

Il teorema di Bayes: una introduzione critica

Roma, dicembre 2002

Paolo Agnoli

Premessa

Questo breve scritto è stato redatto come tesina per l'esame di *Psicologia Generale* presso l'Università di Roma 2 Tor Vergata nell'anno accademico 2002-2003 (prof. Vetrone). Lo studio si propone di introdurre i fondamenti filosofici della cosiddetta statistica Bayesiana e discuterne le implicazioni riguardanti l'interpretazione dell'attività scientifica.

1. Introduzione

Se volete apprendere qualcosa [dagli scienziati] sui metodi che essi usano, vi consiglio di tener fermo un solo principio: non ascoltate le loro parole in proposito ma fissate la vostra attenzione su ciò che fanno.

A. Einstein

*An effect may be produced by the cause a or by the cause b
The effect has just been observed. We ask the probability that
it is due to the cause a. This is an à posteriori
probability of cause. But I could not calculate it, if
a convention more or less justified did not tell me in advance
what is the à priori probability for the cause a
to come into play. I mean the probability of this event to
some one who had not observed the effect.*

Henri Poincaré

I nuovi sviluppi della teoria della misura sono basati sul concetto di probabilità soggettiva e sul *teorema di Bayes*. Per introdurre quest'ultimo cercherò di evitare il più possibile i formalismi, nondimeno dovremo familiarizzare con un minimo di terminologia sulle variabili casuali (nel linguaggio della probabilità soggettiva).

Una *variabile casuale*, o *numero aleatorio*, è qualsiasi numero rispetto al quale si è in stato di incertezza. Facciamo due esempi nel contesto delle misure:

1. Pongo un chilogrammo campione su una bilancia di laboratorio con indicazione (digitale) dei centesimi. Che valore leggerò (in grammi)? 1000.00, 999.95, 1000.03 ...?
2. Leggo su una bilancia di laboratorio 2.315 g. Quanto vale il valore vero della massa del corpo? 2.311, 2.312, ... 2.315, ... 2.319, ...?

Nel primo caso la variabile è la lettura x (subordinatamente ad un certo valore vero); nel secondo caso la variabile è il valore vero, che ora indicherò con μ (subordinatamente ad un certo valore letto).

Ai possibili valori della grandezza viene associata una funzione $f(x)$ che quantifica il grado di fiducia ad essi assegnato. Quindi scrivere che $f(x_1) > f(x_2)$ sta ad indicare che si crede più a x_1 che a x_2 . A seconda che la variabile x sia *discreta* o *continua*, $f(x)$ ha l'accezione di *funzione di probabilità* o di *funzione densità di probabilità*.

Tutte le proprietà di $f(x)$ apprese nei corsi convenzionali (basati cioè sui concetti convenzionali di probabilità) rimangono valide nell'approccio soggettivista. In particolare si ricorda che la deviazione standard σ fornisce la dispersione di valori che è possibile attendersi dalla variabile. Tutte le distribuzioni di variabile casuale sono subordinate ad un certo stato di informazione. Utilizzando i due esempi precedenti possiamo perciò scrivere

$$\begin{aligned} f(x) &\rightarrow f(x|\mu = 1000.00) \\ f(\mu) &\rightarrow f(\mu|x = 2.315), \end{aligned}$$

ove "|" si legge "dato", "subordinatamente a", etc.

L'intero stato di incertezza sui valori della grandezza di interesse è espresso da $f(\mu)$. Da questa funzione è possibile calcolare la probabilità che la grandezza abbia un valore compreso in un certo intervallo.

Ogni misura è eseguita allo scopo di accrescere la conoscenza di chi la esegue, si tratti di uno scienziato, un ingegnere che deve valutare/collaudare un prodotto o un apparato, un medico che ha prescritto una certa analisi. E' anche chiaro che la necessità stessa di eseguire misure indica che ci si trovava in uno stato di incertezza su qualcosa di interesse. Questo "qualcosa" può essere una *costante fisica o una teoria sull'origine dell'universo, l'affidabilità di una apparecchiatura per trasmissione dati, lo stato di salute di un paziente, etc...* In tutti i casi la misura ha lo scopo di *modificare* un certo stato di conoscenza.

Si sarebbe tentati di dire addirittura "acquisire", anziché "modificare", lo stato di conoscenza, come ad indicare che la conoscenza possa essere creata dal nulla nell'atto della misura. Non è difficile convincersi, invece, che nella maggior parte dei casi si tratta invece soltanto di un aggiornamento alla luce di fatti nuovi e di un certo raziocinio. Prendiamo ad esempio la misura della temperatura di una stanza, effettuata con un termometro digitale tanto per escludere contributi soggettivi alla lettura dello strumento - e supponiamo di ottenere 21.7°C. Anche se si potrà dubitare del decimo di grado, indubbiamente la misura è servita a restringere l'intervallo di temperature ritenute plausibili prima della misura - quelle compatibili con la sensazione di "ambiente confortevole". In base alla conoscenza del termometro usato, o dei termometri in generale, ci saranno valori di temperatura in un certo intervallo intorno a 21.7°C ai quali crediamo di più e valori al di fuori ai quali crediamo di meno.

E' però altresì chiaro che se il termometro avesse indicato, a parità, di sensazione fisiologica, 17.3°C si sarebbe tentati a ritenere che esso non funzioni bene. Non si avrebbero invece dubbi sul suo malfunzionamento se avesse indicato 2.5°C!

I tre casi corrispondono a tre diversi gradi di aggiornamento della conoscenza. Nell'ultimo caso, in particolare, l'aggiornamento è nullo. (D'Agostini, Gennaio 1999, 32)

Molto spesso si pensa che l'unico metodo scientifico valido sia quello della falsificazione. Ma non è sempre così.

