore soglia laddove si ha il massimo di sensibilità (1,0), il test risulterebbe totalmente aspecifico (darebbe cioè un risultato positivo anche a tutti i soggetti sani), mentre all'aumentare della specificità del test la sua sensibilità (la capacità cioè di identificare i malati) diminuisce.

Positivo se...	Sensibilità	Specificità	Efficienza
> 0,0	1,00000	0,00000	0,00000
> 8,33333	1,00000	0,00000	0,00000
> 16,66667	1,00000	0,00000	0,00000
> 25,0	0,99648	0,00234	0,00000
> 33,33333	0,99297	0,04338	0,03634
> 41,66667	0,98593	0,21454	0,20047
> 50,0	0,97421	0,47831	0,45252
> 58,33333	0,95311	0,74912	0,70223
> 66,66667	0,92614	0,88277	0,80891
> 75,0	0,89449	0,94842	0,84291
> 83,33333	0,84994	0,98007	0,83001
> 91,66667	0,79484	0,99531	0,79015
> 100,0	0,74091	0,99883	0,73974
> 108,33333	0,66002	1,00000	0,66002
> 116,66667	0,58734	1,00000	0,58734
> 125,0	0,49824	1,00000	0,49824
> 133,33333	0,40914	1,00000	0,40914
> 141,66667	0,34115	1,00000	0,34115
> 150,0	0,27315	1,00000	0,27315
> 158,33333	0,21805	1,00000	0,21805
> 166,66667	0,17351	1,00000	0,17351
> 175,0	0,14068	1,00000	0,14068
> 183,33333	0,09496	1,00000	0,09496
> 191,66667	0,07503	1,00000	0,07503
> 200,0	0,05627	1,00000	0,05627
> 208,33333	0,03165	1,00000	0,03165

> 216,66667	0,02227	1,00000	0,02227
> 225,0	0,01524	1,00000	0,01524
> 233,33333	0,00703	1,00000	0,00703
> 241,66667	0,00234	1,00000	0,00234
> 250,0	0,00000	1,00000	0,00000

Nella tabella che segue le caratteristiche del test il valore predittivo del test positivo e il valore predittivo del test negativo sono calcolati per ciascuno dei livelli soglia possibili..

			Valore predittivo del test	
Positivo se...	Sensibilità	Specificità	test positivo	test negativo
> 0,0	1,00000	0,00000	0,50000	###
> 8,33333	1,00000	0,00000	0,50000	###
> 16,66667	1,00000	0,00000	0,50000	###
> 25,0	0,99648	0,00234	0,49971	0,40000
> 33,33333	0,99297	0,04338	0,50932	0,86047
> 41,66667	0,98593	0,21454	0,55659	0,93846
> 50,0	0,97421	0,47831	0,65125	0,94884
> 58,33333	0,95311	0,74912	0,79163	0,94109
> 66,66667	0,92614	0,88277	0,88764	0,92279
> 75,0	0,89449	0,94842	0,94548	0,89989
> 83,33333	0,84994	0,98007	0,97709	0,86722
> 91,66667	0,79484	0,99531	0,99413	0,82910
> 100,0	0,74091	0,99883	0,99842	0,79404
> 108,33333	0,66002	1,00000	1,00000	0,74628
> 116,66667	0,58734	1,00000	1,00000	0,70788
> 125,0	0,49824	1,00000	1,00000	0,66589
> 133,33333	0,40914	1,00000	1,00000	0,62859
> 141,66667	0,34115	1,00000	1,00000	0,60283
> 150,0	0,27315	1,00000	1,00000	0,57909
> 158,33333	0,21805	1,00000	1,00000	0,56118
> 166,66667	0,17351	1,00000	1,00000	0,54750
> 175,0	0,14068	1,00000	1,00000	0,53783
> 183,33333	0,09496	1,00000	1,00000	0,52492
> 191,66667	0,07503	1,00000	1,00000	0,51949
> 200,0	0,05627	1,00000	1,00000	0,51448
> 208,33333	0,03165	1,00000	1,00000	0,50804
> 216,66667	0,02227	1,00000	1,00000	0,50563
> 225,0	0,01524	1,00000	1,00000	0,50384
> 233,33333	0,00703	1,00000	1,00000	0,50176
> 241,66667	0,00234	1,00000	1,00000	0,50059
> 250,0	0,00000	1,00000	###	0,50000

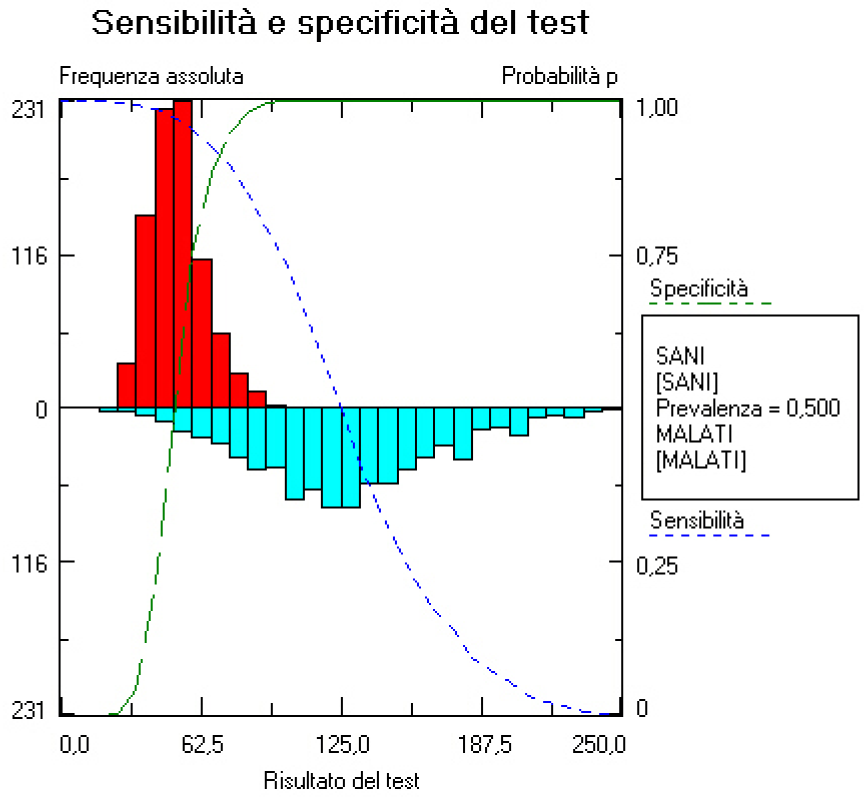
Quindi dato che, sulla base del criterio di massima efficienza del test, è stato scelto come livello di discriminazione tra sani e malati il valore di 75,0, è possibile fornire una misura quantitativa, e oggettiva, delle caratteristiche del test, che sono:

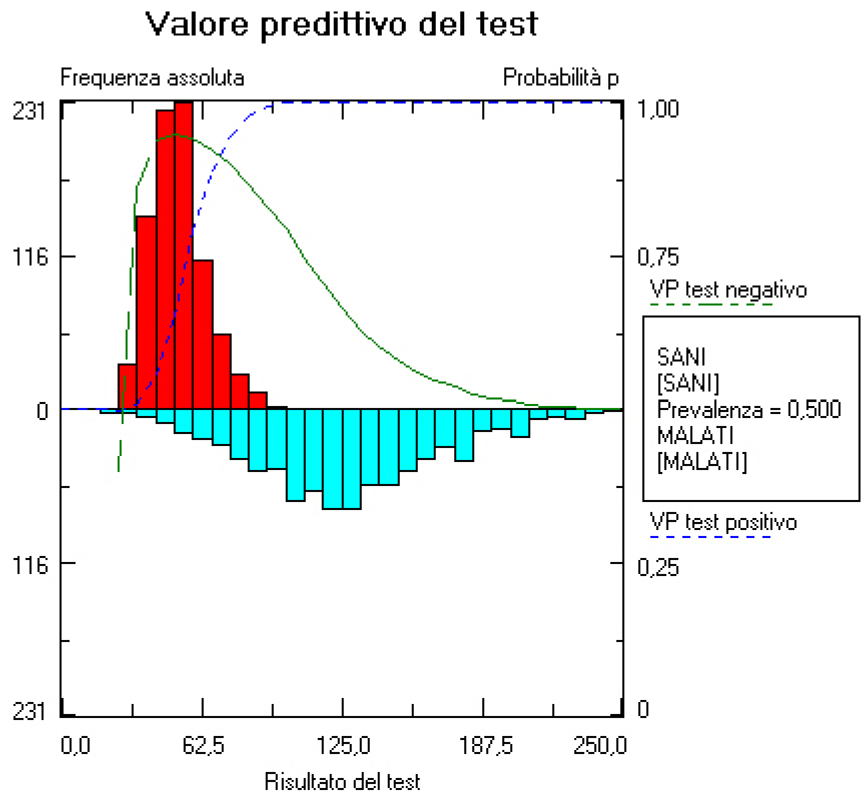
- sensibilità = 0,89449
- specificità = 0,94842
- valore predittivo del test positivo = 0,94548
- valore predittivo del test negativo = 0,89989

(chi preferisce i valori percentuali, deve solo moltiplicare tali valori per cento. I valori indicati con # non sono computabili per il numero di osservazioni effettuate).

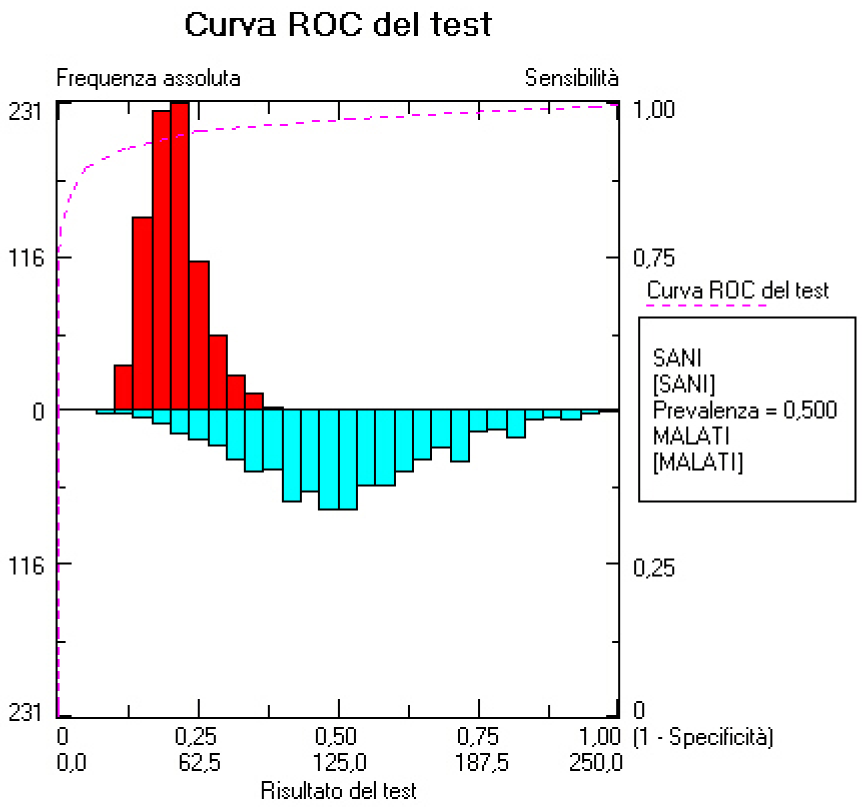
La rappresentazione grafica dell'andamento della sensibilità e della specificità in funzione del valore soglia prescelto è riportata nella prima delle due figure successive.

La rappresentazione grafica dell'andamento del valore predittivo del test positivo e del valore predittivo del test negativo in funzione del valore soglia prescelto è riportata nella seconda delle due figure successive.

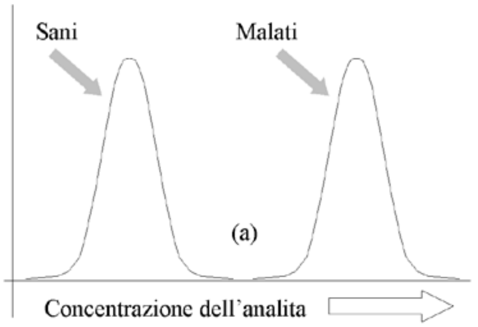


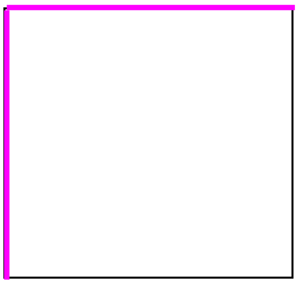


In fine riportando in un grafico in ordinate la sensibilità e in ascisse (1- specificità) si può ottenere la curva caratteristica (cirva ROC) del test

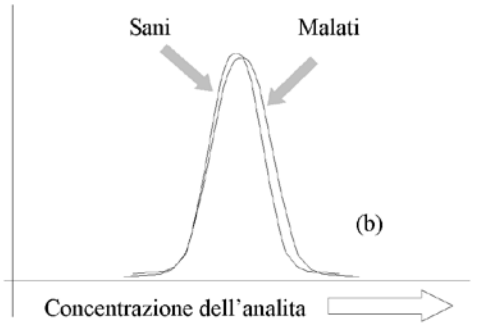


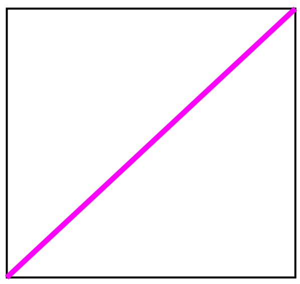
In pratica le curve ROC [1,2] riassumono in un unico indice le grandezze bayesiane illustrate analiticamente in precedenza, e forniscono una descrizione cumulativa del grado di sovrapposizione tra la distribuzione dei sani e dei malati, come indicato dalla corrispondenza tra questo e le curve ROC qui riportata.



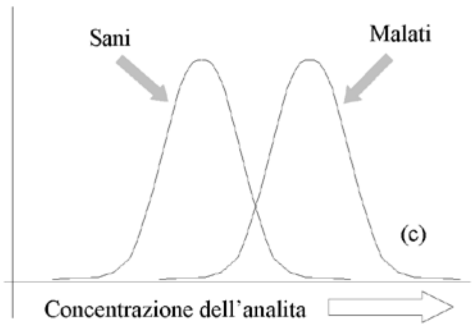


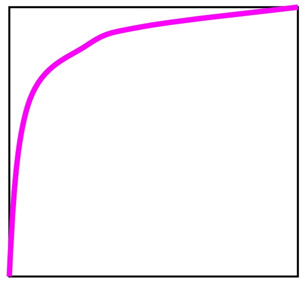
Curva ROC di un test ideale





Curva ROC di un test "inutile"





Curva ROC di un test reale

Fatte queste ulteriori premesse, possiamo ora esaminare quattro aspetti del teorema di Bayes che si rivelano della massima importanza per la sua comprensione:
- il collegamento che esiste tra l'espressione del teorema di Bayes in termini di probabilità e l'espressione in termini di numero dei casi osservati;
.- la risposta all'obiezione che in una classificazione dicotomica si perde il senso del risultato numerico;
- le strategie per la scelta della soglia che discrimina tra sani e malati;
- gli effetti della specificità e delle prevalenza della malattia sul valore predittivo del test.

Per illustrare il primo aspetto, ovvero il collegamento che esiste tra l'espressione del teorema di Bayes in termini di probabilità e l'espressione in termini di numero dei casi osservati, è necessario riprendere la tabella di classificazione (6.1)

	T+	T-
M+	*P(T+M+)* [sensibilità]	*P(T-M+)* [1 – sensibilità]
M-	*P(T+M-)* [1 – specificità]	*P(T-M-)* [specificità]

e ricordare che *P*(*M*+) = prevalenza [della malattia] e *P*(*M*-) = 1 – prevalenza.

Si passa dalle probabilità al numero di casi osservati moltiplicando le grandezza della prima riga per P(M+) e le grandezza della seconda riga per P(M-), ottenendo la seguente tabella

	T+	T-	
M+	*P(T+M+) · P(M+)* [sensibilità · prevalenza] [VP]	*P(T-M+)· P(M+)* [(1 – sensibilità) · prevalenza] [FN]	(7.1)
M-	*P(T+M-)· P(M-)* [(1 – specificità) · (1 – prevalenza)] [FP]	*P(T-M-)· P(M-)* [(specificità) · (1 – prevalenza)] [VN]	

che mette in evidenza le corrispondenze che ci permettono di ricalcolare tutte le grandezze in termini di numero dei casi osservati, che sono riassunte nella tabella (7.2) che segue:

Grandezza	*Calcolata come…*
sensibilità (positività del test nei malati)	VP / (VP+FN)
specificità (negatività del test nei sani)	VN / (VN+FP)
prevalenza (numero dei malati)	(VP+FN) / (VP+FN+FP+VN)
valore predittivo del test positivo (probabilità di essere malato per un soggetto con il test positivo)	(VP / (VP+FP)
valore predittivo del test negativo (probabilità di essere sano per un soggetto con il test negativo)	VN / (VN+FN)

(7.2)

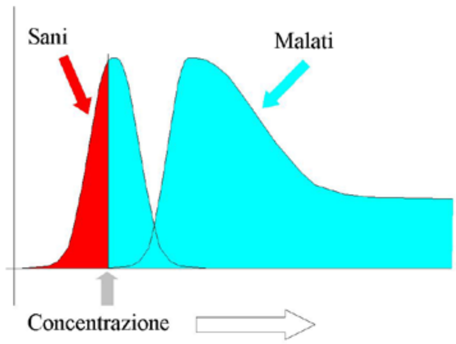
Il seguente esempio tratto da Galen e Gambino [3] si riferisce ai risultati della determinazione dell'alfa-fetoproteina nel cancro del fegato (malati, M+) e in altri disordini (sani, M-)

	T+	T-
M+	*90*	*17*
M-	*39*	*2079*

dai quali è possibile calcolare mediante le definizioni fornite nella (7.2):
- sensibilità = VP / (VP+FN) = 90 / (90+17) = 0,841
- specificità = VN / (VN+FP) = 2079 / (2079+39) = 0,982
- prevalenza = (VP+FN) / (VP+FN+FP+VN) = (90+17) / (90+17+39+2079) = 0,048
- valore predittivo del test positivo = VP / (VP+FP) = 90 / (90+39) = 0,698
- valore predittivo del test negativo = VN / (VN+FN) = 2079 / (2079+17) = 0,992

L'obiezione che in una classificazione dicotomica si perde il senso del risultato numerico assume più o meno questa forma: "Perfetto, ma avendo fissato un valore soglia di 40 U/L per la ALT, per la diagnosi di epatite sapere che, a parità di valore predittivo del test positivo, il mio paziente ha 80 U/L o 800 U/L di ALT è molto diverso!" Questa obiezione è valida solo apparentemente: infatti è possibile in ogni momento cambiare la soglia scegliendo quella più opportuna in relazione all'obiettivo clinico.

