MANUALI

UMANISTICA

– 10 –

MARIA GRAZIA SANDRINI

Filosofia dei metodi induttivi e logica della ricerca

Firenze University Press
2009

Filosofia dei metodi induttivi e logica della ricerca/
Maria Grazia Sandrini. – Firenze : Firenze University
Press, 2009.
(Manuali . Umanistica ; 10)

http://digital.casalini.it/9788864530376

ISBN 978-88-6453-034-5 (print)
ISBN 978-88-6453-037-6 (online)

Immagine di copertina:
© Starfotograf (Franz Pfluegl) | Dreamstime.com

Progetto grafico di Alberto Pizarro Fernández

Università degli Studi di Firenze
Firenze University Press
Borgo Albizi, 28
50122 Firenze, Italy
http://www.fupress.com/

Printed in Italy

Sommario

Maria Grazia Sandrini, *Filosofia dei metodi induttivi e logica della ricerca*, ISBN 978-88-6453-034-5 (print) ISBN 978-88-6453-037-6 (online)

Introduzione

Si è scritto molto sulla metodologia della ricerca, ma quasi esclusivamente in relazione al problema dello sviluppo, o crescita, della conoscenza: la scienza procede per rivoluzioni o per generalizzazioni sempre più ampie? Per rotture traumatiche o per progresso lineare? Il metodo della scienza è induttivista o deduttivista? verificazionista o falsificazionista?

Moltissimo si è scritto anche a proposito del concetto di probabilità, come se in esso si esaurisse il senso induttivo della conoscenza. Che induzione e probabilità siano entrati, da un certo tempo in poi, in strettissima connessione, è un fatto innegabile; come lo è la vivace discussione sorta attorno alla definizione ed alla natura del concetto di probabilità. Tuttavia appare eccessivamente limitativo ricondurre tutto il discorso intorno all'induzione al concetto ed al calcolo della probabilità.

Molto è stato scritto attorno all'induzione in un senso generico e legato a una immagine spesso obsoleta della scienza, e molto è stato scritto, da una diversa prospettiva, per criticare radicalmente quella concezione. In altro ambito di ricerca, ci si è occupati dell'esplicazione di singoli, per quanto rilevanti, concetti scientifici, come quello di probabilità, innescando talora polemiche in gran parte prive di senso. Quasi nulla, invece, è stato scritto sulla effettiva metodologia in uso nella scienza empirica e sperimentale, senza la considerazione della quale ogni concetto di conoscenza appare sfocato e impreciso.

Che un 'rompicapo' possa essere risolto all'interno del paradigma in carica, o inneschi un processo che sfocerà nell'affermazione di un nuovo paradigma, sarà la ricerca stessa a deciderlo e la ricostruzione storica e razionale a tentare poi di comprenderlo. Scienza 'normale' e scienza 'rivoluzionaria' si intrecciano e si confondono nella pratica di ogni giorno: cercare di rinchiudere il processo scientifico in un rigido paradigma di lettura e di interpretazione – o, peggio ancora, di norme generali che la

Maria Grazia Sandrini, *Filosofia dei metodi induttivi e logica della ricerca*, ISBN 978-88-6453-034-5 (print) ISBN 978-88-6453-037-6 (online) © 2009 Firenze University Press

scienza, per essere 'razionale', *dovrebbe* seguire –, mi sembra fuorviante e sclerotizzante, e, in ogni caso, sterile.

Se la filosofia in genere, e la filosofia della scienza in specie, non vogliono rinnovare la frattura tra scienza e filosofia, devono evitare anzitutto, credo, di fare della prima l'oggetto di una nuova metafisica: solo se la scienza potrà riconoscersi nell'immagine che di essa offre la filosofia, senza tuttavia fare di se stessa una nuova metafisica rifiutando a priori di confrontarsi con prospettive filosofico-epistemologiche nuove e stimolanti, il dialogo tra filosofia e scienza potrà essere fecondo.

La riflessione filosofica può contribuire a vari livelli allo sviluppo della scienza, non fosse altro che attraverso l'esercizio di chiarificazione dei significati dei termini teorici della scienza, quel compito di 'ricostruzione razionale' da Carnap efficacemente intrapreso.

Peraltro, la scienza non appare più così 'neutrale' come la volevano i positivisti; essa 'costruisce' immagini del mondo. Mettere in evidenza quali siano queste immagini, e che cosa esse implichino, fa parte del lavoro critico della filosofia e dei contributi che essa può fornire al procedere della scienza.

D'altra parte, appare impensabile, oggi, una filosofia che non si senta tenuta a fare onestamente i conti con i risultati ed i metodi della scienza; non di una scienza astratta e mummificata, ma della scienza effettiva, nella sua eventuale disunitarietà e disorganicità, nella sua concretezza teorica e pratica e quindi anche nella sua complessità, contraddittorietà e molteplicità di aspetti. A tale proposito, vorrei osservare che l'immagine della scienza che spesso i filosofi hanno eretto sulla base della considerazione di uno solo dei suoi campi di indagine, anche se, fino a qualche decina di anni addietro, il più avanzato, cioè la fisica, non è che una immagine per molti aspetti deformante, oltre che parziale.

Un'altra considerazione da tener presente è che la concezione della conoscenza e della stessa scienza è riflessa nei metodi che la scienza usa. Questo era ben chiaro a scienziati come Newton e Galileo, nei quali concezione del mondo e procedure sperimentali e d'osservazione erano strettamente connesse. Diviene così di rilevante interesse studiare le implicazioni epistemologiche dei metodi della scienza empirica e discutere il contributo effettivo che esse possono apportare tanto allo sviluppo della conoscenza quanto alla riflessione filosofica.

I metodi della scienza empirica sperimentale sono per lo più metodi di inferenza statistica. Sono metodi induttivi, nel senso proprio del termine, e sono utilizzati tanto ai fini della previsione quanto ai fini della generalizzazione. Essi non forniscono algoritmi per inventare nuove teorie o per prendere decisioni in modo automatico; più modestamente, essi sono soltanto metodi escogitati per aiutare, se usati correttamente, ad analizzare e 'pesare' i dati in nostro possesso, secondo regole prestabilite ed in accordo con criteri di scientificità storicamente dati, ma non esimono dalla responsabilità delle conclusioni. Strumenti, dunque, ma tuttavia non utensili adattabili ad ogni situazione: come ogni metodologia, riflettono quei presupposti generali, quella concezione del mondo, in relazione a cui sono stati creati e organizzati; abbastanza flessibili, tuttavia, per adattarsi anche ad ampi slittamenti e cambiamenti del quadro di riferimento, almeno finché le finalità, di tempo in tempo perseguite dalla scienza, non si differenzino troppo fra loro, fino ad un punto di rottura prevedibile, in cui essi si mostreranno ormai inadeguati e anacronistici e saranno soppiantati da nuovi metodi.