Non ci sono dubbi che, se una teoria non è in grado di descrivere i risultati di un esperimento, essa vada scartata o modificata. Ma poiché non è possibile dimostrare la certezza di una teoria, diventa impossibile decidere fra tutte le (infinite) ipotesi non falsificate. Il metodo probabilistico permette di fornire una scala di credibilità a tutte le ipotesi considerate (o rapporti di credibilità fra ogni coppia di ipotesi). Un caso in cui il metodo di falsificazione è completamente inadeguato è quello relativo alle incertezze di misura. Infatti, prendendo alla lettera tale metodo, si sarebbe autorizzati soltanto a verificare se il valore osservato sullo strumento è compatibile o no con un valore vero, niente di più. Si capisce come, con queste premesse, non si possa fare molta strada. (Ibidem)

In generale, non si può applicare il metodo della falsificazione alle ipotesi statistiche, l'area in cui peraltro la metodologia è di maggior importanza per gli scienziati praticanti.

Many scientific theories are explicitly probabilistic and, for this reason, have no logical consequences of a verifiable character. An example is Mendel's theory of inheritance. This states the probabilities with which certain combinations of genes occur during reproduction; but, strictly speaking, the theory does not categorically rule out, nor predict, any particular genetic configuration. Nevertheless, Mendel obtained impressive confirmation from the results of his

plantgrowing trials, results which this theory did not entail but stated to be relatively probable. (Howson, Urbach, 1993, 7)

In altre parole, i principi sui quali si basa Popper sono troppo deboli per restringere sufficientemente il campo di ipotesi alternative, laddove, in pratica, gli scienziati hanno un sistema di classificazione delle ipotesi secondo la loro eleggibilità per una seria considerazione. Tale sistema è stato tradizionalmente caratterizzato come uno spettro di probabilità.

Have we any right, for instance, to enunciate Newton's Law? No doubt numerous observations are in agreement with it, but is not that a simple fact of chance? And how do we know, besides, that this law which has been true for so many generations will not be untrue in the next? To this objection the only answer you can give is: It is very improbable[...] From this point of view all the sciences would only be unconscious applications of the calculus of probabilities. And if this calculus be condemned, then the whole of the sciences must also be condemned. (Poincaré, 1952, 186; cited in Howson, Urbach, 1993, 9)

Per formalizzare il discorso intrapreso, occorre associare variabili casuali sia ai possibili valori delle grandezze fisiche che ai valori osservabili sullo strumento. Fatto ciò, si tratterà di imparare come inferire la distribuzione di probabilità del valore vero, ossia come valutare, per ogni possibile valore della grandezza, un corrispondente *grado di fiducia*.

2. Il teorema di Bayes

Our knowledge is the amassed thought and experience of innumerable minds.
Ralph Waldo Emerson

Come appena introdotto,

$f(x|\mu)$

sta per la funzione densità di probabilità (x è una variabile continua, dal punto di vista pratico) di osservare un certo valore x , dato un determinato valore vero μ . Tutti i possibili valori di μ possono essere visti come le infinite *cause* responsabili del valore x osservato (il loro *effetto*).

La funzione $f(x|\mu)$ ci dà la verosimiglianza che μ possa causare x e per questo è chiamata semplicemente *verosimiglianza*. Essa va stimata dalla conoscenza del comportamento dello strumento e, più in generale, dell'insieme di tutte le procedure di misura. Molto spesso si utilizza per la verosimiglianza un modello gaussiano, giustificato dalle aspettative teoriche basate sul teorema del limite centrale (vedi per esempio Severi, 1986).

Una volta fissata la funzione di verosimiglianza e un valore osservato x , si tratta di costruire la $f(\mu|x)$. Per arrivare in modo euristico alla formula generale, consideriamo soltanto due possibili valori di μ . Se, in base alle nostre conoscenze, riteniamo i due valori ugualmente probabili, ci sembrerà naturale protendere per il valore per il quale la verosimiglianza di osservare x è maggiore. Ad esempio, se $\mu_1 = -1$, $\mu_2 = 10$ e $x = 2$, si è tentati a credere che l'osservazione sia dovuta più verosimilmente alla causa μ_1 che alla causa μ_2 . Se però la grandezza di interesse è definita positiva, la causa μ_1 crolla da causa più probabile a causa impossibile. Ci sono poi casi intermedi in cui, per motivi legati all'esperienza precedente, si tende a credere *a priori* più ad una causa che all'altra. Ne segue che il grado di fiducia risultante di un certo valore di μ sarà proporzionale sia alla verosimiglianza che esso produca il valore osservato che al grado di fiducia che si attribuiva a μ prima dell'osservazione:

$$f(\mu|x) \propto f(x|\mu) \cdot f_0(\mu)$$

Questo è uno dei modi di scrivere il teorema di Bayes, che ha peraltro un ruolo centrale in tutte le inferenze probabilistiche (per approfondimenti su questo tipo di introduzione del teorema vedi per esempio D'Agostini, 2002 e Box, 1992). Il teorema prende il nome da un ministro della chiesa Presbiteriana Riformista che nel 1763 pubblicò un documento(*Essay toward solving a Problem in the Doctrine of Chances*) nei *Philosophical Transactions of the Royal Society* in cui ci si serve di un metodo atto ad elaborare operazioni statistiche basandosi su una prima comprensione di un fenomeno e combinando formalmente quella prima visione con dati misurati al momento in modo da aggiornare l'opinione scientifica di colui che conduce l'esperimento. Il punto centrale dell'articolo (vedi Lad, 1996) è una teorema che identifica come le prove miste di una sequenza di esperimenti possano essere utilizzate per derivare le probabilità circa il risultato del prossimo esperimento, dettagliando precisamente come le prove di 100 osservazioni possano essere riconosciute come diverse dalla prova di una singola di esse.