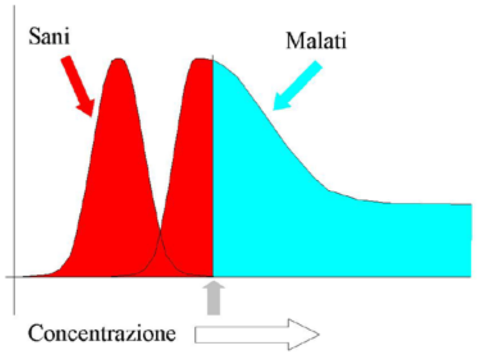
Le strategie per la scelta della soglia che discrimina tra sani e malati hanno una notevole rilevanza, e sono discusse ampiamente da Galen e Gambino [4].



La massima sensibilità (idealmente del 100%) è richiesta quando (i) la malattia è grave e non si vuole correre il rischio di non diagnosticarla, (ii) la malattia è curabile, e (iii) i falsi-positivi non comportano gravi danni psicologici e/o economici.

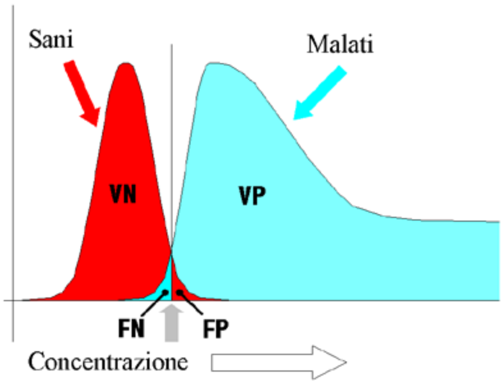
Un esempio è rappresentato dal feocromocitoma. La malattia può essere fatale se misconosciuta. Se diagnosticata è curabile al 100%. In caso di positività è possibile ripetere il test e/o confermarla con test di secondo livello. Ma lo stesso vale per l'epatite C, per la quale vale il principio per cui non si vuole correre il rischio di non diagnosticarla, la ricerca degli anticorpi anti-epatite C può dare dei falsi positivi, ma esistono test di secondo livello (ricerca del RNA del virus dell'epatite C nel sangue con tecniche di biologia molecolare).

La massima specificità (idealmente del 100%) è richiesta quando (i) la malattia è grave ma non è curabile, (ii) il sapere che la malattia è esclusa ha valore psicologico e/o per la salute pubblica, e (iii) i falsi-positivi possono comportare gravi danni psicologici e/o economici.



Un esempio è rappresentato dalla corea di Huntington. La malattia non è curabile. Se un caso non viene diagnosticato, l'evoluzione della malattia porterà il paziente di nuovo all'attenzione del medico. D'altra parte un falso positivo porterebbe a danni rilevanti al paziente.

La massima efficienza è richiesta quando (i) la malattia è grave ma curabile, inoltre i falsi-positivi e i falsi-negativi sono parimenti gravi e/o dannosi. Un esempio è rappresentato dall'infarto del miocardio. Se un caso viene trascurato, ne può derivare un danno. Un danno grave può derivare anche da una diagnosi di infarto in assenza di questo. La massima efficienza rappresenta la scelta più oculata.



Infine il massimo valore predittivo è essenziale quando il trattamento di un falso positivo potrebbe avere conseguenze oltremodo gravi. Un esempio è rappresentato dal cancro del polmone. Un test per questa malattia dovrebbe avere un valore predittivo (del test positivo) dal 100%, in quanto i solo trattamenti conosciuti sono la lobectomia e la radioterapia. Ma se tali terapie fossero eseguite su un paziente senza il cancro, le conseguenze sarebbero oltremodo gravi.

<u>Gli effetti della specificità e delle prevalenza della malattia sul valore predittivo del test</u> possono essere illustrati con il seguente esempio.

PROBLEMA 10
E' stato messo a punto un nuovo test per la diagnosi prenatale, su liquido amniotico, di una malattia genetica, per la quale è nota una prevalenza di 3 su mille neonati. Il test ha una sensibilità del 100%. Per la specificità del test, che è pari al 99,5%, sono riportati anche gli intervalli di confidenza al 95%, che rappresentano i limiti all'interno dei quali si colloca l'incertezza della stima: essi vanno da 99,0 a 99,9.
Quale è il grado di incertezza che può essere attribuito al valore predittivo del test positivo?

La risposta può essere ricavata dalla seguente tabella, nella quale il valore predittivo del test positivo è stato calcolato anche per i due valori di specificità (99,95 e 99,0%) che rappresentano i limiti superiore e inferiore all'interno dei quali si colloca l'incertezza con cui è stata ottenuta la misura della specificità.

Sensibilità	***Specificità***	***Prevalenza***	***Valore predittivo del test positivo***	***Valore predittivo del test negativo***
1,000	1,000	0,003	1,000	1,000
1,000	0,999	0,003	0,751	1,000
1,000	0,990	0,003	0,231	1,000
1,000	0,900	0,003	0,029	1,000

Come si vede se, a fronte di una specificità stimata del 99,5%, la specificità "vera" fosse del 99,9% il valore predittivo del test positivo sarebbe del 75,1%, mentre se la specificità "vera" fosse del 99,0% il valore predittivo del test positivo sarebbe del 23,1%. Ora dire che un test di diagnosi prenatale positivo ha un valore predittivo del 75% o che ha un valore predittivo del 23% sono due cose ben diverse.

PROBLEMA 11
E' stato messo a punto un nuovo test per la diagnosi dell'artrite reumatoide. Il test viene provato in via preliminare una clinica universitaria, presso la quale la metà circa dei pazienti è affetta da artrite reumatoide Per questo motivo vi è la rara opportunità di provarlo su ben 500 pazienti affetti da artrite reumatoide e su 500 pazienti affetti da altre patologie reumatologiche diverse dall'artrite reumatoide. Il test dimostra di avere una sensibilità del 100% e una specificità del 99%.
Il test viene immesso sul mercato includendo nel bugiardino i dati di sensibilità e di specificità, che sono in assoluto i migliori per un test per la diagnosi di artrite reumatoide. Dopo l'immissione in commercio, il test si diffonde rapidamente proprio per queste sue notevoli caratteristiche. Tuttavia l'utilizzo del test da parte dei medici di medicina generale dimostra risultati meno soddisfacenti di quanto atteso.
Esiste un razionale che spieghi perché nella pratica il test risulta meno soddisfacente di quanto prevedibile in base ai dati ottenuti nella fase preliminare?

La risposta può essere ricavata dalla seguente tabella, nella quale il valore predittivo del test positivo è stato calcolato a diversi valori di prevalenza della malattia.

Sensibilità	***Specificità***	***Prevalenza***	***Valore predittivo del test positivo***	***Valore predittivo del test negativo***
1,000	0,990	0,500	0,990	1,000
1,000	0,990	0,050	0,840	1,000
1,000	0,990	0,005	0,334	1,000
1,000	0,990	0,001	0,048	1,000

Consideriamo il valore predittivo del test positivo. Nel corso delle prove presso la clinica universitaria, il test è stato applicato ad una casistica di 500 malati e di 500 sani: quindi con una prevalenza della malattia del 50%. Il valore predittivo del test positivo corrispondente, quindi la probabilità di essere malato per un soggetto con un test positivo, era del 99%. Una volta entrato nell'uso, il test è stato poi applicato a casistiche non selezionate, con prevalenza della malattia minore. Dalla tabella appare evidente come l'applicazione del test con una prevalenza della malattia più "realistica", con valori della prevalenza che vanno dal 5 all'1 per mille, comporta un valore predittivo del test positivo che va dal 33,4% al 4,8%, o in altre parole comporta il fatto che su 100 soggetti con un test positivo la maggior parte (rispettivamente 67 e 95) sono soggetti sani. Quindi non un falso positivo su 100

Le possibili conseguenze determinate dalla dipendenza del valore predittivo del test dalla prevalenza della malattia, sono ben descritte in Gerhardt e Keller [5], e comportano la necessità nel corso della messa a punto e della applicazione di un test diagnostico di valutare tutti gli aspetti.

Bibliografia e riferimenti

[1] Besozzi M. Ministat: un primer di biostatistica (http://www.bayes.it/html/download.html).
[2] Altman DG, Bland JM. Statistics Notes: Diagnostic tests 3: receiver operating characteristic plots *BMJ* 1994;309;188 (http://www.bmj.com/content/309/6948/188.full).

[3] Galen RS, Gambino RS. Oltre il concetto di normalità: il valore predittivo e l'efficienza delle diagnosi mediche. Piccin Editore, 1980, p. 37.
[4] Galen RS, Gambino RS. Oltre il concetto di normalità: il valore predittivo e l'efficienza delle diagnosi mediche. Piccin Editore, 1980, pp. 49-80.
[5] Gerhardt W, Keller H. Evaluation of test data from clinical studies. Scand J Clin Lab Invest, 46, 1986 (supplement 181), pp. 39-42.

8. Teorema di Bayes e decisioni mediche

"There are three kinds of lies: lies, damned lies, and statistics."
(Mark Twain)

Per analisi il cui risultato può essere espresso in una scala numerica continua, tradizionalmente, la decisione medica in merito allo stato di salute di un paziente (sano o malato?) presuppone di cambiare l'orientamento diagnostico in corrispondenza di valori soglia prefissati, i *valori decisionali*.

Valori decisionali	***Esempio***
Intervalli di riferimento	concentrazione del glucosio compresa tra 60 e 100 mg/dL (milligrammi per decilitro di plasma)
Intervalli terapeutici dei farmaci	concentrazione della digossina nel siero compresa tra 1,0 e 2,0 μg/L (microgrammi per litro di siero)
Valori che definiscono la malattia	concentrazione dell'emoglobina inferiore a 13,0 g/dL (grammi per decilitro di sangue) in soggetti di sesso maschile con età superiore a 17 anni* = anemia
Valori desiderabili	concentrazione del colesterolo totale inferiore a 200 mg/dL (milligrammi per decilitro di siero)
Valori di allarme	concentrazione del potassio superiore 6,7 mmol/L (millimoli per litro di siero)
Differenze critiche	colesterolo ridotto del 12% dopo terapia

* definizione dell'Organizzazione Mondiale della Sanità

I criteri utilizzati per definire i vari tipi di valori decisionali sono semplici:

⇒ gli *intervalli di riferimento* rappresentano l'ambito di valori che include il 95% dei risultati osservati in un gruppo di controllo di soggetti sani. Sono quelli riportati nel referto di laboratorio. Come conseguenza della definizione, un soggetto sano che esegue una analisi di laboratorio ha il 95% di probabilità che il proprio valore risulti all'interno degli intervalli di riferimento, e un soggetto sano che esegue 20 analisi di laboratorio ha una probabilità di $(0,95)^{20} = 0,358486$ che tutti i risultati rientrino all'interno degli intervalli di riferimento. O, in altri termini, su 100 soggetti sani che eseguono ciascuno 20 analisi, solamente il 36% avrà tutti i valori all'interno degli intervalli di riferimento;

⇒ gli *intervalli terapeutici dei farmaci* rappresentano l'ambito di valori che caratterizza l'efficacia terapeutica. Concentrazioni inferiori non garantiscono l'efficacia. Concentrazioni superiori indicano che la dose somministrata è eccessiva;

⇒ i *valori che definiscono la malattia* forniscono, per l'appunto, la definizione della malattia. Un altro esempio oltre a quello dell'anemia: una glucosio a digiuno uguale o superiore a 126 mg/dL (milligrammi per decilitro di plasma) fornisce la diagnosi di diabete mellito. Da notare la differenza, oltre che numerica, concettuale rispetto agli intervalli di riferimento del glucosio, che vanno da 60 a 110 mg/dL;

⇒ i *valori desiderabili* sono valori raccomandati in relazione a specifici obiettivi. Nel caso del colesterolo totale nel siero, essendo nota la correlazione tra aumento del colesterolo e aumento del rischio di malattia cardiovascolare, il valore di 200 mg/dL è stato scelto sulla base di un consenso internazionale in merito al fatto che il mantenimento di una concentrazione del colesterolo inferiore a tale limite consente di ridurre in modo significativo l'incidenza di malattie cardiovascolari;

⇒ i *valori di allarme* sono valori che impongono un intervento. Nel caso di un potassio superiore a 6,7 mmol/L, quale si può riscontrare in caso di grave insufficienza renale, l'intervento si rende necessario a causa del rischio imminente di aritmie cardiache mortali;

⇒ le *differenze critiche* rappresentano la minima differenza tra due risultati consecutivi che può essere ritenuta statisticamente significativa [vedere appendice E]. Si supponga di sottoporre a trattamento dietetico un paziente con un colesterolo di 243 mg/dL, e che un successivo controllo a distanza fornisca un valore di 225 mg/dL. La terapia è stata efficace? Probabilmente no, in quanto per il colesterolo è noto che la differenza critica è del 12%, quindi il valore avrebbe dovuto scendere almeno a 214 mg/dL perché la differenza potesse essere attribuita al trattamento e non al caso.

I valori decisionali fissi sono degli stereotipi: ovviamente utili (senza di essi la diagnostica di laboratorio non si sarebbe sviluppata), ma che tendono a forzare l'interpretazione del dato di laboratorio, e a prescindere dal fatto che ogni individuo, e quindi ogni caso clinico, differisce sottilmente da tutti gli altri.
Per capire come si sia resa sempre più evidente la necessità di introdurre nella decisione medica un criterio di flessibilità, che consenta di adattare il risultato della misura (oggettivo) al caso clinico specifico, unico e irripetibile, si consideri l'esempio degli enzimi (transaminasi) per la diagnosi dell'epatite. La situazione apparsa ai primi "esploratori" della biochimica, può essere emblematicamente così descritta (risultati in U/L)

Controlli "sani"		***Pazienti***
12	31	80

Questa prima tabella vuole evidenziare il fatto che un tempo erano poche le analisi eseguite, e pochi i controlli sani utilizzati come confronto. Inoltre i pazienti erano selezionati in base alla presenza di segni clinici importanti, per esempio, nel caso dell'epatite, in base alla presenza di ittero (un danno del fegato elevato al punto tale da determinare una diminuzione della capacità di eliminare la bilirubina nella bile). Il che implica una compromissione più consistente del fegato. Nella distribuzione dei valori, esemplificati in questa prima tabella, appariva, ai primi "esploratori" della biochimica, una discontinuità, a livello della quale era "intuitivo" porre il limite per distinguere tra sani e malati.
Oggigiorno sono molte le analisi eseguite, molti i controlli sani di cui disponiamo, i pazienti solo raramente (tipicamente solo quando ricoverati in ospedale) sono selezionati in base alla presenza di segni clinici importanti, come la presenza di ittero. La spinta verso una diagnosi sempre più precoce comporta il fatto che oggigiorno la situazione delle transaminasi nei confronti della diagnosi di epatite può essere emblematicamente così descritta (risultati in U/L)

Controlli "sani"								***Pazienti***							
12	15	21	27	31	38	44	46	38	45	51	58	65	73	75	80

Data la sovrapposizione tra le distribuzioni dei valori nei sani e nei malati, il valore decisionale in questa seconda tabella risulta tutt'altro che evidente. E allora come stabilirlo?

Il concetto di "mucchio di sabbia" è intuitivo. Il problema sorge quando, dal mucchio, incominciamo a togliere i granelli di sabbia uno ad uno. Quando è che il mucchio non è più tale, e diventa un "non-mucchio"? Reciprocamente, se prendiamo valori di transaminasi intuitivamente "non-normali", e incominciamo a ridurli senza soluzioni di continuo, di una U/L per volta, quale è il valore che segna la transizione tra "non-normali" e "normali"?

L'approccio bayesiano consente di introdurre, tra il bianco e il nero, i livelli di grigio. Se *probabilità (p) di malattia = 0* significa "certamente sano", e *probabilità (p) di malattia = 1* significa "certamente malato", ciò avviene raramente nella pratica medica. Prevalgono i casi di $0 < p < 1$. Il confine tra salute e malattia è sfumato. L'approccio bayesiano, consentendo di definire il valore decisionale di volta in volta, in funzione di specifiche esigenze mediche, permette di formalizzare in modo scientificamente, metodologicamente e numericamente rigoroso quanto il "buon senso" medico aveva da tempo intuito [1].

Tuttavia una cosa ce la dobbiamo domandare: perché il teorema di Bayes non è riuscito ad "automatizzare" il buon senso medico, quindi ad "automatizzare" la diagnosi clinica?

Innanzitutto la forma presentata per un test diagnostico e una malattia dal punto di vista computazionale è banale. Ma bisogna ricordare che se teorema di Bayes è generalizzato a *n* ipotesi concomitanti (come nel caso di una diagnosi differenziale tra più malattie), nascono rapidamente seri problemi computazionali [sic!][68].