A chi guardi all'effettiva prassi scientifica, le metodologie induttive oggi in uso – oggetto tutte di studio della statistica e della statistica matematica – appaiono riconducibili a due grossi approcci, di differente origine storica sebbene tra loro correlati, ricollegantisi l'uno al metodo sperimentale, ipotetico-deduttivo, e l'altro alla teoria delle probabilità ed alle applicazioni laplaceane del teorema di Bayes.

I due approcci induttivi non sempre convivono pacificamente tra gli statistici teorici, anche se in questi ultimi decenni la vecchia polemica sembra un po' smorzata. Sopravvivono invece ancora posizioni contrapposte relative alla definizione ed alla natura, soggettiva od oggettiva, della probabilità. Per più di un secolo, tuttavia, a partire dalla metà dell'Ottocento, la polemica tra i due approcci metodologici è stata piuttosto accesa e talora addirittura veemente. Sono stati autori quali Boole, Venn, Fisher, ad avanzare le critiche più violente all'approccio bayesiano all'induzione; ed è proprio attorno alla figura di Fisher che si sono concentrate, anche dopo la sua morte, le punte di maggiore asprezza della polemica. Tale polemica, che vede impegnata una mentalità positivistica, per molti versi acritica sul piano epistemologico, contro una mentalità matematico-razionalista, più spregiudicata sul piano ontologico, è stata a mio parere ingiustamente trascurata da quanti, tra i filosofi, si sono occupati dell'induzione e del metodo scientifico.

Oggetto di questo studio sarà anche quello di tentare di chiarire i termini di questa polemica e di trarne un bilancio complessivo, utile, spero, tanto al filosofo che si interessi ai problemi della conoscenza, quanto allo scienziato e allo statistico che vogliano riflettere sulla portata dei propri metodi di analisi e di inferenza.

Nascerà anche, da questo studio, una nuova prospettiva che, recuperando la distinzione tra contesto della giustificazione e contesto della scoperta, permette di guardare ai due approcci contrapposti come a due metodi viceversa complementari.

Un ulteriore esito del presente lavoro porterà ad una limitazione dell'applicazione sensata dei metodi induttivi, e quindi al suggerimento che tra conoscenza empirica e induzione vi possa non essere quella relazione di identificazione proposta e sostenuta dalla maggior parte dei neopositivisti. Questa questione è connessa con il problema dell'accettazione delle teorie scientifiche, da molti ricondotta ad una questione induttiva. La tesi qui sostenuta è che le teorie scientifiche abbiano una struttura ed una funzione che rende insensato, oltre che impossibile, il tentativo di racchiuderle all'interno di una procedura induttiva.

Capitolo 1
Induzione e conoscenza

1.1 Conoscenza e generalizzazione

Benché ogni più precisa connotazione di ciò che chiamiamo «conoscenza» sia strettamente collegata con la cultura ed i metodi logici e tecnologici sviluppati da una società, nonché con le sue aspettative e con le finalità da essa assegnate alla stessa scienza, è asserzione comunemente condivisa, ancora oggi, che 'conoscere' significhi anzitutto *generalizzare*. Bisognerà dunque chiedersi, in primo luogo, cosa sia una generalizzazione, quale funzione essa assolva nell'attività conoscitiva ed infine in quale modo essa possa essere formulata e stabilita. Affrontare tali questioni significa arrischiarsi in quel complesso nodo problematico, epistemologico e metodologico, compreso generalmente sotto il termine *induzione*.

La questione dell'induzione, che ha una sua prima formulazione classica di cui diremo, abbraccia oggi una riflessione ben più ampia e articolata sulle metodologie induttive, soprattutto probabilistiche e statistiche, in uso nella scienza, e sulle loro implicazioni epistemologiche. Vi è infatti una sorta di solidarietà tra approccio metodologico e concezione generale della conoscenza che deve essere tenuta presente per una chiara comprensione della questione.

Che cos'è, dunque, una generalizzazione? In prima approssimazione possiamo dire che una generalizzazione è un enunciato 'generale' circa un qualche stato di cose. In senso formale, un enunciato è detto *generale* quando non contiene costanti individuali, ma solo variabili individuali e quantificatori; in senso più informale un enunciato è *generale* quando intende riferirsi ad un intero 'universo', finito od infinito che sia.

Facciamo degli esempi. Se dico che *tutte* le palline di quest'urna, dalla quale ho estratto un certo numero di palline, sono bianche, formulo una generalizzazione, anche se relativa ad un universo piuttosto delimitato e numericamente finito qual è quello delle palline contenute nell'urna.

Passando al campo dei fenomeni naturali, infinito per definizione, perché infinite appaiono le ripetizioni possibili degli eventi, sono generalizzazioni gli enunciati del tipo «tutti i corvi sono neri»; ma lo sono anche enunciati statistici, come «il fumo è responsabile del 90% dei tumori al polmone», perché anche questo tipo di enunciati asserisce qualcosa, una proporzione o una regolarità, che riguarda l'intero universo di discorso.

Gran parte dei comportamenti usuali della vita quotidiana sottintendono generalizzazioni: «il cibo nutre», «il fuoco ustiona», «il cane abbaia», e così via. Tali enunciati, a dispetto della loro forma apparentemente 'particolare', si riferiscono invece ad universi infiniti. «Il cibo nutre», ad esempio, intende affermare che sempre, ogni volta che qualcosa sia 'cibo', esso avrà la proprietà di nutrire: $(x)(Cx \rightarrow Nx)$.

La scienza mostra specifico interesse soprattutto per enunciati generali di un tipo più complesso, come ad esempio la legge di inerzia («ogni corpo persegue nel suo stato di quiete o di moto rettilineo uniforme finché non intervengano forze esterne») o la legge di gravitazione di Newton («due corpi si attraggono in modo direttamente proporzionale al prodotto delle loro masse ed inversamente proporzionale al quadrato della loro distanza»).

Lasciamo per il momento da parte la questione se questo tipo di asserzioni generali, che più interessa alla scienza, debbano considerarsi sullo stesso livello di generalizzazioni empiriche del tipo «tutti i corvi sono neri», o se invece non dovremmo rilevare qualche importante differenza.

Constatiamo, invece, che, sebbene tutte le generalizzazioni siano suggerite dall'osservazione di casi particolari, esse prescindono dal riferimento a casi particolari. La loro formulazione suona contemporaneamente sia come una *descrizione* generale di un aspetto del mondo, sia come *legge* cui ogni particolare caso implicato debba sottostare. In questa ambiguità di significato descrittivo/prescrittivo degli asserti generali si cela gran parte del senso della loro funzione nella conoscenza.