Questo tipo di approccio all'inferenza statistica ha quindi una storia lunga, e cominciò ad essere messo in discussione solo con l'avvento del neopositivismo per essere messo in disparte all'inizio del XX secolo.

A large international school of scientists preceded, supported, expanded, and developed Bayesian thinking about science. These included such famous scientists as James Bernoulli in 1713 (even before the paper by Bayes was published), Pierre Simon de Laplace in 1774, and many twentieth century scientists such as Bruno de Finetti, Harold Jeffreys, L.J. Savage, Dennis V. Lindley, and Arnold Zellner. Bayesian methodology was the method of statistical inference generally used from the time of Bayes until the early part of the twentieth century, when Sir Ronald A. Fisher and others introduced the *frequentist approach* to statistical inference.(Press, Tanur, 2001, 204)

Forse è di un certo interesse sapere che uno dei primi fisici del XX secolo a riflettere sull'approccio Bayesiano alle misure fu Enrico Fermi. Ricorda un suo allievo alla University of Chicago nel 1947.

I have one more personal example of how Fermi left his mark on the entire international physics community[...]most physicist were not very knowledgeable about statistical inference. In my thesis I had to find the best 3-parameter fit to my data and the errors of those parameters in order to get the 3 phase shifts and their errors. Fermi showed me a simple analytical method. At the same time other physicists were using and publishing other cumbersome methods. Also Fermi taught me a general method, which he called Bayes Theorem, where one could easily derive the best-fit parameters and their errors as a special case of the maximum-likelihood method. I remember asking Fermi how and where he learned this. I expected him to answer R.A. Fisher or some textbook on mathematical statistics. Instead he said «perhaps it was Gauss.»[...]Frank Solmitz, a fellow grad student, and I felt we should get down on paper what we were learning from Fermi. So with help from Fermi and Frank Solmitz I pulled all this together a few years later in a 1958 UCRL report called «Notes on Statistics for Physicists».(Orear, 2001, 32-33)

Dal teorema di Bayes nasce la statistica Bayesiana, *basata appunto sull'idea intuitiva che la probabilità quantifica il grado di fiducia nell'occorrenza di un evento*. E la statistica diventa allora *uno strumento logico* per aggiornare la probabilità *alla luce di tutti i differenti tipi di informazione*. Forse è utile mostrare un ragionamento intuitivo che aiuta a capire meglio il contributo delle probabilità a priori nelle inferenze.

Consideriamo un cacciatore che si aggira in un bosco con il suo cane sempre in continuo movimento intorno a lui. Supponiamo che la probabilità che il cane si trovi entro un raggio di 100 m dal cacciatore sia del 50 %. Osserviamo il cane in una certa posizione: cosa possiamo dire sulla posizione dove si trova il cacciatore? Nessuno esiterà a dire che, al 50 %, si troverà entro 100 m dal cane. Chiaramente il cacciatore sta per μ e il cane per l'osservazione x . Ma non è difficile convincersi che per arrivare in modo intuitivo a questo risultato, si sta tacitamente assumendo che il cacciatore possa essere, a priori, in ogni punto del bosco. Le cose cambiano se il cane sta costeggiando un fiume, se corre in una certa direzione con la preda in bocca o se è dentro un terreno recintato (ad esempio a oltre 100 metri dal filo spinato) in cui lui può entrare e il cacciatore no. Detto più chiaramente, si sta assumendo una *distribuzione iniziale uniforme* (del cacciatore nella foresta) e una verosimiglianza simmetrica. Ogni variazione da questo modello porta a conclusioni

differenti. Spesso, in semplici misure di routine il modello cacciatore → cane e cane → cacciatore funziona secondo l'inversione intuitiva che abbiamo descritto. Ci sono però dei problemi (specialmente in fisica) in cui questo è tutt'altro che vero.

Molti possono rimanere perplessi al pensiero che le conclusioni scientifiche possano dipendere dal "pregiudizio" sulla grandezza fisica. Ma anche se pregiudizio ha nel linguaggio corrente un significato prevalentemente negativo, in realtà significa semplicemente un giudizio basato su una esperienza acquisita in precedenza. Secondo il punto di vista Bayesiano, il ragionamento scientifico, e almeno in parte quello quotidiano, vengono condotti in termini probabilistici. In altre parole, quando una persona valuta un'affermazione incerta, lo fa calcolando la probabilità dell'ipotesi alla luce delle informazioni conosciute.

Notiamo ora, come esempio che può aiutare a comprendere in che modo *i pregiudizi siano comunque sempre presenti nelle nostre valutazioni*, che la stessa esperienza visiva che l'osservatore ha quando guarda un oggetto dipende dalla sua esperienza passata, dallo stato delle sue conoscenze e dalle sue aspettative.

Ecco due esempi che illustrano bene questo punto.

In un esperimento molto noto, a dei soggetti venivano mostrate in rapida successione delle carte da gioco e veniva loro chiesto di identificarle. Finché si utilizzava un normale mazzo di carte, tali individui erano capaci di soddisfare questa richiesta. Non appena, però, venivano introdotte carte irregolari, ad esempio un asso di picche rosso, allora sulle prime quasi tutti i soggetti identificavano da principio quelle carte irregolari come carte normali: vedevano così un asso di picche rosso come se fosse un normale asso di quadri o un normale asso di picche. Le impressioni soggettive esperite dagli osservatori erano influenzate dalle loro attese. Quando, dopo un periodo di confusione, i soggetti cominciarono a rendersi conto oppure venivano avvisati della presenza nel mazzo di carte anomale, allora non avevano più alcuna difficoltà a identificare correttamente tutte le carte che venivano loro mostrate, sia che fossero anomale sia che non lo fossero. Un mutamento nelle loro cognizioni e nelle loro aspettative era accompagnato da un mutamento in ciò che vedevano, benché guardassero sempre gli stessi oggetti fisici.