Limitiamoci quindi al caso più semplice, quello di un test diagnostico e una malattia. Concettualmente (e teoricamente) abbiamo una macchina matematica ineccepibile. Tanto che oggi il data mining bayesiano rappresenta uno degli strumenti più utilizzati: e filtri bayesiani stanno alla base dell'efficienza nella ricerca dell'informazione che chiunque è in grado di sperimentare utilizzando un motore di ricerca come Google[69]. Ma un sistema di supporto alle decisioni, quale è quello che pretendiamo di avere applicando ai nostri dati il teorema di Bayes, è influenzato sia dalla qualità della macchina che ne costituisce il motore inferenziale (una macchina peraltro ineccepibile), sia dalla qualità dell'informazione in ingresso.

Ecco dunque il problema: l'informazione. Perché i dati in ingresso nella macchina inferenziale sono per definizione caratterizzati da un certo grado di indeterminazione, che sicuramente è diverso da zero. Che nessuna macchina è in grado di eliminare. E che per definizione tende ad espandersi nell'arco del trattamento dell'informazione.

---

[68] L'inferenza bayesiana ha a lungo rappresentato una corrente minoritaria nella teoria della statistica. Ciò è il larga parte dovuto alle difficoltà algebriche che essa pone; la computazione delle probabilità a posteriori è basata sul calcolo di integrali, per i quali spesso non si hanno espressioni analitiche. Queste difficoltà hanno fino a pochi anni fa limitato la capacità della statistica bayesiana di produrre modelli realistici della realtà. Al fine di evitare di incorrere in problemi algebrici, gran parte dei risultati erano basati sulla teoria delle coniugate, particolari famiglie di distribuzioni per cui la probabilità a posteriori risulta avere la stessa forma di quella a priori. Chiaramente un approccio di questo tipo cozzava con l'ambizione dei bayesiani di fare statistica a partire da ipotesi meno restrittive di quelle dell'inferenza classica. Grazie alla maggiore disponibilità di risorse informatiche a partire dagli anni '90, è stato possibile superare tali difficoltà. È infatti possibile risolvere gli integrali in via numerica, aggirando i problemi algebrici, nella maggior parte delle applicazioni su un qualsiasi personal computer. Questa possibilità ha inoltre stimolato l'applicazione alla statistica Bayesiana di metodi numerici sviluppati in altri contesti, come quelli basati sulla simulazione (metodo Monte Carlo, algoritmi del *Gibbs sampler* e di Metropolis-Hastings), nonché lo sviluppo di metodi nuovi nell'ambito della statistica bayesiana stessa (ad esempio i popolari metodi basati sul Markov Chain Monte Carlo, o MCMC). Ciò ha notevolmente incrementato la popolarità dell'inferenza bayesiana tra gli statistici; sebbene i bayesiani costituiscano ancora una minoranza, si tratta di una minoranza in rapida crescita. Al di là delle difficoltà numeriche che hanno a lungo reso impopolare l'inferenza bayesiana o delle problematiche epistemologiche che i metodi bayesiani sollevano, l'approccio bayesiano ha il merito di aver stimolato, nella statistica come in altre discipline (un recente esempio è dato dall'economia), la riflessione su cosa sia un modello e su che lettura un ricercatore ne deve dare.

[69] Negli ultimi vent'anni la disponibilità di dati ed informazioni digitali è cresciuta vertiginosamente grazie alla progressiva ed inesorabile diffusione dei computer e al drastico abbassamento del rapporto prezzo/capacità dei supporti di memorizzazione unito ad un aumento della potenza di calcolo dei moderni processori. Infatti non è insolito che un sistema di supporto alle decisioni contenga milioni o anche centinaia di milioni di record di dati. I sistemi classici di memorizzazione dei dati, i DBMS (database management system), offrono un'ottima possibilità di memorizzare ed accedere ai dati con sicurezza, efficienza e velocità ma non permettono un'analisi per l'estrazione di informazione utili come supporto alle decisioni. Un approccio recente è il *data mining*. Il data mining grazie ad un approccio esplorativo evidenzia relazioni che non solo erano nascoste e sconosciute, ma che spesso non si era nemmeno mai ipotizzato potessero esistere. Storicamente, lo sviluppo dei metodi statistici ha prodotto un certo numero di tecniche di analisi dei dati utili nel caso in cui si debbano confermare delle ipotesi predefinite. Tali tecniche risultano però inadeguate nel processo di scoperta di nuove correlazioni e dipendenze tra i dati, che crescono in quantità, dimensione e complessità. Le reti bayesiane sono uno strumento flessibile ed adatto per estrarre informazioni dai dati. Sono modelli grafici di probabilità in cui i nodi rappresentano le variabili aleatorie e gli archi le dipendenze casuali tra le variabili, permettendo, quindi, di apprendere le relazioni causali.

Normalmente per i test di laboratorio troviamo definite una sensibilità e una specificità. Anche se praticamente mai è indicato il criterio in base al quale esse sono state calcolate, si può presumere che, essendo entrambe diverse da 1, vi sia stato un qualche compromesso tra le due, in genere basato sul concetto che abbiamo visto di efficienza. Ovviamente entrambe le due grandezze sono misurate, quindi caratterizzate da un certo grado di indeterminazione.

Ma per il calcolo del valore predittivo è necessaria anche la prevalenza della malattia. Anch'essa, come misura, è caratterizzata da un certo grado di indeterminazione. Ma esiste per la prevalenza un ulteriore problema. Essa varia moltissimo sia nel tempo (basti pensare alla fluttuazione della prevalenza dell'influenza e del morbillo nel corso dell'anno), sia in relazione al contesto (la prevalenza di sieropositivi per l'HIV è molto diversa in Africa e in Europa, in Italia è molto diversa tra la popolazione generale e la popolazione carceraria, eccetera). La fluttuazione nel tempo e la sensibilità al contesto della prevalenza creano ovviamente una enorme difficoltà per molte malattie (paradossalmente questo è un problema minore per le malattie genetiche come la sindrome di Down per la quale la prevalenza è stabile ed è nota anche in relazione alla classe di età cui appartiene la gravida).

Combinando le incertezze di sensibilità, specificità e prevalenza, arriviamo a conclusioni ancor più incerte. Aggiungiamo che ogni caso clinico è unico. Quindi vorremmo cambiare in modo dinamico la soglia che discrimina tra sani e malati in relazione allo specifico caso clinico, e quindi allo specifico obiettivo clinico: per fare questo è necessario disporre delle distribuzioni dei risultati del test nei sani e nei malati. Una dato che sostanzialmente non esiste (non esistono se non negli studi clinici controllati database contenenti sia i risultati delle analisi sia le diagnosi). Tutto questo rende per ora scarsamente praticabile affrontare in modo rigorosamente computazionale un approccio che non solo il laboratorio, ma anche i clinici hanno riconosciuto come connaturato con il "modo di pensare diagnostico" (vedere ad esempio il recente articolo "Why clinicians are natural bayesians").

Per ora sembra che, in attesa di sistemi di supporto alla diagnosi ancora al di là da venire, l'unico approccio praticabile sia rappresentato da un compromesso epocale: coniugare il rigore del teorema all'imprinting ancora legato al Pleistocene della razionalità umana. Vediamo come questo sia possibile.

Se partiamo dalla forma classica (5.13) del teorema di Bayes

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}$$

che possiamo anche riscrivere come

$$P(A|B) = \frac{P(B|A)}{P(B)} \cdot P(A) \qquad (8.1)$$

e se consideriamo il rapporto

$$\frac{P(B|A)}{P(B)}$$

globalmente come un *fattore* di moltiplicazione, la (8.1) si può semplificare riscrivendola come

$$P(A|B) = fattore \cdot P(A) \qquad (8.2)$$

che può essere letta come

$$P(della\ causa|dato\ l'effetto) = fattore \cdot P(della\ causa) \qquad (8.3)$$

ovvero come

$$P(a\ posteriori) = fattore \cdot P(a\ priori) \qquad (8.4)$$

Questa ultima espressione è particolarmente interessante in relazione all'uso, diffuso nel mondo anglosassone, di utilizzare in luogo della probabilità *P* gli *odds*, che sono la probabilità espressa dal bookmaker come rapporto tra la vincita e la posta giocata.

La relazione che intercorre tra *odds* e *probabilità* è

$$odds = P / (1 - P) \qquad (8.5)$$

e inversamente

$$probabilità = odds / (1 + odds) \qquad (8.6)$$

PROBLEMA 12
Trasformare in odds la probabilità dell'80% cioè $P = 0{,}8$[70]

PROBLEMA 13
Trasformare in probabilità gli odds 4:1[71]

La (8.1) può essere riformulata in termini di odds come

$$O(A|B) = \Lambda(A|B) \cdot O(A) \qquad (8.7)$$

La grandezza $\Lambda(A|B)$ si chiama rapporto di verosimiglianza[72] e, dall'inglese likelihood-ratio, viene indicato sovente come LR, e pertanto la (8.7) può essere letta in analogia con la (8.3) come

$$odds(della\ causa|dato\ l'effetto) = rapporto\ di\ verosimiglianza \cdot odds(della\ causa) \qquad (8.8)$$

ovvero in analogia con la (8.4) come

$$odds(a\ posteriori) = rapporto\ di\ verosimiglianza \cdot odds(a\ priori) \qquad (8.9)$$

[70] *odds* = 0,8 / (1 – 0,8) = 0,8 / (0,2) = 4 cioè = *4:1*

[71] probabilità = 4 / (1 + 4) = 0,8

[72] altrimenti detto *fattore di Bayes*

ove il rapporto di verosimiglianza è la misura della verosimiglianza che collega la probabilità a priori alla probabilità a posteriori [2].

Nel caso di un test di laboratorio, il rapporto di verosimiglianza è per definizione il rapporto tra la probabilità che si verifichi lo specifico risultato in un individuo che ha la malattia e la probabilità che si verifichi lo specifico risultato in una persona che non ha la malattia:
- un test con rapporto di verosimiglianza di 1,0 non fornisce alcuna informazione, e i sui risultati non influiscono sulla probabilità post-test di malattia (che rimane identica alla probabilità pre-test della malattia);
- un rapporto di verosimiglianza > 1,0 incrementa la probabilità post-test di malattia; più grande è il rapporto di verosimiglianza, maggiore è l'informazione fornita da un risultato positivo del test (maggiore è la probabilità di essere malato dopo avere eseguito il test);
- un rapporto di verosimiglianza < 1,0 diminuisce la probabilità post-test di malattia; più piccolo è il rapporto di verosimiglianza, maggiore è l'informazione fornita da un un risultato negativo del test (minore è la probabilità di essere malato dopo avere eseguito il test).

In base alla definizioni date di odds e di rapporto di verosimiglianza (semplicemente sostituendo ad A la malattia M e a B il test T), possiamo formulare dalla (8.7) il valore predittivo di un test positivo ovvero la probabilità di essere malato per un soggetto con un test positivo

$$O(M+|T+) = \Lambda(M+|T+) \cdot O(M+) \qquad (8.10)$$

nella quale il rapporto di verosimiglianza per un test positivo è espresso come

$$LR+ = \Lambda(M+|T+) = \frac{P(T+|M+)}{P(T+|M-)} \qquad (8.11)$$

PROBLEMA 14
Il rapporto di verosimiglianza LR+ (LR per un test positivo) è uguale al rapporto tra il risultato del test nei malati e nei sani.
Esprimere LR+ in termini di sensibilità e di specificità[73].

Possiamo formulare dalla (8.7) anche il valore predittivo di un test negativo ovvero la probabilità di essere malato per un soggetto con un test negativo

$$O(M+|T-) = \Lambda(M+|T-) \cdot O(M+) \qquad (8.12)$$

nella quale il rapporto di verosimiglianza per un negativo è espresso come

$$LR- = \Lambda(M+|T-) = \frac{P(T-|M+)}{P(T-|M-)} \qquad (8.13)$$

PROBLEMA 15
Il rapporto di verosimiglianza LR- (LR per un test negativo) è uguale al rapporto tra il risultato del test nei malati e nei sani.
Esprimere LR- in termini di sensibilità e di specificità[74].

[73] Dalle definizioni della tabella (6.1) si ricava immediatamente che LR+ = *sensibilità / (1 – specificità).*

Combinando la (8.10) con la (8.11) abbiamo che in definitiva in termini di odds il valore predittivo di un test positivo è

$$O(M+|T+) = \frac{P(T+|M+)}{P(T+|M-)} \cdot O(M+) \qquad (8.14)$$

e combinando la (8.12) con la (8.13) abbiamo che in termini di odds il valore predittivo di un test negativo è

$$O(M+|T-) = \frac{P(T-|M+)}{P(T-|M-)} \cdot O(M+) \qquad (8.15)$$

PROBLEMA 16
Un test del sangue occulto nelle feci per la diagnosi del cancro colorettale ha le seguenti caratteristiche: sensibilità = 50% (0,50), specificità = 97% (0,97). La prevalenza della malattia è uguale allo 0,3% (0,003).
Calcolare valore predittivo del test positivo utilizzando la (6.4) e confrontarlo con quello ottenuto applicando la (8.14). Calcolare valore predittivo del test negativo utilizzando la (6.6) e confrontarlo con quello ottenuto applicando la (8.15).

Vediamo cosa accade per il valore predittivo del test positivo. Applicando la (6.4) avremo il calcolo espresso in dettaglio mediante il teorema di Bayes:

$$P(M+|T+) = \frac{\text{sensibilità} \cdot \text{prevalenza}}{\text{sensibilità} \cdot \text{prevalenza} + (1 - \text{specificità}) \cdot (1 - \text{prevalenza})} \qquad \text{ovvero}$$

$$P(M+|T+) = \frac{0,50 \cdot 0,003}{(0,50 \cdot 0,003) + (0,03 \cdot 0,997)} = 0,048 = \text{valore predittivo del test positivo}$$

Applicando la (8.14) avremo:

$$O(M+) = P(M+) / (1 - (P(M+)) = 0,003 / (1 - 0,003) = 0,003009027$$

$$LR+ = \text{sensibilità} / (1 - \text{specificità}) = 0,50 / (1 - 0,97) = 16,6667$$

quindi

$$O(M+|T+) = 0,003009027 \cdot 16,6667 = 0,05015$$

e riconvertendo gli odds in probabilità avremo

$$P = \text{odds} / (1 + \text{odds}) = 0,05015 / (1 + 0,05015) = 0,048 = \text{valore predittivo del test positivo}$$

Quindi le due soluzioni sono, come atteso, numericamente identiche.

---

[74] Dalle definizioni della tabella (6.1) si ricava immediatamente che LR- = *(1 – sensibilità) / specificità.*

Vediamo cosa accade per il valore predittivo del test negativo. Applicando la (6.6) avremo il calcolo espresso in dettaglio mediate il teorema di Bayes:

$$P(M\text{-}|T\text{-}) = \frac{specificità \cdot (1 - prevalenza)}{specificità \cdot (1 - prevalenza) + (1 - sensibilità) \cdot prevalenza} \quad \text{ovvero}$$

$$P(M\text{-}|T\text{-}) = \frac{0{,}97 \cdot 0{,}997}{(0{,}97 \cdot 0{,}997) + (0{,}50 \cdot 0{,}003)} = 0{,}998 = valore\ predittivo\ del\ test\ negativo$$

Applicando la (8.15) avremo:

$$O(M+) = P(M+) / (1 - P(M+)) = 0{,}003 / (1 - 0{,}003) = 0{,}003009027$$

$$LR- = (1 - sensibilità) / specificità = (1 - 0{,}50) / 0{,}97 = 0{,}515464$$

quindi

$$O(M+|T-) = 0{,}003009027 \cdot 0{,}515464 = 0{,}001551$$

e riconvertendo gli odds in probabilità avremo

$$P = odds / (1 + odds) = 0{,}001551 / (1 + 0{,}001551) = 0{,}002 = valore\ predittivo\ del\ test\ negativo$$

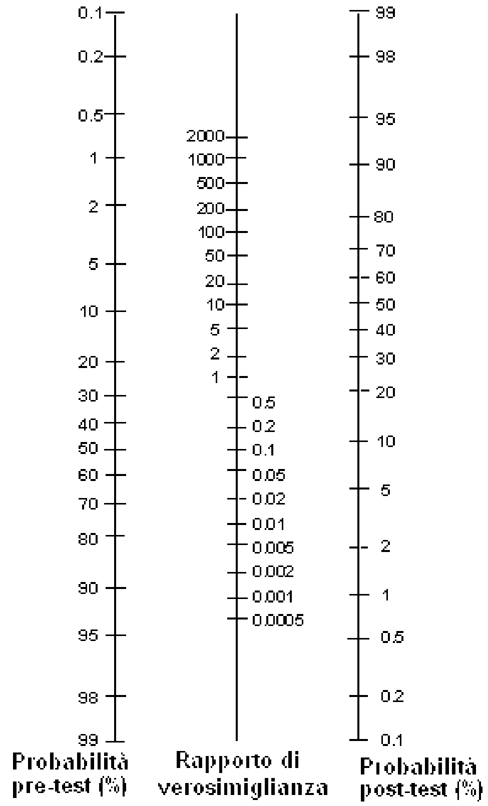
Quindi le due soluzioni sono, come atteso, identiche tranne che per il fatto che, come probabilmente sarà stato notato, con la (6.6) il valore predittivo del test negativo è la probabilità di essere sano per un soggetto con un test negativo, mentre con gli odds (8.15) il valore predittivo del test negativo sono gli odds di essere malato per un soggetto con un test negativo (ovvero, ritrasformando gli odds in probabilità, la probabilità di essere malato per un soggetto con un test negativo). Le differenze nell'espressione dei risultati del test sono riassunte nella seguente tabella:

Formulazione del teorema	***Valore predittivo del test positivo***	***Valore predittivo del test negativo***
In termini di probabilità	P(M+|T+) probabilità di essere malato per un soggetto con un test positivo	P(M-|T-) probabilità di essere sano per un soggetto con un test negativo
In termini di odds e likelihood ratio	O(M+|T+) odds di essere malato per un soggetto con un test positivo	O(M+|T-) odds di essere malato per un soggetto con un test negativo

Anche se l'approccio in termini di odds, con l'utilizzo del rapporto di verosimiglianza, é basato su una espressione della probabilità da bookmaker, e piuttosto lontana dalla mentalità di chi non è abituato alle scommesse, va notato che:
- gli odds sono immediatamente riconvertibili in probabilità, e questo consente di superare facilmente il problema (come vedremo appena avanti con il diagramma di Fagan);
- l'approccio ha il vantaggio di esprimere i risultati del test positivo e del test negativo sotto forma della stessa grandezza (odds di essere malato), semplificando almeno da questo punto di vista la comprensione delle conclusioni;
- gli odds forniscono un modo di utilizzo dell'approccio bayesiano alternativo allo sviluppo

analitico delle formule del teorema (uno sviluppo peraltro oggigiorno reso banale dalla possibilità di implementarne le formule anche sul tabellone elettronico di un dispositivo palmare).