Lo scienziato, così come l'uomo comune nella sua esperienza quotidiana, ha generalmente davanti a sé un numero finito di osservazioni, o esperienze, singole ed isolate fra loro: oggetti ed eventi della vita quotidiana o risultati di laboratorio, ottenuti talora mediante sofisticate procedure, appaiono come le isolate pietre miliari di un cammino che non può essere contenuto, nella sua interezza, nell'esperienza di nessun uomo, e neppure, forse, in quella dell'intera umanità. Non solo: se, infatti, tali singole ed isolate esperienze non potessero essere in qualche modo collegate e raccolte tra loro, esse non potrebbero dar luogo che a descrizioni tautologiche di un'esperienza soggettiva, effimera e ormai trascorsa. La conoscenza di casi particolari isolati e non collegabili non sarebbe affatto una conoscenza e risulterebbe del tutto inutilizzabile.

Di fronte alla transeunte immediatezza e soggettività dell'esperienza, l'attività conoscitiva si pone dunque anzitutto il compito di ricondurre ogni singolo oggetto di esperienza sotto enunciazioni generali, la cui portata è intesa estendersi ben oltre le osservazioni effettive, o anche solo possibili, di ogni uomo. Solo in tal modo si rende possibile assolvere all'ulteriore compito di *spiegare* i singoli fenomeni.

La generalizzazione è dunque l'essenza della spiegazione, anche se, a rigore, non dovrebbe esaurirlo: spiegare un fenomeno significa, in primo luogo, dunque, ricondurlo sotto un asserto generale. Generalizzare e spiegare sono tuttavia compiti ancora troppo astratti e, d'altra parte, insufficienti, da soli, a garantire la scientificità degli stessi asserti generali, se non vengono connessi a quella fondamentale funzione del conoscere che è la *previsione*. Se assumiamo l'enunciato generale «tutte le palline di quest'urna sono bianche», possiamo prevedere con certezza che ogni pallina estratta dall'urna sarà bianca; sulla base della legge di gravitazione universale è possibile inferire il comportamento dei corpi, e così via. In altre parole, una generalizzazione stabilita permette di compiere previsioni, più o meno certe, ma non arbitrarie, circa eventi futuri e fenomeni non ancora osservati, gettando un ponte tra l'attività teoretica e quella pratica dell'uomo.

1.2 Il problema dell'induzione

Con il termine «induzione» si è inteso originariamente il passaggio logico da enunciati particolari ad un enunciato generale; in breve, l'inferenza da «A è un corvo nero», «B è un corvo nero», ..., «R è un corvo nero» all'enunciato «tutti i corvi sono neri».

Dal punto di vista meramente logico, l'inferenza induttiva, a differenza di quella deduttiva, non è un'inferenza valida. Dall'enunciato «tutti gli uomini sono mortali» consegue *necessariamente* che ogni specifico individuo umano è mortale; ma dalle premesse «A è N», «B è N», ecc., *non* consegue *necessariamente* che tutti gli individui, della classe cui A, B, ..., appartengono, abbiano la proprietà N, per quanto numerosi siano gli individui osservati. La certezza della conclusione potrebbe essere raggiunta soltanto nel caso in cui si fosse in grado di procedere ad una enumerazione completa degli individui della classe in questione; ma in tal caso la generalizzazione diverrebbe una descrizione banalmente vera, corrispondendo ad una congiunzione di tutti gli enunciati particolari veri.

La differenza tra deduzione ed induzione può essere ulteriormente chiarita mediante i diagrammi illustrati nelle Figg. 1 e 2.

Figura 1

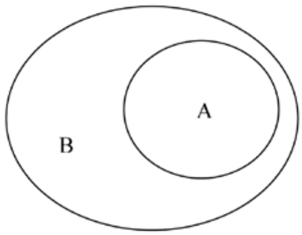
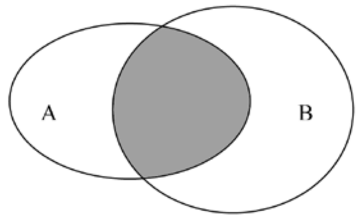


Figura 2



Il caso della deduzione rientra sempre nella Fig. 1: se sappiamo che la classe A è inclusa interamente nella classe B, allora possiamo dedurne che ogni singolo membro di A è anche membro di B. Nel caso induttivo, invece, il problema consiste proprio nel fatto che non sappiamo mai se ci troviamo in una situazione illustrata dalla Fig. 1 oppure in quella illustrata dalla Fig. 2. Siano stati osservati ad esempio *n* individui di A, tutti con la proprietà B: se il nostro caso rientrasse nella Fig. 1 l'inferenza «tutti gli A sono B» sarebbe vera. Ma la situazione potrebbe invece essere quella illustrata dalla Fig. 2, e solo un caso fortunato può averci fatto incontrare finora solo individui di A compresi nell'intersezione tra le due classi, e quindi aventi la proprietà B, per quanto grande possa essere stato il numero delle osservazioni; ma in tal caso, ovviamente, l'inferenza dell'enunciato universale dai casi osservati sarebbe falsa.

Già Aristotele aveva rilevato la fallacia dell'inferenza induttiva. Nel campo della conoscenza empirica, tuttavia, ci troviamo perennemente in una situazione induttiva: le nostre osservazioni sono sempre particolari e di numero finito; ma la conoscenza (anzi, la vita stessa!) esige che si traggano da esse inferenze generalizzanti, possibilmente esplicative, in grado di guidare le nostre previsioni.

Tuttavia, se la conoscenza empirica è strettamente collegata con l'inferenza induttiva, e se, d'altra parte, il ragionamento induttivo non è in grado di garantire le conclusioni delle proprie inferenze, come è possibile giustificare la conoscenza stessa? Come possiamo nutrire fiducia nei risultati della ricerca scientifica? Il «problema dell'induzione» è in sostanza esattamente quello della possibilità di giustificare razionalmente la conoscenza.

La scienza dei secoli XVII-XVIII credeva di poter trarre direttamente dall'esperienza la prova delle conclusioni raggiunte. Il metodo sperimentale, introdotto da Galileo, costituiva non solo uno strumento di indagine, ma anche uno strumento di giustificazione razionale dei risultati.

Tale metodo poggiava, più o meno esplicitamente, su due grossi assunti: la causalità e l'uniformità della natura. La causalità sembrava indispensabile per autorizzare il passaggio da leggi generali ai fatti sperimentali e poi di nuovo da questi ultimi alle leggi. L'uniformità della natura sembrava indispensabile a garantire una conoscenza che permettesse di trascendere il piano effimero del presente nella previsione di eventi futuri.