Un altro esempio è fornito da un gioco infantile che consiste nel trovare i caratteri di un volto umano confusi tra il fogliame nel disegno di un albero. In questo caso ciò che si vede, vale a dire l'impressione soggettiva esperita da una persona che osserva il disegno, corrisponde sulle prime a un albero con relativi fusto, foglie e rami. Ma tutto questo cambia non appena viene rintracciato il volto. Ciò che prima veniva visto come fogliame e parti di rami è visto adesso come volto umano. Ancora una volta, lo stesso oggetto fisico è stato osservato prima e dopo la soluzione dell'indovinello ed è presumibile che l'immagine sulla retina dell'osservatore non cambi quando si è trovata la soluzione e il volto è stato scoperto. Inoltre, se la figura viene guardata dopo qualche tempo, un osservatore che abbia già sciolto una volta l'indovinello riconoscerà di nuovo il volto facilmente. In questo esempio, quel che uno spettatore vede è condizionato dallo stato delle sue conoscenze e dalla sua esperienza. (Chalmers, 1986, 34)

Non sarebbe difficile produrre esempi tratti dalla pratica scientifica e dalla storia della scienza che mettono in chiaro questo stesso punto, vale a dire che ciò che gli osservatori vedono, le esperienze soggettive che hanno quando osservano un oggetto o una scena, non è determinato soltanto dalle immagini sulle loro retine, ma dipende anche dall'esperienza, dalle cognizioni, dalle attese e dallo stato interiore complessivo di chi osserva.

Un ricercatore che sperava di sapere qualcosa circa l'opinione degli scienziati a proposito della natura della teoria atomica, chiese a un famoso fisico e a un eminente chimico se un singolo atomo di elio fosse una molecola o no. Entrambi risposero senza esitare, ma le loro risposte non furono identiche. Per il chimico l'atomo di elio era una molecola, perché si comportava come tale rispetto alla teoria cinetica dei gas. Per il fisico, d'altra parte, l'atomo di elio non era una molecola, perché non presentava nessuno spettro molecolare. Presumibilmente entrambi parlavano della stessa particella, ma essi la consideravano secondo le differenti prospettive delle loro rispettive formazioni ed attività di ricerca. Le loro esperienze nel risolvere problemi avevano insegnato loro che cosa doveva essere una molecola. Senza dubbio le loro esperienze avevano avuto molti punti in comune, ma in questo caso, portavano i due specialisti a conclusioni diverse. (Kuhn, 1969, 73-74)

Ci volle più di un secolo e mezzo dopo la scoperta degli spermatozoi perché si arrivasse a capirne il ruolo nella fecondazione, e dovettero passare duecento anni dalla scoperta dei gameti maschile e femminile perché il concetto di fecondazione venisse formulato correttamente. Si vede qui bene quale parte giocassero le rappresentazioni mentali dei vari studiosi nell'interpretazione delle loro osservazioni, *cioè, in ultima analisi, nell'edificazione del sapere scientifico.*

[...]pure observation does not create, or increase, knowledge without personal inputs which are needed to elaborate the information. In fact, there is nothing really objective in physics, if by objective we mean that something follows necessarily from observation, like the proof of a theorem. There are, instead, beliefs everywhere. Nevertheless, physics is objective, or at least that part of it that is at present well established, if we mean by 'objective', that a rational individual cannot avoid believing it. This is the reason why we can talk in a relaxed way about beliefs in physics without even remotely thinking that it is at the same level as the stock exchange, betting on football scores, or... New Age. (D'Agostini, July 1999, 124)

Gli scienziati più famosi del mondo hanno condizionato le loro metodologie di ricerca con le loro credenze, le loro opinioni e le loro intuizioni personali in diversi modi, spesso senza rendersene conto o senza darne esplicita e formale comunicazione. Tale soggettività comprendeva idee, intuizioni, credenze forti, interpretazioni scientifiche di principi di base, che essi avevano elaborato e acquisito durante i propri studi o su cui avevano ricevuto adeguate informazioni: si fondavano, cioè, su osservazioni precedentemente elaborate da loro stessi o da altri.

Le “credenze” fanno parte naturale del “fare” scienza e l'ammettere che esistono non necessariamente pregiudica la percepita oggettività delle scienze stesse.

In other words, one needs only to look closely at how frontier science makes progress, instead of seeking refuge in an idealized concept of objectivity[...]My preferred motto on this matter is "*no one should be allowed to speak about objectivity unless he has had 10-20 years working experience in frontier science, economics, or any other applied field*"[...]the statistician D. Berry [...] has amused himself by counting how many times Hawking uses 'belief', 'to believe', or synonyms, in his '*A brief history of time*'. The book could have been entitled '*A brief history of beliefs*', pointed out Berry in his talk. (Ivi, 123)

Le conclusioni scientifiche sono, pertanto, *il risultato della fusione delle credenze e delle interpretazioni presenti prima di effettuare esperimenti con l'analisi ben più oggettiva di dati sperimentali.*

E' necessario imparare a guardare con occhio esperto nel telescopio o al microscopio: i raggruppamenti non strutturati di macchie scure e luminose che vede il principiante sono altra cosa dalle forme e spettacoli minuziosi che l'osservatore competente è in grado di discernere. Qualcosa di simile deve essere accaduto quando Galileo per primo introdusse il telescopio come strumento utile all'esplorazione dei cieli. Le riserve avanzate dai rivali di Galileo nell'accettare fenomeni quali le lune di Giove da lui scoperte devono essere attribuite in parte non a pregiudizi, ma a positive difficoltà nel vedere attraverso quello che, dopo tutto, era un telescopio alquanto rudimentale. (Chalmers, 1986, 35)

Quest'ultima citazione mi permette inoltre di sottolineare che la stessa scelta della strumentazione di laboratorio è ovviamente influenzata dalle nostre aspettative.