Per chi opera in condizioni tali per cui è impensabile utilizzare uno strumento di calcolo elettronico, il nomogramma di Fagan [3] semplifica i calcoli: è sufficiente collegare con una retta il valore della probabilità a priori, sulla sinistra, con il valore di LR+ (o il valore di LR-), al centro, per ottenere, nel punto di intersezione con la retta di destra, la probabilità a posteriori, ovvero la probabilità a priori aggiornata sulla base della verosimiglianza che caratterizza il test. Da notare che, proprio per superare il problema della espressione in odds, nel diagramma di Fagan le grandezze sia in ingresso (pre-test) sia in uscita (post-test) sono espresse in termini di probabilità.



Al di là di questi aspetti di tipo tecnico, resta tra la formulazione classica del teorema di Bayes e la formulazione basata sugli odds una <u>differenza nell'interpretazione dell'informazione a priori</u> (prima di avere eseguito il test) che è una differenza strutturale all'attuale epistemologia [4] e al mondo della statistica[75]. Per chiarire meglio cosa si intenda con ciò, ricordiamo l'esempio dovuto a De Finetti per illustrare la differenza tra le tre principali definizioni di probabilità. Immaginiamo una partita di calcio per la quale gli eventi possibili sono:

[75] *"…Nel mondo statistico si distinguono due diverse scuole di pensiero: la scuola classica o frequentista e la scuola bayesiana. La motivazione di fondo che divide i due approcci statistici riguarda il significato che essi danno alla nozione di probabilità. Dal punto di vista frequentista la probabilità è concepita in modo oggettivo. Essa rappresenta una caratteristica intrinseca degli eventi per i quali viene calcolata e deve essere depurata di ogni elemento soggettivo che la riguarda e la caratterizza. Dal punto di vista Bayesiano, invece, tali elementi non vengono considerati fonte di disturbo. Al contrario, sono il punto di partenza della definizione soggettiva di probabilità, secondo cui essa esprime un'opinione personale dell'individuo nei confronti di un determinato evento. Ciò sta a significare che la probabilità si colloca tra l'individuo e il mondo esterno e non già all'interno dell'evento… Se si chiede, ad esempio, ad un gruppo di persone di valutare la probabilità che la squadra di calcio A superi la squadra B nella prossima partita di campionato, è lecito aspettarsi tante differenti risposte: un tifoso della squadra A può reputare più probabile la vittoria di A rispetto ad un tifoso della squadra B. Nella statistica classica … l'inferenza … viene effettuata non considerando affatto l'eventualità della presenza di informazioni a priori sul fenomeno che si sta analizzando, assumendo cioè che le probabilità iniziali non esistano e che le valutazioni personali non debbano entrare nella trattazione dell'incertezza. L'identificazione del modello statistico, ovvero l'identificazione dei suoi parametri, è ottenuta esclusivamente basandosi sulle informazioni sperimentali, cioè sulle informazioni che scaturiscono da dati campionari ottenuti con la misura [delle] variabili d'interesse. Il motivo di questo atteggiamento è che nell'impostazione classica della statistica la probabilità è intesa come una frequenza e le sue proprietà sono definite come proprietà asintotiche legate ad un numero infinito di dati e quindi solo per fenomeni replicabili…. L'approccio bayesiano si basa principalmente sulla definizione soggettiva della probabilità. Essa cioè è interpretata come rappresentazione del "grado di fiducia" che un individuo ripone nel verificarsi di un determinato evento e dipende dallo stato di conoscenza, o di ignoranza, di tale evento, che fa parte di ogni individuo in modo diverso. Tale definizione si ripercuote sull'inferenza bayesiana, in quanto la determinazione dei parametri del modello statistico è ottenuta basandosi sulla preconoscenza, ovvero sulla disponibilità di informazioni a priori rispetto ai dati osservati nel campione, che dipende strettamente dall'esperienza precedentemente accumulata dall'individuo... Nelle procedure inferenziali bayesiane le informazioni espresse dalla probabilità a priori dei parametri possono essere aggiornate alla luce dei dati osservati, grazie al teorema di Bayes…."* [5]

- la vittoria della squadra di casa;
- la vittoria della squadra ospite;
- il pareggio.

Secondo la teoria (della probabilità) classica esiste 1 probabilità su 3 che avvenga la vittoria della squadra di casa. Secondo la teoria (della probabilità) frequentista ci si può dotare di un almanacco, controllare tutte le partite precedenti e calcolare la frequenza di un evento. Secondo la teoria (della probabilità) soggettiva, ci si può documentare sullo stato di forma dei calciatori, sul terreno di gioco e così via fino ad emettere un giudizio di probabilità (soggettiva).

E' possibile trasferire questi modelli alla diagnostica di laboratorio/diagnostica medica? Vediamo i tre casi. La teoria (della probabilità) classica non è ovviamente applicabile né alla diagnostica di laboratorio né alla diagnostica medica. La formulazione classica del teorema di Bayes ci collega alla teoria (della probabilità) frequentista. Infatti in questo caso la probabilità a priori viene interpretata come la frequenza a priori della malattia (prevalenza della malattia). Questa è una <u>probabilità oggettiva</u>, nella misura in cui tale è considerabile la frequenza (della malattia). La formulazione del teorema di Bayes basata sul rapporto di verosimiglianza ci collega alla interpretazione soggettiva della probabilità alla De Finetti. Utilizzando l'approccio basato sul rapporto di verosimiglianza e il diagramma di Fagan si vede come, dato un LR+ (o un LR-) fisso, la probabilità pre-test è un concetto che va al di là della prevalenza, e diventa/può diventare, nelle mani di ciascun medico, un giudizio a priori che include le informazioni sullo stato precedente del paziente (anamnesi, malattie pregresse, conoscenza "storica" del paziente da parte dello stesso medico), quelle derivanti dall'esame clinico, da altri esami strumentali. Il tutto diventa un giudizio di <u>probabilità soggettiva</u> da parte del medico, una probabilità a priori (pre-test) che viene quindi corretta dal rapporto di verosimiglianza per dare la probabilità a posteriori (post-test). Il risultato del test diagnostico "inietta" informazione nel motore inferenziale, e fornisce valore aggiunto aumentando l'informazione (diagnostica) a disposizione del medico. Dato uno specifico test e il suo specifico valore di LR, la <u>probabilità soggettiva a priori</u> di ciascun medico diventa il driver che pilota la diagnosi verso la direzione giusta (o verso quella sbagliata), da un lato lasciando lo spazio alla osservazione quotidiana della differente "bravura" di ciascun medico, dall'altro lato indicando come questa componente "intuitiva" e di "soggettività", che spiega la differente bravura, rappresenti ancora l'ostacolo alla formalizzazione rigorosa del processo di diagnosi medica.

<u>Bibliografia e riferimenti</u>

[1] Gill C. J., Sabin L. and Schmid C. H. Why clinicians are natural bayesians. BMJ 005;330;1080-1083.

[2] Deeks J. J. and Altman D. G. Diagnostic tests 4: likelihood ratios. *BMJ* 2004;329;168-169 (http://www.bmj.com/content/329/7458/168.full).

[3] http://www.childrens-mercy.org/stats/definitions/fagan.htm

[4] http://www.uniurb.it/Filosofia/isonomia/2005russo.pdf

[5] http://tesi.cab.unipd.it/archive/00000423/01/scaranaro.pdf

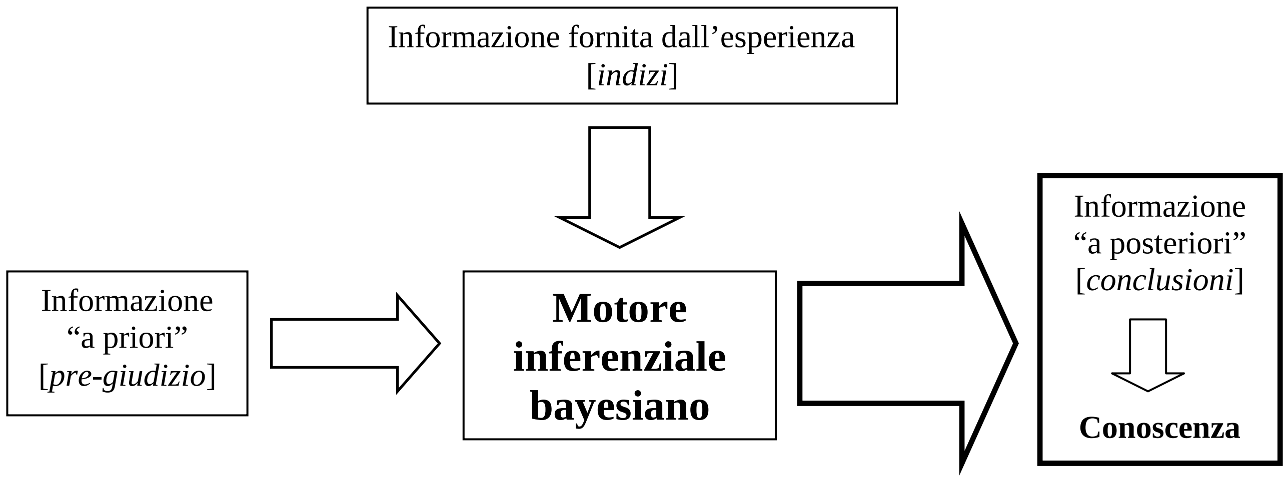
9. Epilogo

Il teorema di Bayes svolgo un ruolo centrale nel pensiero razionale: consente di aggiornare le informazioni sulla base dell'esperienza. Abbiamo una informazione "a priori": il teorema consente di combinarla con l'informazione data dall'esperienza e quindi di ottenere in uscita una informazione "a posteriori" che è quella a priori incrementata dell'informazione che l'esperienza è in grado di fornire.

Lo schema di aggiornamento del grado di fiducia mediante il meccanismo bayesiano

$$\textit{pregiudizio} + \textit{indizi} \rightarrow \textit{conclusioni}$$

nel processo di conoscenza è quello del detective, il cui paradigma è rappresentato dal sagace Sherlock Holmes con il suo ineffabile collega dottor Watson, di Sir Arthur Conan Doyle (1859-1930). Ma lo schema di aggiornamento del grado di fiducia mediante il meccanismo bayesiano è anche quello utilizzato comunemente nella diagnostica medica e più in generale nella ricerca scientifica.



Il problema gnoseologico ha accompagnato dalla notte dei tempi lo sviluppo del pensiero occidentale:
- gli antichi greci si concentrano sul tema della razionalità; scoprono il sillogismo, il principio del terzo escluso, costruiscono la geometria euclidea, e contemporaneamente scoprono i paradossi. Forse per la loro cultura, che disdegnava il lavoro manuale, nel quale va inclusa la sperimentazione, trascurano l'apporto che l'informazione fornita dall'esperienza può dare alla conoscenza;
- Galileo valorizza l'informazione fornita dall'esperienza stabilendo il valore dell'esperimento;
- Cartesio apre il fronte dell'informazione a priori, richiamando all'esigenza di "idee chiare e distinte" sui cui fondare la conoscenza. Se la soluzione che propone non risolve il problema della saldezza delle fondamenta dell'informazione a priori (perché irrisolvibile), di lui restano le esigenze di razionalità nel procedere sulla strada della conoscenza che stanno alla base del metodo;
- la teoria della probabilità introduce l'incertezza nella deduzione. Anche da premesse certe (le facce di un dado), si arriva a conclusioni solo probabili (il risultato di un singolo lancio). Ritorna il tema della razionalità: il teorema di Bayes, per la prima volta, fornisce un motore inferenziale in grado di ribaltare il percorso (puramente deduttivo) dalla causa all'effetto, dall'universale al particolare; il problema inverso, il passare dall'effetto alla causa, dal particolare all'universale, viene risolto utilizzando l'informazione fornita dall'esperienza, anche se a patto di rinunciare alla certezza delle conclusioni/della conoscenza. Ma queste conseguenze del teorema di Bayes rimangono in ombra per due secoli;
- la critica al rapporto di causa e effetto e la critica all'induzione sono messe in secondo piano dalla visione determinista del mondo, favorita dei successi della scienza da Newton fino all'inizio del XX

secolo;
- le scoperte della fisica che si susseguono dal 1905 (anno della pubblicazione da parte di Einstein dei tre famosi scritti su teoria della relatività speciale, effetto fotoelettrico e moto browniano) fino alla cromodinamica quantistica, attivano una serie di considerazioni critiche sull'ottimismo della visione deterministica, che culminano nella rivalutazione della critica all'induzione di Hume da parte di Karl Popper, e nella riscoperta della probabilità come fenomeno ineludibile nel sapere scientifico. Einstein e Popper aggiungono l'importanza dell'immaginazione: la "teoria" è invenzione, che il metodo scientifico prevede di mettere in discussione con l'esperimento. Il XX secolo ci ha trasbordati nel terzo millennio con una consapevolezza: il sapere scientifico non può darci né verità né certezze.

Ritornando a quel piccolo ma significativo ambito del procedere "*more scientifico*" che è rappresentato dalla diagnostica di laboratorio, quelle che possono apparire come difficoltà nell'applicazione del teorema di Bayes sono in realtà i problemi derivanti dalla incertezza che contraddistingue l'informazione che alimenta il motore inferenziale. Dato per evidente il valore aggiunto fornito dall'esperienza (effettuato nelle condizioni opportune, un test di laboratorio fornisce informazione), l'informazione a priori non è certa, come non è certa l'informazione fornita dall'esperienza. Pertanto anche utilizzando un motore inferenziale "razionale", come il teorema di Bayes, l'informazione a posteriori risulta incerta, e tale risulta la conoscenza che ne deriva.

Il rigore del teorema di Bayes deve sempre venire a patti con la realtà: che ci vede "dare forma" ai dati nel momento in cui li trasformiamo in informazione. Alla fine varrà sempre il vecchio adagio di John Milton per cui «ragionare non è altro che scegliere», sapendo però, come amava dire John Looke, che "ci risolviamo per questo o quel partito non nella chiara luce del mezzogiorno, bensì nel crepuscolo della probabilità».

APPENDICE A
GRANDEZZE E UNITÀ DI MISURA NEL LABORATORIO CLINICO

In medicina di Laboratorio, alla base dell'espressione di risultati vi è "*la misura (dell'entità) di una grandezza fisica*". Questa misura consiste nell'esprimere la grandezza in modo quantitativo, dando ad essa un "*valore numerico*" che è un numero puro, ottenuto per confronto della "*(entità della) grandezza* in esame" con la "(entità di una) grandezza di riferimento ad essa omogenea, definita *unità di misura*", essendo quindi

(entità della) grandezza / unità di misura = valore numerico

da cui si ricava

(entità della) grandezza = valore numerico · unità di misura

Da quest'ultima espressione si deduce che il "*risultato di una misura*" è dato dal prodotto di un "*numero*" per la "*unità di misura*": pertanto l'indicazione di quest'ultima non deve mai essere omessa [1].

Il sistema SI

Allo scopo di pervenire ad una immediata comprensione, in qualsiasi Paese, della espressione dei risultati di una misura, le organizzazioni internazionali e nazionali a ciò preposte hanno proceduto alla codificazione di un sistema di unità di misura delle varie grandezze, unificato nella definizione, nella nomenclatura e nella simbologia. Il sistema di base, oggi adottato, discende dal sistema metrico decimale (introdotto alla fine del XVIII secolo[76]) ed ha il nome di "Sistema Internazionale di Unità di Misura" (l'abbreviazione è SI); esso è stato sancito dalla Conferenza Generale dei Pesi e Misure (CGPM) nel 1960 e nel 1971, accettato dalla Comunità Economica Europea (CEE) nel 1980 e divenuto legale in Italia nel 1982[2].

Il sistema SI è fondato su sette grandezze e relative unità di base, indipendenti l'una dall'altra:

Grandezza di base	*Simbolo*	*Unità di base*	*Simbolo*
lunghezza	l	metro	m
massa	m	kilogrammo	kg
tempo	t	secondo	s
intensità di corrente elettrica	I	ampere	A
temperatura termodinamica	T	kelvin	K
quantità di sostanza	n	mole	mol
intensità luminosa	I	candela	cd.