Questi due assunti furono messi in luce da Hume[1] e da lui sottoposti ad una critica demolitrice. Ma già prima di Hume, la questione circa il tipo di certezza, che un numero qualsiasi di osservazioni possa fornire ad asserti di carattere generale, quali le leggi naturali, era stata sentita; è importante, anzi, il fatto che essa fosse già affiorata anche nella mente di alcuni scienziati. Valga per tutti ricordare il fisico olandese Gravesande[2], il quale si rese ben conto di come le famose quattro *Regulae philosophandi* di Newton, con le quali questi intendeva costituire la base di ogni sapere induttivo, non fossero regole logiche valide a priori, e neppure potessero essere giustificate sulla base dell'esperienza, costituendo viceversa, esse stesse, il fondamento di ogni induzione. Per salvare la validità di tali regole, e, con esse, la possibilità del conoscere, Gravesande si vedeva costretto a ricorrere al postulato che Dio regge il mondo con leggi immutabili e a giustificare tale principio facendo appello alla bontà del Creatore, il quale avrebbe così disposto per venire incontro alle necessità dell'uomo, le cui possibilità di azione e di vita sarebbero annullate se le argomentazioni induttive non avessero fondamento alcuno.

La critica più approfondita all'induzione, e agli assunti su cui essa si regge, è dovuta tuttavia al filosofo inglese David Hume. Egli vide chiaramente come tutti i ragionamenti induttivi siano riconducibili in definitiva alla relazione di causalità. Solo per mezzo di questa relazione, egli afferma, è possibile andare al di là di ciò che risulta evidente in base alla testimonianza della memoria e dei sensi. Per analizzare questa relazione, Hume ricorre al famoso esempio delle palle da biliardo. Noi vediamo che la palla A, spinta da noi, va a colpire la palla B, mettendola in movimento. La palla A è la *causa* del movimento di B e tra questi due eventi (A colpisce B e il movimento di B) ha luogo una connessione costante: ogni volta che una palla ne colpisce un'altra, quest'ultima si mette in movimento. Ora, se dal piano del 'vedere' passiamo a quello del 'ragionare', ha luogo una differenza fondamentale. Quando vediamo che A muove verso B, *inferiamo,* prima che l'urto abbia luogo, che certamente la palla A metterà in moto la palla B. Ed è proprio questa inferenza a guidare la mano del giocatore esperto. Tuttavia, un'inferenza del tutto simile viene fatta, secondo Hume, per tutti i fenomeni naturali: siamo certi, prima di vederlo sorgere, che domani mattina sorgerà il sole; siamo certi che il nostro amico è mortale

[1] D. Hume (1711-1776), *A Treatise of Human Nature,* London 1739*;* vol. I, *Of the Understanding*.

[2] W.J. Gravesande (1688-1742), *Physices elementa mathematica experimentis confermata, sive Introductio ad philosophiam newtonianam,* 2 voll., London 1720-21.

prima di averlo visto morire, e così via. Si pone così la seguente questione: la relazione causale è forse, in quanto sembra precederla, indipendente dall'esperienza?

Qui la risposta di Hume è tassativa: la causalità non consegue da alcun ragionamento a priori, ma nasce proprio dall'esperienza, ogni volta che troviamo che certi particolari oggetti sono costantemente connessi tra loro. Prima di ogni esperienza, «il fuoco ustiona» ed «il fuoco non ustiona» sono generalizzazioni parimenti concepibili per l'intelletto umano, come possiamo, ad esempio, constatare nei bambini. Ne deriva l'impossibilità di fornire una dimostrazione razionale del principio di causalità, che non riesce a differenziarsi dalla connessione costante tra eventi. Resta tuttavia il fatto che *sappiamo* che la palla A metterà in moto la palla B, che domani sorgerà il sole, e via di seguito. Come si spiega allora tale fatto? Si spiega, afferma Hume, in base all'esperienza passata, e questa inferenza dal passato al presente e al futuro si regge sul postulato dell'uniformità della natura. Ma questo postulato è indimostrabile; anzi, è esso stesso basato sull'induzione.

Hume perviene così a dimostrare che è impossibile giustificare l'induzione per mezzo dell'esperienza, perché una tale inferenza presuppone l'induzione a un livello più alto, dando così avvio ad un regresso all'infinito.

Non è possibile, quindi, fornire all'induzione, e dunque alla conoscenza, una giustificazione razionale. Una giustificazione, se così possiamo chiamarla, può semmai essere offerta sul piano psicologico: *l'abitudine,* che nasce dal ripetersi di impressioni fra loro simili, sollecita nell'individuo aspettative circa il futuro e credenze generalizzanti. Ciò non offre garanzie di validità all'induzione, né alla conoscenza, ma spiega e giustifica il persistere di un atteggiamento umano.

Malgrado la critica demolitrice di Hume, la questione sui fondamenti e sulla validità dell'induzione fu ancora al centro della discussione che, attorno alla metà dell'Ottocento, vide protagonisti Herschel e Mill, da una parte, e Whewell, dall'altra[3].

Nella suo *Discourse*, opera di alta divulgazione scientifica, Herschel aveva assunto una concezione strettamente empiristica della conoscenza: non soltanto le leggi di natura, ma persino gli assiomi della geometria euclidea, sono frutto di inferenze induttive. L'analisi dei procedimenti induttivi, concernenti la scoperta delle cause dei fenomeni, è articolata da Herschel in una serie di regole; ma in nessun momento del ragionamento scientifico è ammesso l'intervento di concetti a priori. L'enunciato stesso

[3] J.F.W. Herschel (1738-1822), *Preliminary Discorse on the Study of Natural Philosophy,* London 1830; W. Whewell (1794-1866), *Philosophy of the Inductive Science,* 1840, ristampa London 1967; J.S. Mill (1806-1873), *System of Logic*, London 1843; trad. it. *Sistema di logica,* Ubaldini, Roma 1968.

di una legge di natura include ciò che Newton definiva *analogia naturae*, ossia l'uniformità della natura, assunta da Herschel come il fondamento dell'induzione e la garanzia della universalità delle leggi induttivamente scoperte e verificate. Infatti, i soli fatti utili per la ricerca scientifica sono quelli che invariabilmente accadono nelle medesime circostanze: «se non hanno questo carattere non possono essere compresi in leggi: difettano di quella universalità che li rende atti a far parte, in quanto particelle elementari, degli assiomi universali che cerchiamo di scoprire»[4]. Dalla stessa esperienza, quindi, trarrebbe origine, a suo avviso, la nostra nozione di «ordine naturale», che rende possibile l'induzione: la necessità sarebbe inerente alla natura!