In breve, la decisione di usare una particolare apparecchiatura e di usarla in un particolare modo indica che si dà per scontato in forma più o meno cosciente che si dovranno verificare solo circostanze di un certo tipo. Vi sono, oltre alle aspettative teoriche, anche quelle strumentali, e queste hanno spesso svolto un ruolo decisivo nello sviluppo scientifico. (Kuhn, 1969, 83)

Per dirla con una espressione di Koyré che mi piace molto(vedi per esempio Koyré, 1950, 39), gli strumenti sono davvero “*incarnazioni*” della teoria. Con ciò ovviamente non si vuole affermare che la scienza debba abbandonare la strumentazione convenzionale, ma mettere in luce che le relative scelte inevitabilmente restringono il campo dei fenomeni accessibili all’indagine scientifica. Attualmente si sta proponendo un nuovo paradigma scientifico, all’interno del quale si vuole tener conto delle credenze pregresse *in maniera per quanto possibile formale*. Il relativo formalismo parte proprio dal teorema di Bayes e da alcune conseguenze teoriche che da esso scaturiscono. Nel prossimo paragrafo, pur evitando i dettagli tecnici, cercherò di presentare meglio la portata e le possibili conseguenze scientifiche di questo cambiamento. Voglio qui sottolineare che questo nuovo paradigma è anch’esso figlio legittimo di quegli ideali di razionalità e precisione che hanno caratterizzato la nascita della società moderna. E’ questo però un approccio che non cerca rifugio in un concetto ideale di “oggettività”, ma cerca di fare i conti con il modo concreto con cui scienziati o ricercatori operano o possono operare al meglio nel valutare i risultati dei loro esperimenti e delle loro osservazioni.

Posso ancora ricordare che ponendo accento soprattutto sul *momento decisionale* questo approccio sembra poter offrire *una maggiore pragmaticità* nel risolvere problemi in differenti ambienti tecnici, industriali ed aziendali.

Many scientists now believe that a paradigm shift in the sense of Kuhn[...] has been taking place in the way that scientific inference is, and will be, carried out. Many scientists now recognize the advantages of bringing prior beliefs into the inference process in a formal way from the start. The alternative is attempting to achieve total objectivity, or pretending to have done so, even though prior information is often brought to bear on the problem anyway, in surreptitious or even unconscious ways. Subjectivity may enter the scientific process surreptitiously in the form of seemingly arbitrarily imposed constraints, in the introduction of initial and boundary conditions, in the arbitrary levels of what should be called a statistically significant result, in the deemphasizing of certain data points which represent suspicious observations[...], and so on.(Press, Tanur, 2001, 204)

Le conoscenze pregresse e le relative aspettative solo apparentemente non hanno una parte rilevante nel comporre ed analizzare un esperimento. Anche i cosiddetti “dati grezzi” sono per loro natura interpretati. E per quanto riguarda la successiva valutazione di questi dati, anche se spesso tacitamente si suppone che le appropriate procedure siano neutrali rispetto alla teoria ciò *non è evidentemente vero*. Nella complessa valutazione delle prove esse sono applicate sensatamente solo da coloro che comprendono tutti i dettagli dell’esperimento.

Gli esperimenti iniziano e finiscono in una matrice di credenze[...]credenze sul tipo di strumentazione, sui programmi di ricerca, sui giudizi individuali relativi ad ogni comportamento locale di pezzi di apparato.(Galison, 1987, 45)

Nel laboratorio nulla è semplicemente dato. *Le misure sono prese, non date.*
La scienza senza credenze pregresse non ha senso.

3. Possibili campi di applicazione

[...]much qualitative research, both empirical and theoretical, is normally prerequisite to fruitful quantification of a given research field. In the absence of such prior work, the methodological directive, “go ye forth and measure,” may well prove only an invitation to waste time.

Thomas S. Kuhn

Statistical thinking will one day be as necessary for efficient citizenship as the ability to read and write.

H. G. Wells

I possibili campi di applicazione dell'inferenza Bayesiana sono molteplici, nelle scienze come in tutti i processi decisionali. Per avere una interessante panoramica di differenti esempi si può senz'altro vedere Press, Tanur, 2001 e Bernardo, Smith, 2000 e Bradley, Thomas 2000.

The allocation of billions of U.S. federal dollars now depends on Bayesian estimates of population characteristics (e.g. poverty) in small geographic areas. The FDA now not only permits but encourages Bayesian designs and analyses in applications seeking approval for new medical devices such as pacemakers, an area where large clinical trials are infeasible but directly relevant historical information is often available. The Microsoft Windows Office Assistant is based on a Bayesian artificial intelligence algorithm- similarly, a prominent movie-rental chain now offers a web-based Bayesian expert system to help you choose which film to rent, based on your general preferences and the names of other films you have enjoyed.(Bradley, Thomas, 2000, viii)

La statistica Bayesiana si può di fatto applicare ogniqualevolta si voglia valutare la probabilità (grado di fiducia) che un particolare evento possa avverarsi, denotata per esempio con $P\{\text{event}\}$, quando questo evento è generato da un meccanismo *random*. Il *frequentista* valuta le procedure basandosi sull'immaginazione della raccolta ripetuta di campioni da un particolare modello che definisce la distribuzione della probabilità dei dati osservati condizionata da parametri sconosciuti. La statistica Bayesiana richiede un modello e, in più, una distribuzione "a priori" di tutte le quantità sconosciute nel modello (parametri e dati mancanti). Il modello e la distribuzione vengono utilizzati per calcolare la distribuzione degli elementi sconosciuti utilizzando anche i dati osservati (la distribuzione *posteriore*).