A questa vanno aggiunte due grandezze supplementari, che fanno pure esse parte integrante del sistema SI:

[76] *"…che ci sia una sola misura e un solo peso in tutto il Regno…e anche una misura uniforme per i vini, almeno nella stessa provincia" si chiedeva insistentemente nei "Cahiers de Doléances" ai tempi della rivoluzione francese.*

Grandezza supplementare	*Simbolo*	*Unità supplementare*	*Simbolo*
angolo piano	a,b,g...	radiante	rad
angolo solido	w,O	steradiante	sr

Per indicare multipli e sottomultipli delle unità sono previsti i fattori riportati nella seguente tabella:

Nome	*Simbolo*	*Fattore*	*Fattore di moltiplicazione*
yotta	Y	10^{24}	1 000 000 000 000 000 000 000 000
zetta	Z	10^{21}	1 000 000 000 000 000 000 000
exa	E	10^{18}	1 000 000 000 000 000 000
peta	P	10^{15}	1 000 000 000 000 000
tera	T	10^{12}	1 000 000 000 000
giga	G	10^{9}	1 000 000 000
mega	M	10^{6}	1 000 000
kilo	k	10^{3}	1 000
etto	h	10^{2}	100
deca	da	10^{1}	10
deci	d	10^{-1}	0,1
centi	c	10^{-2}	0,01
milli	m	10^{-3}	0,001
micro	µ	10^{-6}	0,000 001
nano	n	10^{-9}	0,000 000 001
pico	p	10^{-12}	0,000 000 000 001
femto	f	10^{-15}	0,000 000 000 000 001
atto	a	10^{-18}	0,000 000 000 000 000 001
zepto	z	10^{-21}	0,000 000 000 000 000 000 001
yocto	y	10^{-24}	0,000 000 000 000 000 000 000 001

Tra le principali regole adottate dal sistema SI, si rammentano le seguenti:

- sono raccomandati i fattori che fanno variare l'unità di un fattore 1000 (kilo, mega, milli, micro, eccetera), come riportato nella precedente tabella;
- è sconsigliato l'uso dei fattori che fanno variare le unità di un fattore 10 o 100 (deca, etto, deci, centi);
- dopo i simboli non si deve mettere il punto (cm e non cm., mol e non mol., eccetera): si tratta appunto di simboli, e non di abbreviazioni;
- non si devono usare i fattori da soli; il nome o il simbolo dell'unità non deve essere omesso (micrometro o μm, e non micron o μ; kilogrammo e non kilo, eccetera);
- non si devono usare unità con nomi d'uso tipo il lambda (λ) per il microlitro (μL) e il gamma (γ) per il microgrammo (μg);
- non devono essere formate unità con più di un prefisso (nanometro e non millimicrometro o, peggio ancora, millimicron, eccetera);
- i multipli e i sottomultipli dell'unità di massa (kilogrammo), che già contiene un prefisso, si formano anteponendo i prefissi al grammo (quindi μg e non nKg, eccetera).

Dalle grandezze e unità di base e supplementari SI è possibile ricavare grandezze e unità SI "derivate", di cui numerose hanno importanza in campo biomedico:

Grandezza	*Nome unità*	*Simbolo unità*
Frequenza (cicli al secondo)	hertz	Hz
Forza	newton	N
Pressione e tensione	pascal	Pa
Energia, lavoro, quantità di calore	joule	J

Potenza, flusso energetico	watt	W
Quantità di elettricità, carica elettrica	coulomb	C
Tensione elettrica, potenziale elettrico, forza elettromotrice	volt	V
Resistenza elettrica	ohm	Ω
Conduttanza	siemens	S
Capacità elettrica	farad	F
Flusso d'induzione magnetica	weber	Wb
Induzione magnetica	tesla	T
Induttanza	henry	H
Temperatura Celsius	grado Celsius	°C
Flusso luminoso	lumen	lm
Illuminamento	lux	lx
Attività (irraggiamento ionizzante)	becquerel	Bq
Dose assorbita	gray	Gy
Equivalente di dose	sievert	Sv

Della precedente tabella si fa notare in particolare che la temperatura nel sistema SI viene misurata in gradi Celsius (°C) e non in gradi centigradi come comunemente (ed erroneamente) si continua a dire.

Infine si ricordano a parte le seguenti grandezze e unità che, in quanto non SI, continuano ad essere ammesse, sono ammesse per usi particolari o, come accade per molte, non sono più ammesse:

Grandezza	***Nome unità***	***Simbolo unità***	***Osservazioni***
Volume	litro	l o L	ammessa
Massa	tonnellata	t	ammessa
Pressione e tensione	bar	bar	ammessa
Tempo	minuto	min	ammessa
	ora	h	ammessa
	giorno	d	ammessa
Pressione	millimetro di mercurio	mmHg	ammessa solamente per pressione sangue e fluidi biologici
Lunghezza	angström	Å	non più ammessa
Lunghezza	micron	μ	non più ammessa
Pressione	atmosfera (standard)	atm	non più ammessa
Quantità di calore	caloria	cal	non più ammessa
Potenza	cavallo vapore	CV o HP	non più ammessa
Attività di radionuclidi	curie	Ci	non più ammessa
Dose assorbita	rad	rad	non più ammessa
Equivalente di dose	rem	rem	non più ammessa
Esposizione (raggi x o γ)	röntgen	R	non più ammessa

Dal confronto di quest'ultima tabella con quella che la precede è facile notare alcuni importanti cambiamenti intervenuti in campo biomedico:

- scomparsa dell'angström (Å);
- scomparsa del micron (μ);
- caloria (cal) sostituita dal joule (J);
- curie (Ci) sostituito dal becquerel (Bq);
- rad (rad) sostituito dal gray (Gy);
- rem (rem) sostituito dal sievert (Sv).

Per una trattazione completa del sistema SI vedere Fazio[3]. Per le applicazioni del sistema SI all'espressione dei risultati delle analisi di laboratorio vedere Besozzi [1].

Espressione dei risultati nel referto

Attualmente per l’espressione dei risultati è raccomandato l’utilizzo del metodo razionale proposto dalla IFCC (International Federation for Clinical Chemistry and Laboratory Medicine), che prevede per ciascun analita di riportare: il sistema (materiale) su cui viene effettuata l’analisi, del componente (sostanza analizzata o analita), della grandezza utilizzata per esprimere il risultato, del metodo impiegato, del valore numerico del risultato, delle unità di misura impiegate per esprimere il risultato, e degli intervalli di riferimento.

Il sistema

Il sistema è il materiale su cui viene effettuata l’analisi. Nel referto viene abbreviato sotto forma di una sigla, con le principali corrispondenze riportate nella tabella che segue

Sigla	*Sistema (materiale su cui viene eseguita l’analisi)*
A	(sangue) arterioso
Er	eritrocita/i
F	feci
Hb	emoglobina
Lc	leucocita/i
LCR	liquido cefalorachidiano
LS	liquido seminale
P	plasma
Ps	piastrina/e
Pt	paziente (nel caso di test)
S	siero
Sd	sudore
Sg	sangue
U	urine
V	(sangue) venoso

Il componente

La descrizione del componente viene fatta impiegando una terminologia sostanzialmente analoga a quella prima impiegata per descrivere le analisi, con una razionalizzazione tesa a introdurre, laddove possbile, denominazioni sistematiche. Così ad esempio, si parla sempre di CONTA DI ADDIS, mentre per rispettare la denominazione sistematica prevista dall'enzimologia, la parte della biochimica che si occupa degli enzimi, la transaminasi glutammico-ossalacetica (sigla GOT) viene oggi denominata ASPARTATO AMMINOTRANSFERASI (sigla AST), e la transaminasi glutammico-piruvica (sigla GPT) viene oggi denominata ALANINA AMMINOTRANSFERASI (sigla ALT).

Nella tabella che segue sono riportati alcuni esempi di cambiamenti effettuati in seguito all'adozione, nell'espressione dei risultati, del criterio "sistema-componente" in luogo della tradizionale e ormai da considerare obsoleta "denominazione dell'analisi"

Vecchia denominazione ('analisi)	*Nuova denominazione (sistema-componente)*
Acido lattico	S-LATTATO
Acido valproico	S-VALPROATO
Amilasi	S-alfa-AMILASI

Azotemia	S-UREA
Bicarbonati	S-IDROGENOCARBONATO
Bilirubinemia totale	S-BILIRUBINA TOTALE
Calcemia	S-CALCIO
Creatinfosfato-chinasi (CPK)	S-CREATINA CHINASI (CK)
Fosforemia	S- FOSFATO
Glicemia	P-GLUCOSIO (se eseguito su plasma) S-GLUCOSIO (se eseguito su siero)
Glicosuria	U-GLUCOSIO
Lattico-deidrogenasi (LDH)	S-LATTATO DEIDROGENASI (LDH)
Ormone luteinizzante (LH)	S-LUTEOTROPINA (LH)
Ormone della crescita (GH)	S-SOMATOTROPINA (STH)
Ormone follicolo stimolante (FSH)	S-FOLLITROPINA (FSH)
Proteinuria	U-PROTEINE TOTALI
Sideremia	S-FERRO
Transaminasi glutammico-ossalacetica (GOT)	S-ASPARTATO AMMINOTRANSFERASI (AST)
Transaminasi glutammico-piruvica (GPT)	S-ALANINA AMMINOTRANSFERASI (ALT)
Uricemia, acido urico	S- URATO

La grandezza

La grandezza con cui sono espressi i risultati deve essere ovviamente una delle grandezze previste dal sistema SI. Il cambiamento principale riguarda la transizione, ogniqualvolta sia possibile, cioè ogniqualvolta si conosca la massa della mole (ex peso molecolare) della sostanza in esame, dall'espressione in termini di *massa* (o di *concentrazione di massa*) all'espressione in termini di *quantità di sostanza* (o di *concentrazione di sostanza*). In altre parole la transizione dai milligrammi alle millimoli (ovvero dai milligrammi/litro alle millimoli/litro), dai grammi alle moli (ovvero dai grammi/litro alle moli/litro), e così via. Questo cambiamento ha un razionale molto forte: infatti le relazioni numeriche tra le diverse molecole in soluzione riflettono importanti correlazioni fisiopatologiche.

Si rammenta che per definizione il numero di moli (mol) è uguale al rapporto tra la massa in grammi (g) della sostanza in esame e la sua massa della mole (ex peso molecolare), ovvero

moli = grammi / massa della mole

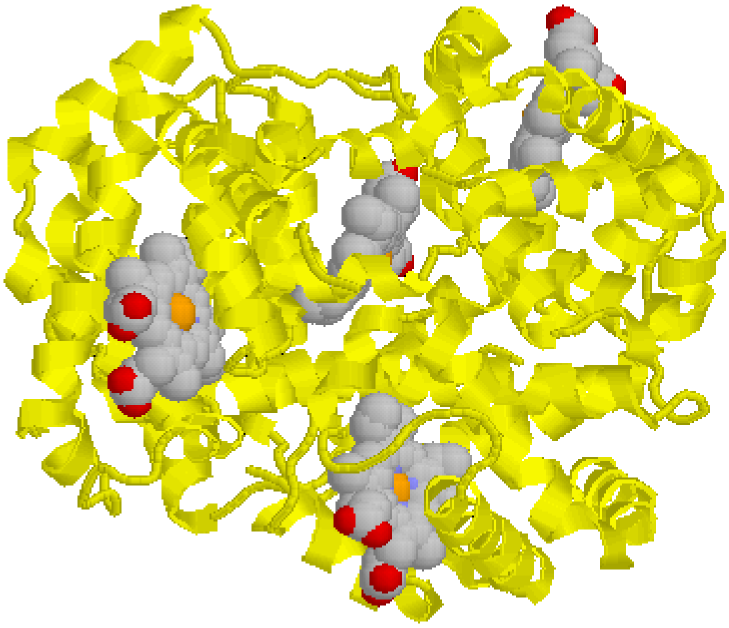
da cui deriva che 1 mole di sostanza è una quantità di sostanza la cui massa in grammi è numericamente uguale alla massa della mole (e ancora 1 millimole di sostanza è una quantità di sostanza la cui massa in milligrammi è numericamente uguale alla massa della mole).

Una mole di qualsiasi sostanza contiene per definizione sempre lo stesso numero di molecole[77].

Si pensi all'espressione "4,0 grammi di albumina sono in grado di legare non più di 34 milligrammi di bilirubina" e all'espressione "580 micromoli (μmol) di albumina sono in grado di legare non più di 580 micromoli di bilirubina". Questa seconda espressione rivela in modo più immediato della prima il fatto che l'albumina, che oltre alla funzione di mantenere la pressione colloidosmotica ha anche una funzione di trasporto, è in grado di legare la bilirubina secondo un rapporto al massimo di 1:1 (una molecola di bilirubina per ogni molecola di albumina).

[77] *Questo numero è espresso dalla costante N (costante di Avogadro), essendo N uguale a $6{,}02252 \cdot 10^{23}$ mole $^{-1}$.*

Si pensi ancora all'espressione "1 grammo di emoglobina trasporta 1,37 millilitri (mL) di ossigeno (O_2), contiene 3,5 milligrammi (mg) di ferro e forma 35,9 milligrammi (mg) di bilirubina". Non risulta evidente da questo il fatto che una molecola di emoglobina, come mostra la successiva immagine, è un tetramero, formato da quattro catene polipeptidiche contenenti altrettanti gruppi eme, ciascuno dei quali è in grado di legare un atomo di O_2, ciascuno dei quali contiene un atomo di ferro (l'atomo arancione al centro dei gruppi eme), e ciascuno dei quali porta alla formazione, nel catabolismo dell'emoglobina, di una molecola di bilirubina. Il fatto quindi che 1 molecola di emoglobina contenga 4 molecole di eme, leghi 4 molecole di ossigeno e dia luogo a 4 molecole di bilirubina, risulta invece assolutamente evidente nell'espressione "15 micromoli (μmol) di emoglobina trasportano 60 micromoli di ossigeno, contengono 60 micromoli di ferro e formano 60 micromoli di bilirubina".



Detto questo va precisato che, mentre in numerosi paesi (europei e non) la transizione dall'espressione in termini di massa (o di concentrazione di massa) all'espressione in termini di sostanza (o di concentrazione di sostanza) è praticamente completata, in Italia non è ancora avvenuto nulla. A discolpa di coloro che (come lo scrivente) ritengono tale transizione necessaria stante la sua razionalità scientifica, ma hanno finora evitato di introdurla nei propri laboratori, c'è peraltro il fatto che l'operare in modo non contemporaneo sul territorio nazionale avrebbe sicuramente ingenerato grande confusione. Per questo si attende finalmente per il terzo millennio un'iniziativa a livello nazionale.

Il metodo

Rappresenta un'indicazione importante, in quanto ancora oggi metodi diversi possono fornire risultati differenti, che dovranno peraltro essere valutati in relazione agli intervalli di riferimento specifici del metodo utilizzato, che quindi non devono mai mancare.

Il valore numerico del risultato

Deve essere espresso utilizzando un numero adeguato di cifre significative, secondo quanto esposto in precedenza. Si rammenta ancora, data l'attuale imprecisione dei metodi analitici per la determinazione del calcio, per la determinazione degli enzimi, l'esprimere un risultato come "0,84 mg/dL" ha scarso significato: un solo decimale (0,8 mg/dL) rappresenta il modo corretto di esprimere un tale risultato.

Le unità di misura

Anche le unità di misura con cui sono espressi i risultati devono essere ovviamente quelle previste dal sistema SI, con la raccomandazione aggiuntiva di esprimere le concentrazioni riportando al denominatore il litro ((L). Questo può essere fatto in molti casi senza determinare alcun cambiamento nel valore numerico del risultato, se si considera il fatto che, ad esempio, i microgrammi/millilitro (µg/mL) sono numericamente identici ai milligrammi/litro (mg/L), che le milliunità/millilitro (mU/mL) sono numericamente identici alle unità/litro (U/L) e così via.

L'intervallo di riferimento

L'*intervallo di riferimento* rappresenta lo strumento fondamentale per giudicare se un risultato è ancora all'interno dei valori tipici dei soggetti sani, ovvero di quanto se ne discosta.

In alternativa rispetto all'intervallo di riferimento, possono comparire nel referto, in funzione dell'analita che viene determinato, il *valore desiderabile*, o l'*intervallo terapeutico* (nel caso dei farmaci), o il *valore che definisce la malattia* [vedere i dettagli relativi a questi e agli altri *valori decisionali* al capitolo 8]

In linea teorica, due risultati consecutivi possono essere confrontati tra di loro anche quando ottenuti in laboratori che utilizzando metodi differenti, con differenti intervalli di riferimento, rapportandoli al limite superiore dell'intervallo di riferimento (ci si riferisce qui al caso in cui la patologia determina un aumento di concentrazione). Così, ad esempio, un valore di 60 U/L dell'AST dato un intervallo di riferimento "inferiore a 30" risulterà sovrapponibile a un valore dell'AST di 80 U/L dato un intervallo di riferimento "inferiore a 40". Questo approccio, che avrebbe potuto portare all'uniformazione dell'espressione dei risultati, e alla loro semplificazione, non ha peraltro avuto successo quando, in passato, è stato proposto, anche a causa del fatto che, in due metodi analitici diversi, le proporzioni tra le grandezze non sono sempre conservate.

Resta quindi il fatto che una completa razionalizzazione nella interpretazione dei risultati passa anche attraverso una uniformazione dei metodi analitici: un processo "per consenso" delle Società Scientifiche e dell'industria dei diagnostici che rimane comunque lento e spesso difficile.

Alcuni segni matematici di largo uso

Si riportano qui di seguito alcuni segni matematici di largo uso, che talvolta vengono impiegati in modo improprio:

Segno	***Significato***
÷	intervallo da… a… (estremi inclusi)
=	uguale a
≡	identico a

$\neq$	diverso da
$\approx$	uguale a circa
$\propto$	proporzionale a
∞	infinito

Bibliografia e riferimenti

[1] Besozzi M, De Angelis G, Franzini C. Espressione dei risultati nel laboratorio di chimica clinica. Milano:Società Italiana di Biochimica Clinica, 1989.

[2] DPR n. 802 del 12 agosto 1982, Attuazione della direttiva (CEE) n. 80/181 relativa alle unità di misura. Suppl Ord Gazz Uff della Repubblica Italiana n. 302 del 3 novembre 1982

[3] Fazio M. Dizionario e manuale delle unità di misura. Bologna:Zanichelli, 1985.