A Whewell, che lo recensì condividendone in gran parte le argomentazioni, il saggio di Herschel apparve come il primo serio lavoro di filosofia della scienza fisica, dopo Bacone. Tuttavia, Whewell insinua una nota personale: una legge è capace di includere una grande quantità di fenomeni, in virtù delle naturali connessioni del nostro pensiero; vi sarebbe cioè, per Whewell, una corrispondenza perfetta tra una legge generale di natura e «una singola concezione della mente»[5]. Già egli si pone, evidentemente, su quel terreno vagamente kantiano, che svilupperà in tutto il suo successivo lavoro: egli delinea il processo di inferenza induttiva come mediazione di due aspetti distinti: l'osservazione dei fatti, che è una semplice registrazione, ed il processo della «ragione pura» che si pronuncia sulla necessità e sull'universalità delle generalizzazioni tratte dai fenomeni. Se i fatti sono un indiscriminato mucchio di perle, la ragione è il filo che le collega insieme[6]. L'esperienza, insomma, non è autosufficiente, per Whewell: egli concorda con le critiche mosse, da opposti punti di vista, da Hume e da Kant al concetto di causa e ritiene che in nessun campo la nuda esperienza possa offrire quei criteri di universalità e di necessità in grado di giustificare razionalmente l'induzione: «in ogni processo induttivo c'è un'idea generale data non dai fenomeni, ma dalla mente»[7]. Così, Whewell, se, da un lato, si allea con Hume nel rifiuto del principio di uniformità della natura e della nozione realistica di causa, dall'altro, cercherà nelle strutture a priori della ragione una giustificazione all'induzione, approdando ad una filosofia della scoperta che privilegia gli aspetti soggettivi e pone l'accento sulla genialità e sulla libertà inventiva dello scienziato. La concezione di Whewell sollevò aspre reazioni antiidealistiche, sia da parte di Herschell che da parte di J.S. Mill. Quest'ultimo inserisce, nel suo *System of Logic*, di cui stava portando a termine la stesura, capitoli

4 J.F. Herschel, *Preliminary Discorse,* cit., p. 119.

5 Cfr. W. Whewell, recensione al *Discourse* di Herschel, «London Quarterly Review», 1831, p. 396.

6 Ivi, p. 377.

7 W. Whewell, *The Philosophy of Inductive Science,* cit., vol. II, p. 49.

dedicati alla critica della concezione di Whewell e ribadisce che le proposizioni generali della scienza sono inferite dai fenomeni per induzione. Anzi, non prestando affatto attenzione né alla critica di Hume né al regresso all'infinito che così si innescava, Mill giunge a sostenere che la definizione stessa dell'induzione reca implicito l'assunto dell'uniformità della natura, la quale è essa stessa un caso di induzione[8]. Naturalmente, il problema del fondamento dell'induzione si era posto e si poneva in modo particolarmente acuto per coloro che desideravano attenersi, nella ricerca scientifica, a canoni empiristici e antimetafisici; non toccava chi, come ad esempio Laplace, accoglieva una visione assolutamente deterministica del mondo. Proprio a Laplace, tuttavia, dobbiamo la separazione tra il piano ontologico ed il piano conoscitivo (cosa che, del resto, aveva fatto anche Kant) e l'inserimento della nozione di probabilità nel concetto stesso di conoscenza.

In realtà, la critica di Hume aveva colpito in profondità; dopo di lui, il problema dell'induzione viene a trovarsi in una *impasse,* e le concezioni meritevoli di attenzione, che si avvicendano, non sono in realtà che riproposizioni di posizioni empiristico-realistiche, ancorate all'uniformità della natura, oppure di posizioni 'quasi-kantiane', che cercano una mediazione tra le esigenze dell'empirismo e quelle della ragione.

Alcuni hanno pensato, anche, di poter eludere il problema, cercando di ridurre il procedimento induttivo alla sola argomentazione deduttiva ed hanno ritenuto di trovare uno strumento idoneo nel metodo ipotetico-deduttivo, di cui diremo tra poco; altri, come Hans Reichenbach, hanno creduto, in tempi a noi più vicini, di trovare nella nozione di «conoscenza probabile», e nel calcolo delle probabilità, la risoluzione dell'annoso problema, fattosi particolarmente acuto per tutti coloro che, come i neopositivisti, perseguivano come prioritaria la questione del fondamento della scienza e del sapere.

Secondo Reichenbach, il problema dell'induzione si dissolve, allorché si rinunci alla idea di una conoscenza 'certa', colpita dalla critica di Hume. Il ricorso al calcolo delle probabilità comporterebbe, a suo parere, la sostituzione dell'idea di una conoscenza certa con una concezione probabilistica della conoscenza, cosicché il processo conoscitivo diventerebbe un processo di approssimazione graduale alla verità, attraverso gradi di probabilità man mano più alti, ed, insieme, un procedimento per prove ed errori, con il quale la probabilità di una ipotesi può aumentare o decrescere, annullarsi o tendere alla *certezza* morale. Così, a suo parere, cadrebbe la questione della validità dell'induzione. Sebbene la concezione epistemologica di Reichenbach presenti, rispetto alle concezioni precedenti, il vantaggio di una visione più dinamica della scienza e di una considerazione più aggiornata

[8] Cfr. J.S. Mill, *System of Logic,* trad. it. cit., pp. 302-303.

degli strumenti metodologici in suo possesso, non è possibile condividere il suo ottimismo al riguardo. Reichenbach non sembra infatti vedere che neppure il ricorso alla nozione di probabilità (e tanto meno di una probabilità frequentistica come la sua!) può affrancare, entro una concezione della conoscenza che si vuol mantenere empiristica, dall'assunzione, pur sempre arbitraria, di una qualche forma di regolarità nel corso naturale degli eventi (una forma statistica, forse, ma pur sempre qualcosa di equivalente all'assunto di uniformità della natura), senza la quale nessuna inferenza predittiva sarebbe possibile. Mi pare sintomatico, al riguardo, il fatto che lo stesso Reichenbach fornisca poi una giustificazione del tutto pragmatica dell'induzione. Tale giustificazione è vividamente espressa nella nota metafora del cieco sperduto per impervi viottoli di una montagna disabitata: fermarsi equivale ad una morte certa. Non resta, per lui, che procedere cautamente, tastando la via con il bastone, consapevole che ogni passo può avvicinarlo alla salvezza, oppure alla morte!