Una prima formula di carattere generale, che deriva direttamente da una delle forme con cui può essere scritto il teorema di Bayes, e che per essere compresa non richiede particolari conoscenze specifiche, può essere ben introdotta come segue.

Bayesian statistical analysis depends fundamentally on a relationship derivable from the axioms of probability theory that enables us to convert the chance that an event A will happen, conditional on our having information about event B into the chance that event B will occur once we know that event A has already occurred. Then, using this symbolism, Bayes' formula asserts that

$$P\{(\text{event } B \text{ will occur}) \text{ given } (\text{event } A \text{ has occurred})\} \\ = P\{B \text{ given } A\} = \frac{P\{A \text{ given } B\} \times P\{B\}}{P\{A \text{ given } B\} \times P\{B\} + P\{A \text{ given not } B\} \times P\{\text{not } B\}}$$

In this formula, $P\{A \text{ given } B\}$ is generally referred to as the *model*; it depends on the observational data, as we will see. $P\{B\}$ is generally referred to as the *prior probability* (since this probability depends only on information known prior to observing the data). $P\{B \text{ given } A\}$ is referred to as the *updated or posterior probability* of B since it is conditioned on having information about event A . (The prior probability of B was not so conditioned).(Press, Tanur, 2001, 206-207)

È forse utile ricordare che, nonostante il teorema di Bayes sia stato formulato in varie maniere diverse durante gli anni, una delle prime formulazioni, presentate da Laplace all'inizio delle sue memorie del 1774 (*Mémoire sur la probabilité des causes par les évènements. Mémoires de l'Académie royale des sciences présentés par divers savants*; vedi Stigler, 2000), è molto simile a quella contemporanea che ho appena presentato.

If an event can be produced by a number n of different causes, then the probabilities of these causes given the event are to each other as the probabilities of the event given the causes, and the probability of the existence of each of these is equal to the probability of the event given that cause, divided by the sum of all the probabilities of the event given each of these causes. (Laplace, quoted in Stigler, 2000, 102)

Vale la pena di sottolineare ancora, anche per la particolare importanza pratica che ciò rappresenta, che negli anni recenti risultati davvero incoraggianti sono stati ottenuti nelle *decisioni in campo medico*(vedi Parmigiani, 2002).

Ciò che segue è un esempio semplice ma appropriato in questo campo, per mostrare un'applicazione dell'approccio generale (preso da D'Agostini, 2002 dove si trovano la presentazione e discussione dettagliate). Viene scelto un cittadino italiano da sottoporre al test dell'AIDS. Supponiamo che l'analisi usata per riscontrare l'infezione da HIV possa dichiarare "positive" persone sane con una probabilità dello 0,2%. Assumiamo che l'analisi dichiari la persona "positiva". Possiamo dire che, visto che la probabilità di un errore d'analisi è solo dello 0,2%, la probabilità che la persona sia veramente infettata è del 99,8%? Certamente no se si calcola sulla base di una stima di 1.000.000 di persone infette su una popolazione di 60 milioni (*informazioni conosciute prima di osservare i dati*). Infatti, una semplice applicazione della formula di Bayes appena presentata, dimostra che la probabilità che la persona sia sana è del 55%! Una maniera semplice per arrivare a tale risultato è d'immaginare di sottoporre al test l'intera popolazione. In tal caso, il numero di persone dichiarate positive sarà uguale a quello di tutte le persone infette più lo 0,2% della popolazione restante. Il totale sarebbe di 1.000.000 persone infette e 120.000 persone sane.

Ovviamente nei differenti campi di applicazione il formalismo matematico raggiunge livelli di una certa complicazione, sui quali non è certo il caso di soffermarci ora da un punto di vista strettamente tecnico.

Per quanto riguarda il campo più prettamente scientifico si può innanzi tutto ipotizzare un contributo allo studio dei sistemi complessi. Come è noto negli ultimi decenni dalla ricerca condotta nel campo della chimica, della fisica, della biologia, dell'informatica e, di recente, delle scienze sociali è emerso un nuovo paradigma scientifico: il paradigma della complessità (vedi Lane, 2002).

I fenomeni complessi sono caratterizzati da una serie di entità che interagiscono l'una con l'altra.

In seguito a queste interazioni, alcune proprietà di un'entità possono cambiare, compresa la sua posizione nella struttura reticolare e le modalità di interazione. Se le interazioni sono locali, gli oggetti di interesse nei fenomeni complessi sono però in genere globali, funzione di un modello di eventi interattivi che rimane relativamente stabile su una scala temporale molto più ampia rispetto a quella degli eventi stessi dell'interazione. Di frequente questi modelli meta-stabili di interazione si auto-organizzano, e le strutture e persino la funzionalità che ne risultano possono spesso essere descritte in un linguaggio che non contiene riferimenti alle entità sottostanti e alle loro interazioni. In quel caso i modelli vengono definiti *emergenti*, e lo studio dell'auto-organizzazione e dell'emergenza costituisce l'obiettivo primario della ricerca sulla complessità. (Lane, 2002, 14-15)

Questi modelli sono in parte simili ai modelli statistici. In particolare, entrambe le classi di modelli sono costituite da famiglie parametrizzate di modelli di probabilità, ed entrambe mirano a estrapolare deduzioni sui fenomeni del mondo reale.

[...]negli ultimi dieci anni, i teorici della complessità hanno dimostrato quanti interessanti problemi di biologia, chimica, fisica e scienze sociali presupponga proprio questo tipo di dinamica. Le teorie e i metodi della deduzione sarebbero di grande aiuto nell'affrontare tali problemi.[...]i metodi data-analitici e deduttivi per studiare i fenomeni emergenti sono difficili da individuare. Le tecniche statistiche disponibili non sono adatte allo scopo e la misurazione delle strutture emergenti e la correlazione di queste misure a descrizioni adeguate degli eventi di interazione a livello locale sottostanti rappresentano problemi di difficile soluzione. Attualmente quasi tutti i ricercatori della complessità si trovano ad agire d'istinto, ed è difficile intuire in quale direzione si muovano e riconoscere quando arrivano a destinazione. Credo ci sia la reale possibilità che giovani scienziati di impostazione statistica offrano contributi reali e importanti ai gruppi di ricerca della complessità[...].(ivi, 37)

Un campo, per esempio, strettamente correlato è quello della genetica statistica.