APPENDICE B
SBAGLI, ERRORI E DISTRIBUZIONE GAUSSIANA

Innanzitutto gli errori devono essere mantenuti distinti dallo *sbagli* (*errori grossolani*), che sono un accidente tecnico, e che come tale si manifesta nel corso dell'applicazione delle conoscenze.

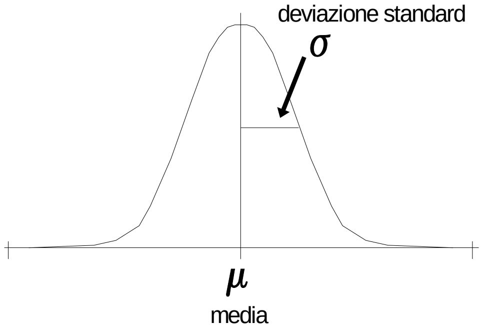
Nel *Vocabolario illustrato della lingua italiana* di G. Devoto e G. C. Oli si trova che "*[sbaglio]...condivide con errore...tutte le determinazioni, ma gnrl. differisce nel senso di un'attenuazione dell'importanza e della gravità...*". A fronte di questo impiego del termine nel linguaggio comune, sta il fatto che lo sbaglio viene dallo stesso vocabolario ulteriormente definito come "..*mancanza nei confronti di un ordine corretto o di una regola...*": questa definizione si avvicina molto a quella qui data di sbaglio come accidente tecnico. Più concisamente ne *Lo Zingarelli 1995* di N. Zingarelli lo sbaglio viene definito come "*...equivoco, disattenzione, svista...*", in un senso che va ancora più chiaramente nella direzione del significato qui adottato.

Gli sbagli sono legati prevalentemente all'organizzazione e quindi ai processi di comunicazione (esempi di sbagli in laboratorio sono l'errata trascrizione di un dato numerico, un risultato sbagliato in seguito all'utilizzo di un reagente scaduto, lo scambio di campioni, una errata interpretazione del risultato di un test di gravidanza acquistato in farmacia ed eseguito a casa propria dalla paziente, che ha frainteso i criteri per l'interpretazione dei risultati del test). Contrariamente a quanto avviene per gli errori, gli sbagli si possono evitare operando con cura, e migliorando il sistema organizzativo. Contrariamente a quanto avviene per gli errori, non e' possibile fissare un livello di tolleranza per gli sbagli, cioè definire una percentuale ammissibile di sbagli: semplicemente, gli sbagli per definizione non si devono verificare.

L'errore è invece legato all'incertezza intrinseca alle nostre conoscenze scientifiche a causa dei limiti inerenti ai sistemi (*strumenti di misura*) impiegati per rilevare i segnali provenienti dalle grandezze fisiche [1]. Nell'approccio all'analisi dell'errore [2] svolge un ruolo centrale la distribuzione gaussiana.

Essendo π uguale a 3,1415 ed essendo e la base dei logaritmi naturali (e = 2,7183), per un dato valore della x l'equazione della distribuzione gaussiana

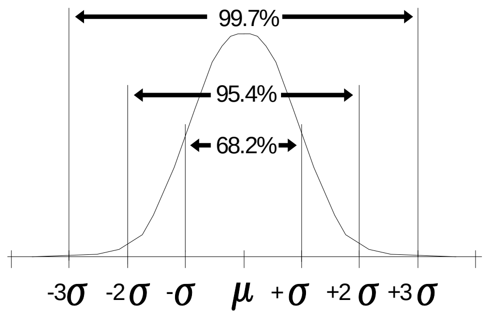
$$y = [\, 1 / (\sigma \cdot \sqrt{2\pi}) \,] \cdot e^{-(x-\mu)^2 / 2\sigma^2}$$



risulta completamente determinata dai due soli valori μ e σ.

Il valore μ (la *media della popolazione*) rappresenta la misura di posizione della distribuzione gaussiana, mentre il valore σ (la *deviazione standard della popolazione*) rappresenta la misura di dispersione della distribuzione gaussiana.

Per convenzione la distribuzione gaussiana teorica viene assunta come avere una media μ = 0 e una deviazione standard σ = 1. In questo modo la superficie sottesa dalla curva gaussiana teorica ha una superficie uguale a 1, che corrisponde al valore di probabilità *p = 1* che include tutte le osservazioni. Risulta pertanto facile calcolare il numero (espresso in percentuale) delle osservazioni che cadono all'interno di un dato intervallo (per esempio entro un numero dato di deviazioni standard rispetto alla media).



Per confrontare una distribuzione gaussiana (o ritenuta tale) con la gaussiana teorica è necessario standardizzare i risultati ottenuti in modo che la media osservata μ venga riportata a 0 e la deviazione standard osservata σ venga riportata a 1. In pratica ciò si ottiene trasformando ogni valore della x osservato nella corrispondente z (*deviata normale standardizzata*), essendo

$$z = (x - \mu) / \sigma$$

L'operazione inversa (trasformazione di una deviata normale standardizzata z nel valore x corrispondente ad un dato appartenente a una distribuzione gaussiana con media μ e deviazione standard σ risulta possibile applicando la trasformazione

$$x = \mu + \sigma \cdot z$$

Infine si ricorda che la media μ e la deviazione standard σ sono denominate globalmente "parametri" della distribuzione gaussiana. Per questo le tecniche statistiche che basano la loro validità (la validità delle conclusioni che mediante esse è possibile trarre) su assunti distribuzionali di gaussianità sono denominate *tecniche statistiche parametriche*. In contrapposizione alle tecniche che basano la loro validità su assunti distribuzionali minimi, e comunque non richiedono una distribuzione gaussiana dei dati per garantire conclusioni valide, che sono denominate *tecniche statistiche non parametriche*. Non è comunque un caso che in presenza di distribuzioni esattamente gaussiane tecniche statistiche parametriche e tecniche statistiche non-parametriche forniscono identici risultati.

I *metodi statistici parametrici* sono basati sull'assunto che i dati campionari siano estratti da una popolazione che ha una distribuzione gaussiana (esistono in realtà come vedremo poco più avanti anche *metodi statistici non-parametrici*, che possono, anzi devono essere impiegati quando i dati non sono distribuiti in modo gaussiano).

In effetti, come riconoscono Snedecor e Cochran [3] "...*è stupefacente che la distribuzione gaussiana abbia dominato sia la teoria che la pratica statistica*...". Tuttavia sono gli stessi Autori a indicare i quattro argomenti a favore dell'utilizzo della statistica parametrica:

⇒ la distribuzioni di molte variabili è approssimativamente gaussiana;
⇒ per distribuzioni non gaussiane, semplici trasformazioni matematiche (come per esempio la radice quadrata e la trasformazione logaritmica dei dati) consentono spesso di ottenere distribuzioni approssimativamente gaussiane;
⇒ la distribuzione gaussiana può essere trattata facilmente in termini matematici;

⇒ anche se la distribuzione della popolazione originaria non è gaussiana, la distribuzione delle medie campionarie tende a divenire gaussiana (teorema centrale del limite). Quest'ultimo è il singolo argomento più consistente a favore della statistica parametrica.

Sono riassunti qui di seguito i concetti e la terminologia utilizzati nell'approccio classico all'analisi dell'errore [2]. Si consideri l'operazione di misura ripetuta della concentrazione del colesterolo nel siero di uno stesso individuo. La misura viene effettuata *n* volte su un unico campione di siero, raccolto in un'unica volta. I concetti di base sono i seguenti:

⇒ a causa dei limiti inerenti ai sistemi (*strumenti di misura*) impiegati per rilevare i segnali provenienti dalle grandezze fisiche, ogni *misura sperimentale* è inevitabilmente accompagnata da un qualche grado di *incertezza;*

⇒ si può assumere che esista, ogniqualvolta viene effettuata una misura sperimentale, un valore teorico, detto *valore vero*: quello che si otterrebbe se la misura non fossa affetta da alcuna incertezza;

⇒ per effetto della incertezza che la caratterizza, la misura sperimentale rappresenta una *stima* più o meno approssimata del valore vero;

⇒ la differenza tra una singola misura sperimentale e il suo valore vero rappresenta l'*errore* ;

⇒ ripetendo più volte una misura, è possibile pervenire ad una migliore caratterizzazione dell'errore;

⇒ in un insieme di misure ripetute, si definisce come *errore casuale* l'errore per cui le singole misure differiscono (casualmente, cioè senza nessuna regola apparente al succedersi delle misure stesse) tra di loro;

⇒ in un insieme di misure ripetute, si definisce come *errore sistematico* l'errore per cui l'insieme (preso globalmente) delle misure ripetute si discosta dal valore vero;

⇒ un insieme di misure ripetute dello stesso fenomeno può essere riassunto sotto forma di una *distribuzione di frequenze*;

⇒ nel caso degli errori di misura si assume generalmente che la distribuzione di frequenze segua un *modello gaussiano*;

⇒ in una *distribuzione gaussiana* la *media aritmetica* è la *misura di posizione* della distribuzione, mentre la *deviazione standard* è la *misura di dispersione* dei dati attorno alla media;

⇒ la *precisione* è il grado di concordanza di una serie di misure ripetute. Convenzionalmente la precisione, che non ha valore numerico, viene misurata in termini di una quantità (*imprecisione*) il cui valore diminuisce all'aumentare dell'attendibilità delle misure;

⇒ l'*imprecisione*, calcolata come *deviazione standard* (*s*), può essere espressa sia come tale (e quindi nelle unità originali) che come percentuale della media (*deviazione standard relativa*, meglio nota come *coefficiente di variazione*, *CV*). L'imprecisione è la *misura di dispersione* dei dati attorno alla media e rappresenta una *stima* dell'errore casuale. Tale stima è tanto più attendibile quanto più numerose sono le misure utilizzate per effettuarla (*numerosità campionaria*);

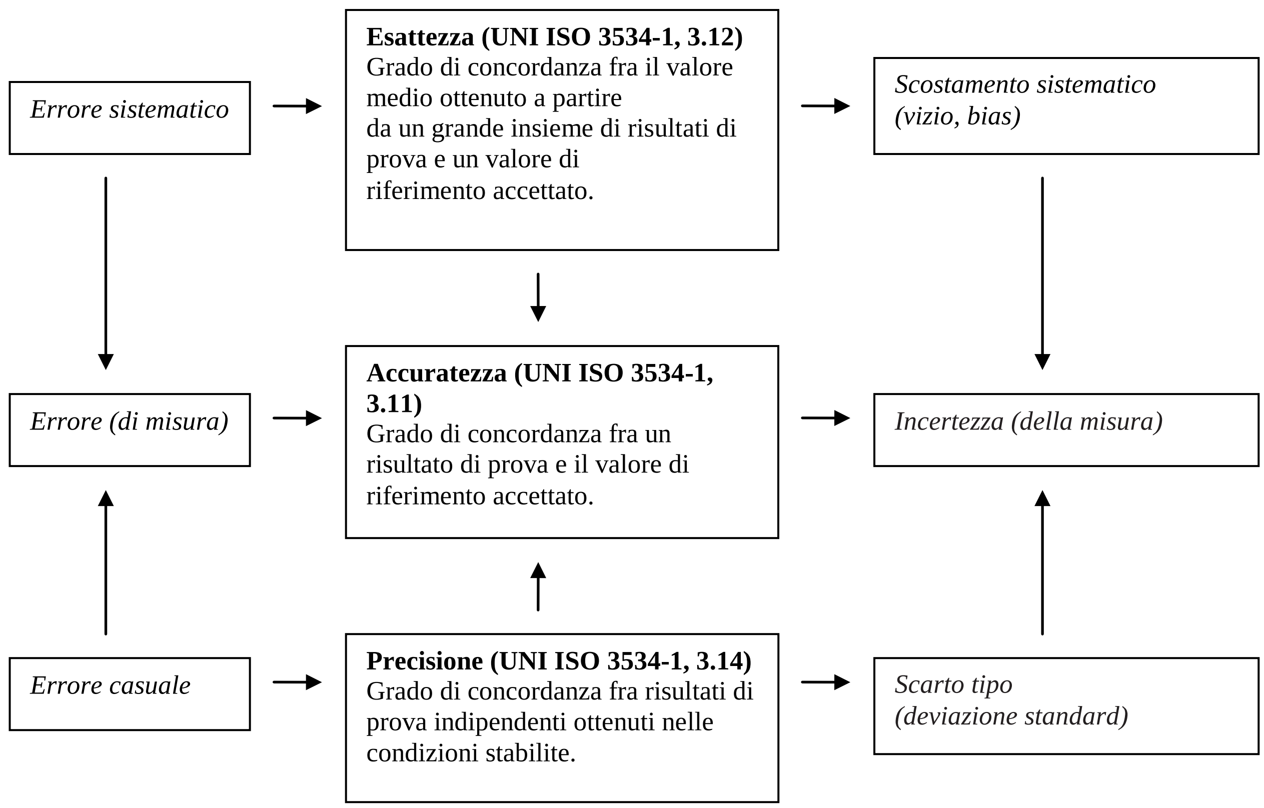
⇒ l'*accuratezza* è il grado di concordanza tra la media di una serie di misure ripetute (la *misura di posizione* della distribuzione) e il valore vero. Convenzionalmente l'accuratezza, che non ha valore numerico, viene misurata in termini di una quantità (*inaccuratezza*) il cui valore diminuisce all'aumentare dell'attendibilità delle misure;

⇒ l'*inaccuratezza*, calcolata come differenza tra la media campionaria e il valore vero, può essere espressa sia come tale (e quindi nelle unità originali) che come percentuale del valore vero. L'inaccuratezza rappresenta una *stima* dell'errore sistematico. Tale stima è tanto più attendibile quanto più numerose sono le misure utilizzate per effettuarla (numerosità campionaria);

⇒ imprecisione e inaccuratezza sono tra loro indipendenti in quanto sono collegate alle due grandezze indipendenti che concorrono a determinare la distribuzione gaussiana.

L'approccio classico sopra riportato è stato negli ultimi anni corretto, introducendo un modello generale per la stima dell'incertezza di misura con *"… la definizione di requisiti il cui fine ultimo è quello di produrre un obiettivo miglioramento della confrontabilità delle misurazioni in tempi e luoghi diversi, in tutti i settori delle attività umane …. sulla base di un ampio consenso internazionale sostenuto da solide basi scientifiche.*" [4]. L'approccio basato sull'incertezza a rigore prevede che l'unico *attributo della misurazione* sia rappresentato dall'accuratezza, e che l'unica *misura dell'attributo* accuratezza sia rappresentata dall'incertezza con la quale è stata ottenuta la misura [5]. Tuttavia tale approccio non è al momento universalmente condiviso, e la via più seguita nella definizione dell'errore e delle sue componenti prevede la sintesi tra approccio classico a approccio basato sull'incertezza che viene illustrata nella schema che segue:

Tipo di errore	***Attributo della misurazione***	***Misura dell'attributo***



L'esattezza è un attributo, quindi un concetto qualitativo. Viene misurata mediante lo scostamento sistematico (vizio), calcolato come "…*differenza tra la speranza matematica dei risultati di prova e un valore di riferimento accettato…*".

La precisione è un attributo, quindi un concetto qualitativo. Viene misurata mediante lo scarto tipo che è la deviazione standard calcolata sui "*risultati di prova indipendenti ottenuti nelle condizioni stabilite*"[78].

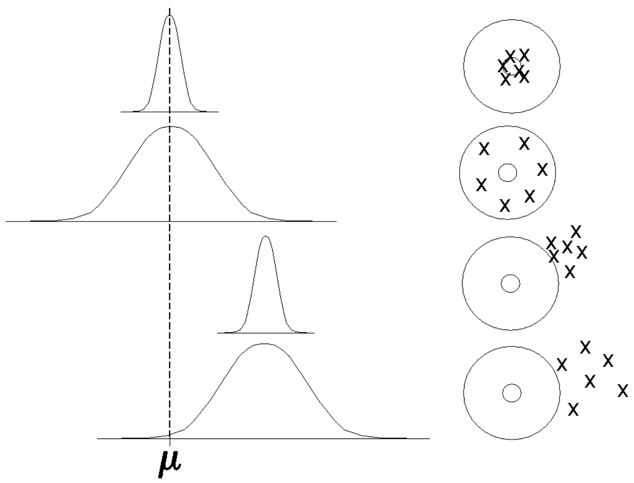
[78] Dato un campione che include *n* dati (osservazioni) x_i, la media campionaria $\bar{x}$ è calcolata come

$$\bar{x} = \Sigma x_i / n$$

ed essendo $\Sigma (x_i - \bar{x})^2$ la devianza, la varianza s^2 dei dati campionari è calcolata dividendo la devianza per i gradi di libertà $(n - 1)$, ovvero come

L'accuratezza è un attributo, quindi un concetto qualitativo. Viene misurata mediante l'incertezza (di misura), che è il "*...parametro, associato al risultato di una misurazione, che caratterizza la dispersione dei valori ragionevolmente attribuibili al misurando...*". Il procedimento di stima dell'incertezza è descritto in [5] e [6] (nel caso più semplice di grandezze tra loro non correlate, l'incertezza è stimata combinando le incertezze sulle grandezze di ingresso in base alla legge della propagazione delle incertezze).

La nomenclatura dell'approccio classico all'analisi dell'errore e quella dell'approccio basato sull'incertezza possono essere messe a confronto con la figura che rappresenta le due grandezze indipendenti che concorrono a determinare la distribuzione gaussiana, la misura di posizione e la misura di dispersione.



In base all'approccio classico all'analisi dell'errore si sarebbe detto che il tiratore (dall'alto verso il basso):
- è preciso e accurato;
- è impreciso ma accurato;
- è preciso ma inaccurato;
- è impreciso e inaccurato.

In base all'approccio basato sull'incertezza si dice (dall'alto verso il basso) che il tiratore:
- è accurato [in quanto tale è ogni singolo colpo];
- non è accurato [in quanto i colpi mancano di precisione];
- non è accurato [in quanto i colpi mancano di esattezza];
- non è accurato [in quanto i colpi mancano sia di precisione sia di esattezza].