1.3 Il metodo ipotetico-deduttivo e la conoscenza probabile

La nascita della scienza moderna segna l'affermazione del metodo sperimentale nella scienza. Nato con Galileo e codificato da Newton, tale metodo, tanto per l'impalcatura logico-deduttiva, quanto per il peso dato all'osservazione sperimentale, sembrò rispondere soddisfacentemente sia alle esigenze di certezza conoscitiva avanzate dal razionalismo seicentesco, che trovavano peraltro motivo di ottimismo negli sviluppi della matematica, sia alle richieste di una scienza che, ormai insofferente delle esasperate e sterili concettualizzazioni dell'aristotelismo, andava costruendo una nuova immagine di razionalità scientifica destinata a rivoluzionare l'intero quadro culturale e sociale.

Del resto, un rinnovamento dei metodi di indagine della natura era auspicato da più parti. Bacone[9], contemporaneo di Galileo, aveva contribuito, in sede filosofica, all'affermarsi di due aspetti che risulteranno fondamentali nel nuovo metodo: una considerazione in chiave pragmatico-operativa della conoscenza ed una nuova concezione della causalità.

Grazie al primo aspetto era posta in evidenza la funzione attiva dello scienziato, cui è affidato il compito di 'interrogare' intelligentemente la natura. Ciò porterà a privilegiare l'esperimento, rispetto all'osservazione passiva e spesso casuale dei fenomeni. Con il secondo, si tendeva a sostituire la concezione di una conoscenza di 'essenze', propria dell'aristotelismo, con quella di una conoscenza delle correlazioni tra fenomeni, aprendo la via alla liberazione del concetto di causa da ogni riferimento teleologico, sostanzialistico, ecc.

[9] F. Bacon (1561-1626), *Novum Organum*, London 1620.

Con Galileo e con Newton si afferma un ideale di razionalità scientifica che privilegia in eguale misura lo strumento matematico ed il fattore empirico-sperimentale. Il compito della scienza si configura quindi come spiegazione dei fenomeni mediante la costruzione di una teoria matematica, dalla quale possa essere dedotto, anche attraverso una lunga catena di implicazioni, il comportamento dei fenomeni stessi. Lo strumento matematico permette di conseguire quel rigore, univocità ed esattezza, che solo gli aspetti quantitativi dei fenomeni possiedono. Ma il processo di generalizzazioni sempre più ampie e di astrazioni matematiche non deve perdere il contatto con il piano empirico dell'osservazione: una teoria non ha valore scientifico senza un'adeguata verifica sperimentale. Grande cura è perciò rivolta alla programmazione degli esperimenti ed alla costruzione di strumenti idonei, sempre più precisi e potenti, nonché all'esattezza delle rilevazioni empiriche e delle misurazioni.

Le indicazioni metodologiche da seguire nella ricerca fisica sono da Newton sintetizzate nelle quattro *Regulae philosophandi* da lui anteposte al terzo libro dei *Principia*[10]:

Regola I – «Delle cose naturali non devono essere ammesse più cause di quelle che sono vere e sufficienti a spiegare i fenomeni». Infatti la natura è «semplice e non abbonda di cause superflue delle cose»: *«Natura nihil agit frustra»*.

Regola II – «Perciò, fin dove è possibile, devono essere assegnate le stesse cause ad effetti naturali dello stesso genere».

Regola III – «Quelle qualità, che non possono essere liberamente aggiunte o sottratte *(quae intendi atque remitti nequeunt)* e che competono a tutti i corpi nei quali è possibile compiere esperimenti, devono essere considerate come qualità di tutti quanti i corpi». Le qualità dei corpi, infatti, non possono essere conosciute se non attraverso gli esperimenti e perciò anche le qualità generali devono essere stabilite in accordo con gli esperimenti: *«natura simplex esse soleat et sibi semper consona»*.

Regola IV – «Nella filosofia sperimentale, le proposizioni ricavate per induzione dai fenomeni, se non si oppongono ipotesi contrarie, devono essere considerate come vere, o rigorosamente vere o quanto più possibile vicine alla verità *(pro veris aut accurate, aut quamproxime haberi debent)*, finché non siano occorsi altri fenomeni mediante i quali esse siano rese più accurate oppure soggette a eccezioni». *«Hoc fieri debet* – aggiunge Newton – *ne argumentum inductionis tollatur per hypoteses»*, intendendo così rifiutare, contro il cartesianesimo, ogni proposizione che non sia ricavata dai fenomeni, bensì presupposta senza prova sperimentale.

[10] I. Newton (1642-1727), *Philosophiae Naturalis Principia Mathematica*, London 1687.

È enunciato, nelle prime tre regole, il principio di induzione basato sulla semplicità e sulla uniformità della natura. La quarta regola sancisce il peso decisivo dell'esperimento nella costituzione della conoscenza, la quale, non solo deve fondarsi sull'osservazione e non deve ammettere nulla che non sia empiricamente suffragato, ma deve anche essere aperta all'estensione ed alle correzioni dell'esperienza.

Le *Regulae* enunciate da Newton, spesso impropriamente condensate nell'emblematica quanto abusata frase *«hypoteses non fingo»,* rappresenteranno per più di due secoli il fondamento metodologico del sapere scientifico. Occorre tuttavia notare che esse esprimono, nella loro enunciazione, criteri assai generali e pertanto generici. È perciò, forse, più nella pratica sperimentale che vanno ricercate le caratteristiche logiche e strutturali di quel metodo scientifico che verrà spesso successivamente indicato come *metodo ipotetico-deduttivo*[11], e che costituisce ancor oggi, sia pure con le dovute integrazioni matematico-probabilistiche, uno dei principali modelli di ragionamento induttivo della metodologia scientifica.

Esempio di tale pratica sperimentale può essere trovato negli esperimenti di Newton sulla rifrangibilità dei raggi luminosi: qui l'ipotesi corpuscolare della luce è ripresa e rigorosamente formulata; le varie conseguenze ne sono dedotte con esattezza matematica; infine, viene dimostrato che tale ipotesi è in grado di spiegare assai bene tutti i risultati sperimentali in precedenza ottenuti.

Esempi illuminanti, riguardo al metodo, troviamo nelle opere di Galileo[12], come mostra il seguente brano:

> L'invenzione fu del caso e mia fu solamente l'osservazione e il far di essa capitale e stima [...]. Raschiando con uno scarpello di ferro tagliente una piastra d'ottone per levarle alcune macchie, nel muovervi sopra lo scarpello con velocità, sentii una volta e due, tra molte strisciate, fischiare e uscirne un sibilo molto gagliardo e chiaro; e guardando sopra la piastra, veddi un lungo ordine di virgolette sottili, tra di loro parallele e per egualissimi intervalli l'una dall'altra distanti. Tornando a raschiar di nuovo più e più volte, m'accorsi che solamente nelle raschiate che fischiavano lasciava lo scarpello le 'ntaccature sopra la piastra; ma quando la strisciata passava senza sibilo, non restava pur minima ombra di tali virgolette. Replicando poi altre volte lo scherzo, strisciando ora con maggiore ed ora con minor velocità, il sibilo riusciva di tuono or più acuto ed or più grave; ed osservai i segni fatti nel suono più acuto esser più spessi, e quelli del più grave più radi.