One of the most important challenges to data analysis in modern statistical genetics and molecular biology is the problem of how to analyze gene expression data in which the number of variables or dimensions, "p", greatly exceeds the number of replicates, "n" (that is, we have p much greater than n). Such data frequently are generated by a piece of equipment called a microarrayer. The data analysis problem is opposite to the usual problem in statistical data analysis in which typically p is much less than n. In the microarrayer context, we might have p = 10,000, for example, while n =

10. In such contexts, to reduce the problem to one where ordinary statistical methods can be applied, it is necessary either to increase the number of replicates (sample size) or to reduce the number of dimensions, or both. The Bayesian approach may offer the best solution to such problems since in Bayesian analysis, as is easy to show, bringing prior information to bear on a model is equivalent to adding more replicates, or increasing the effective sample size. (Press, Tanur, 2001, 212-213).

Inoltre si sta tentando di arrivare ad una nuova teoria della valutazione delle incertezze nel campo strettamente fisico. L'atteggiamento di fondo è a mio avviso ben riassumibile come segue.

As far as physics applications are concerned, the importance of the subjectivist approach stems from the fact that it is the only approach which allows us to speak in the most general way about the probability of hypotheses and true values, concepts which correspond to the natural reasoning of physicists. As a consequence, it is possible to build a consistent inferential framework in which the language remains that of probability. This framework is called Bayesian statistics, because of the crucial role of Bayes' theorem in updating probabilities in the light of new experimental facts using the rules of logics. Subjective ingredients of the inference, unavoidable because researchers do not share the same status of information, are not hidden with the hope of obtaining objective inferences, but are optimally incorporated in the inferential framework. Hence, the prior dependence of the inference should not be seen as a weak point of the theory. On the contrary, it obliges practitioners to consider and state clearly the hypotheses which enter the inference and take personal responsibility for the result. In any case, prior information and evidence provided by the data are properly balanced by Bayes' theorem, and the result is in qualitative agreement with what we would expect rationally. Priors dominate if the data is missing or of poor quality or if the hypotheses favored by the data alone is difficult to believe. They become uninfluential for routine high accuracy measurements, or when the evidence provided by the data in favor of a new hypotheses is so strong that physicists are obliged to remove deeply rooted ideas. (D'Agostini, December 1999, 11-12)

Talvolta, voglio sottolineare, è assolutamente legittimo credere più ai propri pregiudizi che ai dati empirici. Quando ad esempio si usa uno strumento che sappiamo potrebbe non essere totalmente affidabile, o quando si fanno per la prima volta misure in un nuovo campo o in un nuovo range di possibili valori di una specifica grandezza e così via. Per esempio è ovvio che è più facile credere che uno studente abbia fatto un banale errore piuttosto che credere che abbia fatto una nuova scoperta! La seguente riflessione di Poincaré (del 1905) può essere esemplificativa al riguardo.

The impossibility of squaring the circle was shown in 1885, but before that date all geometers considered this impossibility as so 'probable' that the Académie des Sciences rejected without examination the, alas! too numerous memoirs on this subject that a few unhappy madmen sent in every year. Was the Académie wrong? Evidently not, and it knew perfectly well that by acting in this manner it did not run the least risk of stifling a discovery of moment. The Académie could not have proved that it was right, but it knew quite well that its instinct did not deceive it. If you had asked the Academicians, they would have answered: 'We have compared the probability that an unknown scientist should have found out what has been vainly sought for so long, with the probability that there is one madman the more on the earth, and the latter has appeared to us the greater' (Poincaré, 1952, 24)

Le conclusioni hanno, a mio avviso, il pregio di essere chiare e coraggiose.

In conclusion [...], I prefer to state explicitly the naturalness and necessity of subjective priors [...]. If rational people (e.g. physicists), under the guidance of coherency (i.e. they are honest), but each with unavoidable personal experience, have priors which are so different that they reach divergent conclusions, it just means that the data are still not sufficiently solid to allow a high degree of intersubjectivity (i.e. the subject is still in the area of active research rather than in that of consolidated scientific culture). On the other hand, the step from abstract objective rules to dogmatism is very short [...]. (D'Agostini, Luglio 1999, 30)

Ovviamente tutte queste considerazioni possono essere discusse criticamente, e infatti sono state contrastate da tutti color che *credono* in un ideale "oggettivista".

In poche parole, tutti coloro che contestano il metodo Bayesiano si appellano al concetto filosofico di “razionalità” per confermare la necessità e la possibilità di “oggettività”. Ciò è per esempio caratteristico della scuola Popperiana.

The cognitive value of a theory has nothing to do with its psychological influence on people's mind. Belief, commitment, understanding are states of the human mind. But the objective, scientific values of a theory is independent of the human mind which creates it or understands it, its scientific value depends only on what objective support these conjecture have in facts. (Lakatos, 1978, 1)

Popper afferma più di una volta che è impossibile che un'ipotesi possieda informazioni di alto contenuto e contemporaneamente di alta qualità.

They[scientists] have to choose between high probability and high informative content, *since for logical reasons they cannot have both.* (Popper, 1992, 363)

Si può rispondere a questa critica, a mio avviso, in maniera relativamente semplice nel seguente modo.