E' interessante notare come l'indipendenza tra errore sistematico ed errore casuale possa essere collegata al concetto di rapporto segnale/rumore, come illustrato nel seguente esempio. A un paziente sono effettuati due prelievi di sangue, uno al tempo t' e uno al tempo t", ed è determinata in entrambi la concentrazione di uno specifico analita nel siero. L'obiettivo è monitorare l'efficacia di una terapia: pertanto quello che interessa è stabilire se la concentrazione dell'analita nel siero al tempo t" differisca significativamente da quella rilevata al tempo t'. I due campioni sono suddivisi ciascuno in due aliquote, che sono inviate contemporaneamente in due laboratori diversi (laboratorio A e laboratorio B), che ottengono risultati diversi. Le concentrazioni rilevate nel laboratorio A sono molto vicine al valore vero della concentrazione dell'analita nel siero (il valore di concentrazione che si avrebbe se un laboratorio fosse in grado di effettuare le misure dell'analita senza errore), mentre le concentrazioni rilevate nel laboratorio B sono molto lontane al valore vero

---

$$s^2 = \Sigma (x_i - \bar{x})^2 / (n - 1)$$

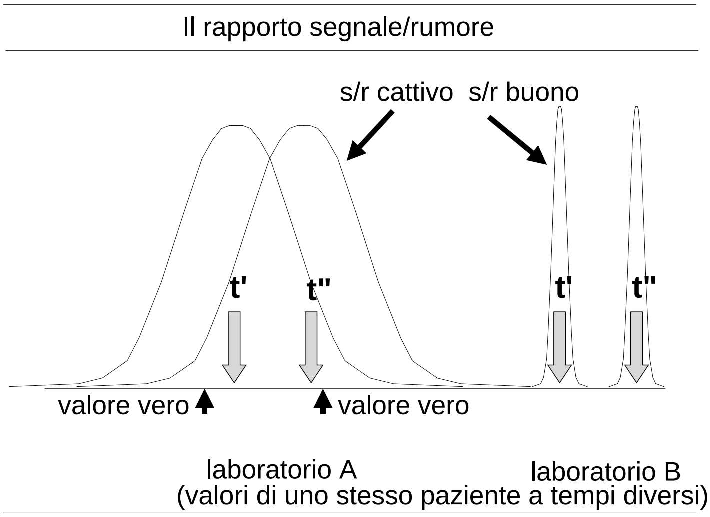
mentre infine la deviazione standard campionaria s è calcolata come

$$s = \sqrt{s^2}$$

Si ricorda che è possibile semplificare il calcolo della varianza s^2 utilizzando per la devianza l'espressione alternativa

$$\Sigma (x_i - \bar{x})^2 = \Sigma x_i^2 - (\Sigma x_i)^2 / n$$

della concentrazione dell'analita. Per converso le misure effettuate nel laboratorio A sono affette da un errore casuale superiore (distribuzione dei valori più allargata) rispetto a qualel effettuate nel laboratorio B. Definiamo ora come segnale la differenza tra la concentrazione dell'analita al tempo t' e quella al tempo t", e come rumore l'incertezza (l'errore casuale) che caratterizza ciascuna misura. Il laboratorio A fornisce risultati che sono affetti da un rapporto segnale/rumore peggiore (più basso) di quello del laboratorio B: l'errore casuale che caratterizza i risultati delle due determinazioni appare di tale entità che le distribuzioni da cui essi sono tratti sono praticamente quasi indistinguibili. Diversa è la situazione del laboratorio B, per il quale il rapporto segnale/rumore è migliore (più alto) di quello del laboratorio A: l'errore casuale che caratterizza i risultati delle due determinazioni appare contenuto, tanto che la concentrazione al tempo t" può essere agevolmente riconosciuta come "diversa" da quella presente al tempo t'. Questo esempio, illustrando come l'errore sistematico possa essere considerato secondario quando ciò che interessa è riconoscere la differenza tra due valori, spiega tra l'altro il paradosso per cui il raggiungimento dell'esattezza dei risultati nel laboratorio clinico (identici risultati ottenuti con metodi diversi nei diversi laboratori) abbia talora finito per essere considerato meno "urgente" del raggiungimento della precisione. E conferma la razionalità, nel caso del monitoraggio, dell'esecuzione delle analisi sempre nello stesso laboratorio.



Bibliografia e riferimenti

[1] Baldini M. L'errore nella scienza. Biochim Clin 1991;15:28-38.
[2] Taylor J. R. Introduzione all'analisi degli errori. Zanichelli, 1986.
[3] Snedecor G. W., Cochran W. G. Statistical Methods. VII Edition. Ames:The Iowa State University Press, 1980.
[4] Patriarca M., Menditto A., Bettinelli M., Minoia C. Stima dell'incertezza di misura nel laboratorio clinico e in medicina ambientale, occupazionale e preventiva. G Ital Med Lav Erg 2004; 26:2, 102-107
[5] UNI CEI ENV 13005:2000. Guida all'espressione dell'incertezza di misura. UNI, 2000[79].
[6] Quantificazione dell'incertezza nelle misure analitiche. Seconda edizione (2000) della Guida EURACHEM / CITAC CG 4. Traduzione italiana a cura di Marina Patriarca, Ferdinando Chiodo, Federica Corsetti, Barbara Rossi, Antonio Menditto, Michela Sega e Margherita Plassa (http://www.iss.it/publ/rapp/cont.php?id=317&tipo=5&lang=1&ccss=000).

[79] La norma UNI CEI ENV 13005:2000 è la traduzione della guida ISO nota come GUM (Guide to the expression of uncertainty of measurement), che è stata recepita come norma europea sperimentale.

APPENDICE C
1000 VALORI DI TRIGLICERIDI

117	92	149	76	86	203	82	50	128	153	79	111	69	87	83	110	114	132	216	151
97	66	221	100	141	102	88	123	59	75	80	205	51	146	86	353	48	180	74	159
75	87	322	79	121	246	211	272	104	81	147	155	79	98	136	179	107	292	115	125
137	104	80	85	200	61	162	161	66	182	176	260	148	231	178	138	63	162	48	125
82	77	71	76	43	209	371	266	125	102	96	122	76	276	104	639	271	741	97	68
92	51	100	139	66	67	431	65	120	121	111	320	110	238	181	73	139	279	65	140
185	62	51	68	52	51	82	70	141	62	95	213	78	201	75	102	135	234	72	107
170	45	136	97	102	140	30	156	102	71	75	90	78	163	57	82	105	216	315	254
127	211	110	98	181	120	108	249	57	108	105	58	154	139	94	38	74	57	182	73
182	100	87	80	135	137	66	64	40	90	115	123	90	260	85	61	77	63	101	54
142	127	59	65	248	108	138	92	77	175	180	126	99	129	114	130	96	92	76	57
81	71	58	90	74	75	77	232	198	114	38	71	40	326	36	164	98	167	101	156
96	56	118	183	39	144	152	519	53	94	125	184	166	146	79	67	104	145	77	77
83	59	61	903	40	268	133	321	47	64	99	56	59	56	80	112	130	226	129	186
119	55	73	185	145	269	153	107	92	123	118	171	100	219	137	164	91	74	63	128
144	101	147	205	141	219	183	64	92	186	97	168	143	169	65	127	113	61	149	86
87	446	79	95	289	85	73	59	81	73	213	179	78	109	144	129	83	612	77	56
47	64	278	215	92	94	47	102	140	111	126	89	86	120	108	57	217	99	176	127
81	100	82	153	61	60	475	333	32	76	72	201	106	135	99	110	88	263	75	118
215	54	165	81	102	115	167	167	164	105	108	102	116	67	149	131	73	63	39	87
120	105	91	63	184	126	355	188	101	241	30	226	80	99	183	71	61	82	116	258
67	84	56	115	48	159	74	57	104	239	227	186	353	113	90	99	97	165	347	75
67	186	64	157	162	126	94	153	136	98	102	138	54	202	33	92	37	110	50	218
87	71	77	266	98	242	74	180	222	332	88	175	164	75	120	106	72	226	115	131
81	97	48	115	122	159	112	118	148	61	119	193	96	100	147	99	100	100	46	90
190	74	93	107	105	61	39	80	126	108	58	177	82	209	116	141	99	383	52	216
191	59	103	121	120	217	99	80	99	190	164	169	143	661	216	115	147	218	129	212
35	324	112	114	117	181	164	121	182	57	90	107	108	144	58	95	57	154	194	72
152	128	42	78	77	210	44	156	74	86	85	136	50	165	96	147	92	113	69	66
170	207	291	218	85	89	158	135	94	48	129	131	301	60	128	152	121	428	78	145
116	131	98	273	49	307	53	151	85	156	64	184	69	144	81	132	196	120	53	290
87	78	210	125	54	59	89	309	396	443	186	84	160	150	60	55	67	153	66	346
77	91	211	149	59	223	114	63	48	449	146	508	129	150	59	421	107	118	91	382
75	58	277	89	88	164	165	78	128	65	66	96	111	111	57	70	149	104	90	101
81	118	113	136	159	419	100	172	191	137	78	223	298	168	137	156	99	139	82	169
78	131	83	113	159	164	166	97	209	164	83	292	77	281	113	95	236	272	98	125
78	87	48	214	224	95	69	82	101	128	180	112	70	47	69	138	100	188	109	625
303	92	130	102	152	82	60	135	95	155	104	80	68	205	137	231	107	39	67	107
102	102	80	99	71	78	185	349	165	117	226	147	106	175	54	271	95	131	153	67
50	115	83	66	85	88	45	61	64	112	67	160	129	369	95	229	102	166	64	171
106	57	129	198	70	201	109	90	242	172	67	304	118	164	85	244	67	175	199	163
93	54	161	85	105	246	128	57	80	561	111	79	275	186	360	171	120	150	44	253
112	71	264	223	122	80	105	144	114	351	102	94	122	475	64	104	120	168	101	130
80	41	178	156	184	92	102	205	131	76	182	127	78	132	114	124	45	132	70	109
68	227	123	221	85	871	151	179	76	69	110	61	46	60	91	266	69	256	118	133

46	205	127	432	42	119	112	61	187	192	88	68	74	30	48	126	74	144	150	131
132	136	404	225	64	172	68	121	81	355	152	113	121	124	156	126	53	207	96	92
107	370	150	80	65	142	163	92	182	82	71	172	140	64	123	87	40	113	118	80
212	206	351	258	55	84	83	145	176	247	273	231	55	230	211	68	146	74	65	141
65	206	83	97	139	72	143	79	167	301	241	157	130	356	113	159	68	74	166	399

APPENDICE D
ALGORITMI PER IL CALCOLO DELLA REGRESSIONE LINEARE

Per adattare la retta ai dati sperimentali viene impiegato il metodo dei minimi quadrati, una tecnica di approssimazione ben nota, che consente di minimizzare la somma dei quadrati delle differenze che residuano fra i punti sperimentali e la retta.

Regressione lineare x variabile indipendente

Il modello matematico impiegato presuppone che la *x*, cioè la variabile indipendente, sia misurata senza errore, e che l'errore con cui si misura la *y* sia distribuito normalmente e con una varianza che non cambia al variare del valore della *x*.

Sia allora *n* il numero dei punti aventi coordinate note (x_i,y_i), e siano $\bar{x}$ la media dei valori delle x_i e $\bar{y}$ la media dei valori delle y_i: il coefficiente angolare b_{yx} e l'intercetta a_{yx} dell'equazione della retta di regressione *x* variabile indipendente

$$y = a_{yx} + b_{yx} \cdot x$$

che meglio approssima (in termini di minimi quadrati) i dati vengono calcolati rispettivamente come

$$b_{yx} = \Sigma (x_i - \bar{x}) \cdot (y_i - \bar{y}) / \Sigma (x_i - \bar{x})2$$

$$a_{yx} = \bar{y} - b_{yx} \cdot \bar{x}$$

Il coefficiente di correlazione *r* viene calcolato come

$$r = \Sigma (x_i - \bar{x}) \cdot (y_i - \bar{y}) / \sqrt{}(\Sigma (x_i - \bar{x})^2 \cdot \Sigma (y_i - \bar{y})^2)$$

E' possibile semplificare i calcoli ricordando che

$$\Sigma (x_i - \bar{x})^2 = \Sigma x_i^2 - (\Sigma x_i)^2 / n$$
$$\Sigma (y_i - \bar{y})^2 = \Sigma y_i^2 - (\Sigma y_i)^2 / n$$
$$\Sigma (x_i - \bar{x}) \cdot (y_i - \bar{y}) = \Sigma x_i \cdot y_i - (\Sigma x_i) \cdot (\Sigma y_i) / n$$

La varianza residua attorno alla regressione viene calcolata come

$$s_0^2 = (\Sigma (y_i - \bar{y})^2 - s_1^2) / (n - 2)$$

essendo

$$s_1^2 = (\Sigma (x_i - \bar{x}) \cdot (y_i - \bar{y}))^2 / \Sigma (x_i - \bar{x})^2$$

Infine l'errore standard della stima s_{yx} e le deviazioni standard del coefficiente angolare (s_b) e dell'intercetta (s_a), che forniscono una misura rispettivamente della dispersione dei dati attorno alla retta calcolata, e del grado di incertezza connesso con i valori ottenuti di a_{yx} e di b_{yx}, sono calcolati come

$$s_{yx} = \sqrt{(s_0^2)}$$
$$s_b = s_{yx} \cdot \sqrt{(1 / \Sigma (x_i - \bar{x})^2)}$$
$$s_a = s_b \cdot \sqrt{(\Sigma x_i^2 / n)}$$

Si consideri che la retta di regressione campionaria

$$y = a_{yx} + b_{yx} \cdot x$$

rappresenta la migliore stima possibile della retta di regressione della popolazione

$$y = \alpha + \beta \cdot x$$

Si consideri che il test t di Student per una media teorica nella forma già vista

$$t = (\bar{x} - \mu) / \sqrt{(s^2 / n)}$$

può essere riscritto tenendo conto delle seguenti identità

$$\bar{x} = a_{yx}$$
$$\mu = \alpha$$
$$\sqrt{(s^2 / n)} = s_a$$

assumendo quindi la forma

$$t = (a_{yx} - \alpha) / s_a$$

Questo consente di sottoporre a test la differenza dell'intercetta a rispetto a un valore atteso (per esempio rispetto a 0, cioè all'intercetta di una retta passante per l'origine). Il valore di p corrispondente alla statistica t rappresenta la probabilità di osservare per caso una differenza della grandezza di quella effettivamente osservata: se tale probabilità è sufficientemente piccola, si conclude per una significatività della differenza dell'intercetta rispetto al valore atteso.

Si consideri che il test t di Student per una media teorica può anche essere riscritto tenendo conto delle seguenti identità

$$\bar{x} = b_{yx}$$
$$\mu = \beta$$
$$\sqrt{(s^2 / n)} = s_b$$

assumendo quindi la forma

$$t = (b_{yx} - \beta) / s_b$$

Questo consente di sottoporre a test la differenza del coefficiente angolare b rispetto a un valore atteso (per esempio rispetto a 0, cioè al coefficiente angolare di una retta orizzontale, oppure rispetto a 1, cioè al coefficiente angolare di una retta a 45 gradi). Il valore di p corrispondente alla statistica t rappresenta la probabilità di osservare per caso una differenza della grandezza di quella effettivamente osservata: se tale probabilità è sufficientemente piccola, si conclude per una significatività della differenza del coefficiente angolare rispetto al valore atteso.

Regressione lineare y variabile indipendente

Il modello matematico impiegato presuppone che la y, cioè la variabile indipendente, sia misurata senza errore, e che l'errore con cui si misura la x sia distribuito normalmente e con una varianza che non cambia al variare del valore della y. Si noti che in questo caso inizialmente la y (variabile indipendente) viene posta in ascisse e la x (variabile dipendente) viene posta in ordinate.

Sia allora n il numero dei punti aventi coordinate note (x_i,y_i), e siano $\bar{x}$ la media dei valori delle x_i e $\bar{y}$ la media dei valori delle y_i: il coefficiente angolare b_{xy} e l'intercetta a_{xy} dell'equazione della retta di regressione y variabile indipendente

$$x = a_{xy} + b_{xy} \cdot y$$

che meglio approssima (in termini di minimi quadrati) i dati vengono calcolati rispettivamente come

$$b_{xy} = \Sigma (x_i - \bar{x}) \cdot (y_i - \bar{y}) / \Sigma (y_i - \bar{y})^2$$

$$a_{xy} = \bar{x} - b_{xy} \cdot \bar{y}$$

Il coefficiente di correlazione r viene calcolato come

$$r = \Sigma (x_i - \bar{x}) \cdot (y_i - \bar{y}) / \sqrt{(\Sigma (x_i - \bar{x})^2 \cdot \Sigma (y_i - \bar{y})^2)}$$

(si noti che, come atteso, esso risulta identico a quello calcolato mediante la regressione x variabile indipendente).

E' possibile semplificare i calcoli ricordando che

$$\Sigma (x_i - \bar{x})^2 = \Sigma x_i^2 - (\Sigma x_i)^2 / n$$
$$\Sigma (y_i - \bar{y})^2 = \Sigma y_i^2 - (\Sigma y_i)^2 / n$$
$$\Sigma (x_i - \bar{x}) \cdot (y_i - \bar{y}) = \Sigma x_i \cdot y_i - (\Sigma x_i) \cdot (\Sigma y_i) / n$$

Per riportare i dati sullo stesso sistema di coordinate cartesiane utilizzato per la regressione x variabile indipendente, si esplicita l'equazione della retta di regressione y variabile indipendente

$$x = a_{xy} + b_{xy} \cdot y$$

rispetto alla y, ottenendo

$$x - a_{xy} = b_{xy} \cdot y$$

e quindi, dividendo entrambi i membri per b_{xy}

$$y = - a_{xy} / b_{xy} + 1 / b_{xy} \cdot x$$

Quindi l'intercetta a e il coefficiente angolare b che consentono di rappresentare la regressione y variabile indipendente sullo stesso sistema di coordinate cartesiane della regressione x variabile indipendente saranno rispettivamente uguali a

$$a = - a_{xy} / b_{xy}$$
$$b = 1 / b_{xy}$$

Componente principale standardizzata

Il modello matematico impiegato presuppone tanto la x quanto la y siano affette da un errore di misura equivalente.