[11] In molta letteratura anglosassone tale metodo è chiamato semplicemente «deduttivo». Cfr. anche J.S. Mill, *System of Logic*, cit., libro III, cap. XI.

[12] G. Galilei (1564-1642), *Opere,* 2 voll., Utet, Torino 1964.

Avendo quindi generalizzato il fenomeno, attraverso l'osservazione di numerosi differenti casi, ed in particolare dei suoni che si ottengono con corde di grossezza, lunghezza e tensione diverse, Galileo enuncia la sua teoria sulle vibrazioni sonore:

> [...] dico che non è la ragion prossima ed immediata delle forme de gl'intervalli musici la lunghezza delle corde, non la tensione, non la grossezza, ma sì bene la proporzione de i numeri delle vibrazioni e percosse dell'onde dell'aria che vanno a ferire il timpano del nostro orecchio, il quale esso ancora sotto le medesime misure di tempi vien fatto tremare[13].

Rileviamo quindi: l'osservazione casuale dei segni lasciati dallo scalpello; la formulazione dell'ipotesi che il suono e la sua acutezza dipendano dalle vibrazioni e dalla loro frequenza; i vari ragionamenti deduttivi e le ripetute esperienze effettuati in base all'ipotesi; la raccolta e l'interpretazione di dati differenti rilevati sia direttamente, contando i segni della piastra scalfita, sia indirettamente sulle corde mettendo i suoni della piastra all'unisono con quello del cembalo; infine la formulazione della legge con la quale il suono, ricondotto nell'ambito dei fenomeni vibratori, è spiegato come effetto delle vibrazioni.

Ulteriori numerosi esempi del metodo ipotetico-deduttivo possono essere trovati nell'intera storia delle scienze sperimentali.

Particolarmente ben sostenuti da una logica rigorosa appaiono gli esperimenti condotti da Pasteur, con i quali fu definitivamente sconfitta la teoria della generazione spontanea. Tale teoria, assai antica – pare possa essere fatta risalire addirittura ad Aristotele –, affermava che la vita, almeno nelle sue forme più minute, come quelle degli insetti e dei vermi, sorgeva dal nulla, spontaneamente. Essa prevalse a lungo, malgrado autorevoli opinioni contrarie, ed ancora nel XVII secolo vi era chi riteneva si potessero produrre rane dal limo di palude e anguille dall'acqua marina! Verso la metà del Settecento la teoria della generazione spontanea tornò in auge, a seguito dei lavori di Needham[14], il quale la sostenne con esperimenti ben immaginati, come quelli condotti con vasi ermeticamente chiusi, sterilizzati all'azione del calore, e nei quali si erano, nondimeno, sviluppati microrganismi.

Pasteur arrivò ad occuparsi della teoria della generazione spontanea attraverso lo sviluppo dei suoi studi sulla fermentazione. La teoria general-

[13] G. Galilei, *Discorsi intorno a due nuove scienze,* in *Opere*, cit., vol. II, pp. 674 sgg.

[14] J.T. Needham (1713-1781), *An account of some new microscopical diiscoveries,* London 1745. Esperimenti contrari alla teoria della generazione spontanea erano stati condotti, tra gli altri, da L. Spallanzani (1729-1799); ma essi non potevano essere considerati conclusivi poiché potevano essere resi compatibili con la teoria della generazione spontanea mediante l'introduzione di ipotesi *ad hoc.*

mente accettata, per spiegare questo tipo di fenomeni, sosteneva che tutte le materie azotate (sempre presenti, in un fenomeno di fermentazione, generalmente come sostanze albuminoidi) subiscono, esposte al contatto dell'aria, una alterazione di natura ignota, che dà loro il carattere di *fermento,* cioè la proprietà di agire sulle sostanze *fermentescibili* (l'altro elemento indispensabile in un fenomeno di fermentazione).

Pasteur era invece di diverso avviso:

> I miei studi mi condussero a conclusioni interamente differenti. Io trovavo che tutte le fermentazioni propriamente dette, viscosa, lattica, butirrica, la fermentazione dell'acido tartarico, dell'acido malico, dell'urea... , erano sempre correlative alla presenza e alla moltiplicazione di esseri organizzati. [...] Secondo me le materie albuminoidi non erano mai dei fermenti, ma l'alimento dei fermenti. I veri fermenti erano degli esseri organizzati.
>
> Ciò posto, i fermenti nascono, si sapeva, per il contatto delle materie albuminoidi con il gas ossigeno. Pertanto, mi dicevo, di due cose, una: dato che i fermenti delle fermentazioni propriamente dette sono organizzati, se il solo ossigeno, in quanto tale, dà loro origine col suo contatto con la materia azotata, tali fermenti sono delle generazioni spontanee; se tali fermenti non sono esseri spontanei, non è tanto il solo ossigeno che interviene come tale nella loro formazione, bensì come eccitante di un germe apportato nel medesimo tempo o esistente nelle materie azotate o fermentescibili. Al punto in cui io mi trovavo negli studi sulle fermentazioni, dovevo dunque formarmi un'opinione sulla questione delle generazioni spontanee[15].

La prima questione sperimentale che si presentò allora a Pasteur fu questa: vi sono dei germi nell'aria? E, se sì, in che numero? Di fatto, era già stato accertato che nella polvere dell'aria vi fossero delle uova di infusori e spore di funghi, ma, si sosteneva, in numero molto limitato o addirittura eccezionalmente. Pasteur studiò un metodo per poter prelevare, e quindi osservare al microscopio, la polvere *in sospensione* nell'aria, anziché quella depositata sulla superficie degli oggetti, come comunemente si era fatto, ritenendo che dalla polvere in riposo potessero essere state asportate dalle correnti d'aria le particelle più leggere.

L'esperimento, tanto ingegnoso quanto semplice, mostrò senza ombra di dubbio «che vi sono sempre in sospensione nell'aria comune dei corpuscoli organizzati del tutto simili ai germi degli organismi inferiori»[16], e che il loro numero è assai elevato.

[15] L. Pasteur (1822-1895), *Mémoire sur les corpuscoles organizés qui existent dans l'atmosphère. Examen de la doctrine des generations spontanées,* 1861; trad. it. in L. Pasteur, *Opere,* Utet, Torino 1972, pp. 401 sgg.

[16] Ivi, p. 417.