Such a charge is quite baseless. There is *nothing* in logic or the probability calculus which precludes the assignment of even probability 1 to any statement, however strong, as long as it is not a contradiction, of course. The only other way in which probabilities depend on logic is in their decreasing monotonically from entailed to entailing statements. But this again does not preclude anybody from assigning any consistent statement as large a probability as they wish. Popper's thesis that a necessary concomitant of logical strength is low probability is simply incorrect. (Howson, Urbach, 1993, 390)

Nel cercare difetti nell'approccio Bayesiano, alcuni avanzano altre obiezioni (una, per esempio, si basa sulla *necessità della probabilità soggettiva a priori*). Data la natura introduttiva di questo libro, non discuterò qui queste obiezioni. Per una buona analisi delle ragioni di chi contesta l'approccio Bayesiano vedi per esempio, Moore, 1997, Mayo, 1996 ed Efron, 1986. Per risposte dettagliate vedi Howson, Urbach, 1993, in particolare l'ultimo capitolo. Vorrei concludere queste specifiche considerazioni con una citazione del fisico Feynman che mi ha fatto riflettere per prima su questi temi.

Some years ago I had a conversation with a layman about flying saucers- because I am scientific I know all about flying saucers! I said “I don't think there are flying saucers”. So my antagonist said, “Is it impossible that there are flying saucers? Can you prove that it's impossible?” “No”, I said, “I can't prove it's impossible. It's just very unlikely”. At that he said, “You are very unscientific. If you can't prove it impossible then how can you say that it's unlikely?” *But this is the way that is scientific.* It is scientific only to say what is more likely and what less likely, and not to be proving all the time the possible and the impossible. To define what I mean, I might have said to him, “Listen, I mean that *from my knowledge* of the world that I see around me, I think that it is much more likely that the reports of flying saucers are the results of the known irrational characteristics of terrestrial intelligence than of the unknown rational efforts of extra-terrestrial intelligence”. *It is just more likely.* That is all. (Feynman, 1967, 47)

Il problema dell'analisi tecnica di dettaglio della statistica Bayesiana a seconda dei differenti campi di applicazione, come appena accennavo e come si può facilmente intuire, si fa piuttosto complesso per il formalismo matematico richiesto, e va ben al di là degli scopi di questo scritto. Chi fosse interessato comunque può vedere un'ottima introduzione in Press, Tanur, 2001, ed una esauriente esposizione degli attuali risultati in Fisica delle Alte Energie in D'Agostini, Luglio 1999.

BIBLIOGRAFIA

- Bernardo J. M., Smith A. F. M., *Bayesian Theory*, New York, John Wiley & Sons INC, 2000
- Box G. E. P., *Bayesian Inference in Statistical Analysis*, New York, John Wiley & Sons INC, 1992
- Bradley P., Thomas A. L., *Bayes and Empirical Bayes Methods for Data Analysis*, Boca Raton, CRC Press, 2000
- Chalmers B., *Che cos'è questa scienza?*, Milano, Est Mondadori, 1987
- D'Agostini G., *Bayesian Reasoning in High-Energy Physics: Principles and Applications*, Geneve, CERN, July 1999
- D'Agostini G., *Errori e incertezze di misura – rassegna critica e proposte per l'insegnamento -*, Roma, Università La Sapienza Dipartimento di Fisica, gennaio 1999
- D'Agostini G., *Teaching Statistics in the Physics Curriculum: Unifying and Clarifying Role of Subjective Probability*, in *American Journal of Physics Special Theme Issue on Thermodynamics, Statistical Mechanics, and Statistical Physics*, Gould and Tobochnik eds, December 1999
- D'Agostini G., *Bayesian Reasoning in Physics. A Critical Introduction*, Singapore, World Scientific, 2002
- Efron B., *"Why isn't everyone a Bayesian?"*, Am. Stat. 40,1986
- Feynman R., *The character of the physical law*, Cambridge MA, The MIT Press, 1967
- Galison, P.L., *How experiments end*, Chicago, Chicago University Press, 1987
- Howson C., Urbach P., *Scientific Reasoning. The Bayesian Approach*, Chicago, Open Court, 1993
- Koyré A., *Apport scientifique de la Renaissance*, in *Revue de Synthèse*, XLVII , 1950
- Kuhn T.S., *La struttura delle rivoluzioni scientifiche*, Torino, Einaudi, 1969
- Lad F., *Operational Subjective Statistical Methods. A Mathematical, Philosophical, and Historical Introduction*, New York, John Wiley & Son INC., 1996
- Lakatos I., *Philosophical Papers*, vol. 1, Cambridge, Cambridge University Press, 1978
- Lane D. A., *Complessità: modelli e inferenza*, in AA.VV., *Complessità e biologia*, a cura di Biava P.M., Milano, Bruno Mondadori, 2002, pp. 13-41
- Mayo D. G., *Error and the Growth of Experimental Knowledge. Science and its Conceptual Foundation*, Chicago, Chicago University Press, 1996
- Moore D. S., *"Bayes for beginners? Some reasons to hesitate"*, Am. Stat. **51**, 1997
- Orear J., *Enrico Fermi, The Man*, in *Nuovo Saggiatore* vol.17 anno 2001 no 5-6, pp. 30-38

Parmigiani G., *Modeling in Medical Decision Making. A Bayesian Approach*, New York, John Wiley & Sons INC, 2002

Poincaré H., *Science and Hypothesis*, London, Dover Publications, 1952

Popper K., *The Logic of Scientific Discovery*, London, Routledge, 1992

Press S. J., Tanur J.M., *The Subjectivity of Scientists and the Bayesian Approach*, New York, John Wiley & Sons, INC, 2001

Severi M., *Introduzione all'esperienza fisica*, Bologna, Zanichelli, 1986

Stigler S. M., *The History of Statistics. The Measurement of Uncertainty before 1900*, Cambridge, The Belknap Press of Harvard University, 2000