Sia allora n il numero dei punti aventi coordinate note (x_i,y_i), siano $\bar{x}$ la media dei valori delle x_i e $\bar{y}$ la media dei valori delle y_i, sia b_{yx} il coefficiente angolare dell'equazione della retta di regressione x variabile indipendente, e sia b_{xy} il coefficiente angolare dell'equazione della retta di regressione y variabile indipendente.

Il coefficiente angolare b_{cps} dell'equazione della retta di regressione calcolata come componente principale standardizzata è allora uguale a

$$b_{cps} = \sqrt{(b_{yx} \cdot b_{xy})}$$

cioè alla media geometrica tra il coefficiente angolare b_{yx} della regressione x variabile indipendente e il coefficiente angolare b_{xy} della regressione y variabile indipendente, cioè

$$b_{xy} = \sqrt{(\Sigma (y_i - \bar{y})^2 / \Sigma (x_i - \bar{x})^2)}$$

mentre l'intercetta a_{cps} dell'equazione della retta di regressione calcolata come componente principale standardizzata è uguale a

$$a_{cps} = \bar{y} - b_{cps} \cdot \bar{x}$$

Infine il coefficiente di correlazione r viene calcolato come

$$r = \Sigma (x_i - \bar{x}) \cdot (y_i - \bar{y}) / \sqrt{(\Sigma (x_i - \bar{x})^2 \cdot \Sigma (y_i - \bar{y})^2)}$$

(si noti che, come atteso, esso risulta identico sia a quello calcolato mediante la regressione x variabile indipendente sia a quello calcolato mediante la regressione y variabile indipendente).

E' possibile semplificare i calcoli ricordando che

$$\Sigma (x_i - \bar{x})^2 = \Sigma x_i^2 - (\Sigma x_i)^2 / n$$
$$\Sigma (y_i - \bar{y})^2 = \Sigma y_i^2 - (\Sigma y_i)^2 / n$$
$$\Sigma (x_i - \bar{x}) \cdot (y_i - \bar{y}) = \Sigma x_i \cdot y_i - (\Sigma x_i) \cdot (\Sigma y_i) / n$$

APPENDICE E
DIFFERENZE CRITICHE [1]

Le differenze critiche consentono di calcolare in modo rigoroso la differenza minima che due risultati in tempi successivi della stessa analisi devono avere perché si possa concludere che i risultati sono tra loro significativamente diversi.

Le differenze critiche possono essere calcolate facilmente, conoscendo la variabilità analitica e la variabilità biologica che caratterizzano un dato analita, come

$$\textit{Differenza critica} = 2{,}77 \cdot (CV_a^2 + CV_b^2)^{1/2}$$

Conoscendo il valore della variabilità analitica CV_a di un dato metodo analitico e il valore della variabilità biologica CV_b dell'analita in questione, è quindi elementare calcolare il valore della differenza critica corrispondente. Se la differenza tra due valori consecutivamente osservati in un dato soggetto eccede il valore della differenza critica, si è autorizzati a ritenere, con una probabilità del solo 5% di sbagliare (è questo il livello di confidenza statistica qui impiegato) che i due valori differiscono significativamente tra loro. I valori delle differenze critiche dei principali analiti sono qui calcolati assumendo che la variabilità analitica sia la metà della variabilità biologica. Questo assunto fa sì che, al termine del procedimento analitico, la variabilità totale osservata per lo specifico risultato analitico sia solamente l'11,8% in più di quella che si avrebbe in presenza della sola variabilità biologica: o, in altri termini, che la variabilità analitica sia praticamente trascurabile rispetto alla variabilità biologica.

Nella tabella che segue sono riportati i valori delle differenze critiche dei principali analiti. Dopo la tabella, sono riportate alcune note interpretative relative a questo importante argomento.

Analita	***Variabilità biologica %***	***Variabilità analitica %***	***Differenza critica %***
17-ALFA IDROSSIPROGESTERONE	15,0	7,5	46
ALANINA AMMINOTRANSFERASI (ALT,GPT)	10,0	5,0	31
ALBUMINA	2,5	1,3	8
ALDOSTERONE	15,0	7,5	46
ALFA-1-GLICOPROTEINA ACIDA	10,0	5,0	31
ALFA-AMILASI	7,5	3,8	23
ANTIGENE CARBOIDRATICO 125	10,0	5,0	31
ANTIGENE CARBOIDRATICO 15.3	7,5	3,8	23
ANTIGENE CARBOIDRATICO 19.9	15,0	7,5	46
ANTIGENE CARCINO-EMBRIONALE	12,5	6,3	39
ANTIGENE PROSTATICO SPECIFICO (PSA)	12,5	6,3	39
APOLIPOPROTEINA A-I	5,0	2,5	15
APOLIPOPROTEINA B	5,0	2,5	15
APTOGLOBINA	15,0	7,5	46
APTT	5,0	2,5	15
ASPARTATO AMMINOTRANSFERASI (AST, GOT)	15,0	7,5	46
BETA-2-MICROGLOBULINA	4,0	2,0	12

BICARBONATO (Idrogenocarbonato)	2,5	1,3	8
BILIRUBINA DIRETTA	15,0	7,5	46
BILIRUBINA TOTALE	12,5	6,3	39
CALCIO TOTALE	1,5	0,8	5
CERULOPLASMINA	4,0	2,0	12
CLORURO	1,5	0,8	5
COLESTEROLO HDL	5,0	2,5	15
COLESTEROLO LDL	5,0	2,5	15
COLESTEROLO TOTALE	4,0	2,0	12
COLINESTERASI	5,0	2,5	15
COMPLEMENTO/C3	5,0	2,5	15
COMPLEMENTO/C4	7,5	3,8	23
CORTISOLO	10,0	5,0	31
CREATINA CHINASI-MB	15,0	7,5	46
CREATINCHINASI	12,5	6,3	39
CREATININA	4,0	2,0	12
EMOCROMOCITOMETRICO			
Emoglobina	1,5	0,8	5
Eritrociti	1,5	0,8	5
Leucociti	7,5	3,8	23
Piastrine	5,0	2,5	15
Ematocrito	1,5	0,8	5
EMOGASANALISI			
pH	1,5	0,8	5
pCO2	2,5	1,3	8
pO2	2,5	1,3	8
EMOGLOBINA GLICATA	4,0	2,0	12
ESTRADIOLO	12,5	6,3	39
FARMACI DIGITALICI	5,0	2,5	15
FERRITINA	7,5	3,8	23
FERRO	12,5	6,3	39
FIBRINOGENO	7,5	3,8	23
FOSFATASI ALCALINA	7,5	3,8	23
FOSFATO INORGANICO	5,0	2,5	15
FRUTTOSAMMINA	4,0	2,0	12
GAMMA-GLUTAMMIL TRANSPEPTIDASI	12,5	6,3	39
GLUCOSIO	2,5	1,3	8
GLUCOSIO-6-FOSFATO DEIDROGENASI	7,5	3,8	23
IgA	7,5	3,8	23
IgG	4,0	2,0	12
IgM	7,5	3,8	23
INSULINA	10,0	5,0	31
LATTATO DEIDROGENASI	7,5	3,8	23
LIPASI	7,5	3,8	23
LIPOPROTEINA(a)	12,5	6,3	39
LITIO	5,0	2,5	15
LUTEOTROPINA (LH)	12,5	6,3	39
MAGNESIO TOTALE	2,5	1,3	8
MICROALBUMINURIA	10,0	5,0	31
OSTEOCALCINA	10,0	5,0	31
POTASSIO	2,5	1,3	8
PROLATTINA	10,0	5,0	31
PROTEINA C REATTIVA	7,5	3,8	23
PROTEINE (ELETTROFORESI)			

Albumina	2,5	1,3	8
Alfa-1-globuline	7,5	3,8	23
Alfa-2-globuline	5,0	2,5	15
Beta-globuline	5,0	2,5	15
Gamma-globuline	5,0	2,5	15
PROTEINE TOTALI	2,5	1,3	8
RAME	4,0	2,0	12
SODIO	1,5	0,8	5
TEMPO DI PROTROMBINA	4,0	2,0	12
TEOFILLINA	10,0	5,0	31
TESTOSTERONE	10,0	5,0	31
TIREOTROPINA (TSH)	10,0	5,0	31
TIROXINA LIBERA (fT4)	7,5	3,8	23
TRANSFERRINA	4,0	2,0	12
TRIGLICERIDI	15,0	7,5	46
TRIIODOTIRONINA LIBERA (fT3)	7,5	3,8	23
URATO	7,5	3,8	23
UREA	7,5	3,8	23
VES (1 ora)	12,5	6,3	39

Si consideri il caso di un paziente al quale una prima volta sia stata rilevata una concentrazione del colesterolo nel siero pari a 243 mg/dL. Il paziente, dopo un periodo di dieta di opportuna durata, effettua una seconda determinazione del colesterolo, che risulta ora pari a 225 mg/dL. La differenza tra la prima e la seconda determinazione è uguale a 18 mg/dL. La differenza critica riportata nella tabella precedente è pari al 12%. Il 12% di 243 mg/dL è uguale a 29,16 mg/dL (arrotondato a 29 mg/dL).

Conclusione: la differenza osservata (18 mg/dL) non è significativa, in quanto risulta inferiore alla differenza critica (29 mg/dL). Per un paziente con una concentrazione iniziale di 243 mg/dL, il successivo valore avrebbe dovuto essere pari o inferiore a 214 mg/dL (243 - 29) per consentire di concludere, con un adeguato grado di confidenza (solo il 5% di probabilità di sbagliare nel trarre la conclusione), che la dieta è stata efficace nel ridurre la concentrazione del colesterolo nel siero.

Simmetricamente, con una concentrazione iniziale di 243 mg/dL, il successivo valore avrebbe dovuto essere pari o superiore a 272 mg/dL (243 - 29) per consentire di concludere, con un adeguato grado di confidenza, che la dieta ha indotto piuttosto un aumento della concentrazione del colesterolo nel siero (come può avvenire veramente: esiste un meccanismo di retroazione negativa sulla sintesi endogena del colesterolo mediato appunto dal colesterolo di origine alimentare, e la ridotta introduzione di colesterolo con gli alimenti può pertanto determinare un'attivazione della sua sintesi endogena, con un conseguente, ancorché indesiderato aumento della concentrazione del colesterolo nel siero).

Al fine di consentire un'adeguata familiarizzazione con le differenze critiche, la tabella che segue contiene alcuni ulteriori esempi. Nella prima colonna viene riportato l'analita, nella seconda colonna il valore della differenza critica corrispondente, nella terza colonna un possibile obiettivo clinico del monitoraggio, nella quarta colonna il valore (ipotetico) riscontrato alla prima osservazione, e nella quinta il valore al quale (e oltre il quale) si deve considerare significativa la differenza alla seconda osservazione, con un rischio minimo di sbagliare (solo il 5%) accettando la conclusione che la concentrazione dell'analita si è effettivamente modificata. Alla seconda osservazione, qualsiasi valore compreso tra il primo ed il secondo (colonne (a) e (b) rispettivamente) deve essere attribuito al caso, impersonato nella fattispecie dalla variabilità analitica (ridotta) e dalla variabilità biologica (prevalente).

Un'ultima chiave di lettura della tabella. Un paziente che, nel corso di indagini multiple successive, si presentasse con valori che oscillano, ma sempre compresi tra quello della colonna (a) e quello della colonna (b), deve essere visto come un paziente nel quale non vi sono cambiamenti significativi: quello che si sta osservando è in sostanza la sua variabilità biologica.

Analita	***Differenza critica %***	***Obiettivo clinico***	***Valore (ipotetico) riscontrato alla prima osservazione (a)***	***Valore al quale (oltre il quale) considerare significativa la differenza alla seconda osservazione (b)***
ALANINA AMMINOTRANSFERASI (ALT,GPT)	31	Valutare la riduzione della concentrazione dell'enzima	120 U/L	83 U/L
ANTIGENE PROSTATA SPECIFICO (PSA)	39	Valutare il possibile aumento in un soggetto con problemi specifici in atto	10,0 μg/L	13,9 μg/L
CREATININA	12	Monitorare l'eventuale innalzamento in paziente con insufficienza renale	8,5 mg/dL	9,5 mg/dL
EMOCROMOCITOMETRICO Emoglobina	5	Valutare la correzione dell'anemia	10,8 g/dL	11,3 g/dL
FERRO	39	Monitorare la sideremia in un donatore abituale di sangue	86 μg/L	53 μg/L
GLUCOSIO	8	Valutare il miglioramento della glicemia	180 mg/dL	166 mg/dL
MICROALBUMINURIA	31	Valutare il peggioramento della funzione renale in un paziente diabetico	40 mg/L	52 mg/L
VES (1 ora)	39	Monitorare la scomparsa della flogosi	36 mm	22 mm

Per gli analiti non riportati nelle tabelle precedenti è agevole, conoscendone le rispettive variabilità analitica e biologica, calcolare le differenze critiche mediante la formula riportata all'inizio, ovvero con la tabella seguente, nella quale all'incrocio tra la variabilità biologica e la variabilità analitica osservate compare la differenza critica corrispondente.

	CVa																		
CVb(intra)	1	1.5	2	2.5	3	3.5	4	4.5	5	5.5	6	6.5	7	7.5	8	8.5	9	9.5	10
1	3.9	5.0	6.2	7.5	8.8	10.1	11.4	12.8	14.1	15.5	16.8	18.2	19.6	21.0	22.3	23.7	25.1	26.5	27.8
2	6.2	6.9	7.8	8.9	10.0	11.2	12.4	13.6	14.9	16.2	17.5	18.8	20.2	21.5	22.8	24.2	25.5	26.9	28.2
3	8.8	9.3	10.0	10.8	11.8	12.8	13.9	15.0	16.2	17.4	18.6	19.8	21.1	22.4	23.7	25.0	26.3	27.6	28.9
4	11.4	11.8	12.4	13.1	13.9	14.7	15.7	16.7	17.7	18.8	20.0	21.1	22.3	23.5	24.8	26.0	27.3	28.6	29.8
5	14.1	14.5	14.9	15.5	16.2	16.9	17.7	18.6	19.6	20.6	21.6	22.7	23.8	25.0	26.1	27.3	28.5	29.7	31.0
6	16.8	17.1	17.5	18.0	18.6	19.2	20.0	20.8	21.6	22.5	23.5	24.5	25.5	26.6	27.7	28.8	30.0	31.1	32.3
7	19.6	19.8	20.2	20.6	21.1	21.7	22.3	23.1	23.8	24.7	25.5	26.5	27.4	28.4	29.4	30.5	31.6	32.7	33.8
8	22.3	22.5	22.8	23.2	23.7	24.2	24.8	25.4	26.1	26.9	27.7	28.6	29.4	30.4	31.3	32.3	33.4	34.4	35.5
9	25.1	25.3	25.5	25.9	26.3	26.7	27.3	27.9	28.5	29.2	30.0	30.8	31.6	32.5	33.4	34.3	35.3	36.2	37.3
10	27.8	28.0	28.2	28.6	28.9	29.3	29.8	30.4	31.0	31.6	32.3	33.0	33.8	34.6	35.5	36.4	37.3	38.2	39.2
11	30.6	30.8	31.0	31.2	31.6	32.0	32.4	32.9	33.5	34.1	34.7	35.4	36.1	36.9	37.7	38.5	39.4	40.3	41.2
12	33.4	33.5	33.7	34.0	34.3	34.6	35.0	35.5	36.0	36.6	37.2	37.8	38.5	39.2	39.9	40.7	41.6	42.4	43.3
13	36.1	36.2	36.4	36.7	37.0	37.3	37.7	38.1	38.6	39.1	39.7	40.3	40.9	41.6	42.3	43.0	43.8	44.6	45.4
14	38.9	39.0	39.2	39.4	39.7	40.0	40.3	40.7	41.2	41.7	42.2	42.8	43.4	44.0	44.7	45.4	46.1	46.9	47.7
15	41.6	41.8	41.9	42.1	42.4	42.7	43.0	43.4	43.8	44.3	44.8	45.3	45.9	46.5	47.1	47.8	48.5	49.2	49.9
16	44.4	44.5	44.7	44.9	45.1	45.4	45.7	46.0	46.4	46.9	47.3	47.8	48.4	48.9	49.6	50.2	50.9	51.5	52.3
17	47.2	47.3	47.4	47.6	47.8	48.1	48.4	48.7	49.1	49.5	49.9	50.4	50.9	51.5	52.0	52.6	53.3	53.9	54.6
18	49.9	50.0	50.2	50.3	50.5	50.8	51.1	51.4	51.7	52.1	52.6	53.0	53.5	54.0	54.6	55.1	55.7	56.4	57.0
19	52.7	52.8	52.9	53.1	53.3	53.5	53.8	54.1	54.4	54.8	55.2	55.6	56.1	56.6	57.1	57.7	58.2	58.8	59.5
20	55.5	55.6	55.7	55.8	56.0	56.2	56.5	56.8	57.1	57.5	57.8	58.3	58.7	59.2	59.7	60.2	60.8	61.3	61.9

Bibliografia e riferimenti

[1] Besozzi M. Dal laboratorio alla clinica: le differenze critiche. In: Cabrini E, Ottomano C, Pagano A. Argomenti di medicina di laboratorio per il medico pratico. Volume II. Edizioni Recordati, Milano, 1998.