Come secondo passo, Pasteur provò che l'acqua di lievito di birra zuccherata, sostanza eminentemente alterabile, portata all'ebollizione per due o tre minuti entro palloni di vetro e quindi messa a contatto con aria calcinata, immessa nei palloni che venivano quindi sigillati al calore, non si altera per niente e può essere conservata indefinitamente, senza che mai si verifichi la nascita di infusori o muffe.

Si trattava ora di procedere al passo decisivo, di vedere cioè cosa sarebbe successo introducendo, in quegli stessi palloni, della polvere d'aria raccolta con il metodo già precedentemente collaudato. Tralascerò di descrivere le precauzioni adottate da Pasteur per deporre la polvere d'aria nei liquidi racchiusi nei palloni, in modo che non venisse introdotto nient'altro, all'infuori di questa polvere. Il risultato dell'esperimento dette definitivamente ragione all'idea di Pasteur: inserita la polvere, comparvero nei palloni delle produzioni organizzate nel giro di 24-48 ore, esattamente il tempo necessario per la comparsa di queste stesse produzioni nell'acqua di lievito di birra zuccherata quando quest'ultima è posta a contatto con l'aria comune[17].

Nell'esempio ora esposto, Pasteur è mosso, come risulta dalla sua *Memoria,* dai dati osservativi e dai risultati di precedenti esperienze, alla formulazione di una ipotesi esplicativa per i fenomeni di fermentazione. Egli avanza quindi l'idea che organismi organizzati fossero presenti nell'aria e che fossero proprio questi, e non l'ossigeno in quanto tale, a dar luogo al fenomeno fermentativo. Procede quindi, sulla base di un ragionamento deduttivo, a costruire degli esperimenti di verifica, accuratamente controllati. Tali esperimenti dovevano dare una risposta affermativa alle seguenti questioni: che vi fossero organismi nell'aria; che un liquido fortemente alterabile rimanesse intatto se conservato a contatto con aria calcinata in recipiente sterile, mentre vi comparissero produzioni organizzate se messo a contatto con i microrganismi introdotti con la polvere d'aria.

Un altro illuminante esempio di processo ipotetico-deduttivo è quello, che Hempel ha contribuito a rendere famoso, del dr. Semmelweiss e della febbre da parto[18].

Come è noto, Ignaz Semmelweiss, medico presso il primo reparto di maternità dell'Ospedale Generale di Vienna, si trovò, negli anni 1844-1848, di fronte al problema di cercare una spiegazione al fatto che, presso il proprio reparto, il tasso di mortalità per tale tipo di febbre fosse assai più alto di quello del secondo reparto dello stesso ospedale. Si trattava di una differenza costante di circa il 5%, troppo alta per essere puramente accidentale. Nessuna ipotesi fu trascurata, come narra lo stesso

[17] Cfr. Ivi, pp. 417-419.

[18] C. Hempel, *Philosophy of Natural Science,* 1966; trad.it. *Filosofia delle scienze naturali*, Il Mulino, Bologna, 1968, pp. 15 sgg.

Semmelweiss, neppure la più strampalata: cambiamenti climatici, sovraffollamento, eventuali lesioni prodotte da visite mal condotte, la posizione assunta dalla donna nel parto, perfino un possibile trauma psicologico, causato dal passaggio del prete col viatico. Tutte queste ipotesi, ed altre ancora, furono esaminate e, ove possibile, sottoposte a prova, mediante cambiamenti del protocollo ospedaliero. Tutto risultò inutile, finché un malaugurato incidente capitato ad un collega, morto con sintomi analoghi a quelli della febbre da parto, a causa della ferita prodottagli da un bisturi usato per un'autopsia, mise Semmelweiss sulla giusta pista. Egli ipotizzò allora che le pazienti morissero a causa di una infezione, di cui egli stesso ed i suoi studenti erano portatori: era infatti abitudine passare a visitare le pazienti direttamente dalla sala di autopsia, dopo un lavaggio superficiale delle mani. Stabilito l'obbligo che le mani fossero accuratamente lavate mediante una soluzione di ipoclorito di sodio, prima di procedere alla visita delle puerpere, il tasso di mortalità calò rapidamente, assestandosi sul livello dell'altro reparto, dove le pazienti erano assistite da levatrici, le quali, ovviamente, non avevano nulla a che fare con le autopsie.

Semmelweiss volle successivamente rafforzare la propria ipotesi con un ulteriore, sciagurato, esperimento: egli, con i suoi assistenti, visitò dapprima una partoriente affetta da tumore cervicale in suppurazione e procedette poi alla visita di altre venti donne senza eseguire la disinfezione delle mani. Di queste puerpere, ben undici morirono per febbre da parto (inutile dire che oggi quest'ultimo esperimento verrebbe perseguito come atto criminale!).

Ulteriori e numerosi esempi di applicazione del metodo ipotetico-deduttivo possono essere facilmente reperiti: la fisica a cavallo tra il XIX ed il XX secolo, ad esempio, offre esempi famosi con gli esperimenti sui fenomeni elettrici ed elettromagnetici, sui raggi catodici, sulla diffrazione dei raggi X, sui fenomeni radioattivi, sulla trasmissione e sulla natura della luce, ecc.

Possiamo a questo punto schematizzare la struttura del metodo ipotetico-deduttivo, che si configura come una successione di quattro fasi in cui il fattore logico e quello empirico (le «certe dimostrazioni» e le «sensate esperienze» di Galileo) risultano strettamente connessi, sebbene la parola decisiva spetti infine ai risultati sperimentali:

1. Formulazione di una ipotesi generale, suggerita dalle osservazioni empiriche.
2. Deduzione, da tale ipotesi, delle conseguenze empiriche osservabili.
3. Esecuzione degli esperimenti o delle osservazioni empiriche.
4. Controllo dell'ipotesi, alla luce dei risultati sperimentali: sua convalida o confutazione.

Si noti che, mentre le fasi 2 e 3, strettamente interconnesse, rappresentano i momenti rigorosamente deduttivo e sperimentale del metodo, le fasi 1 e 4 presuppongono entrambe un ragionamento induttivo: è pertanto illusoria la fiducia di chi crede di aver trovato, in tale metodo, uno strumento per eludere la fallacia propria del ragionamento induttivo!

Nella sua forma originaria, il metodo ipotetico-deduttivo fondava la sua pretesa di validità conoscitiva sui due assunti criticati da Hume, la causalità e l'uniformità della natura. La causalità, in particolare, appare come il cemento indispensabile per saldare insieme le varie fasi del procedimento. All'assunto ontologico, si accompagna tuttavia la concezione della conoscenza come descrizione veritiera della realtà.

Quando, con Laplace, si opererà una frattura tra piano ontologico, deterministico, e piano epistemologico, connesso con lo stato di ignoranza dell'uomo, l'ideale di una conoscenza certa cederà il passo a quella di una conoscenza «probabile»: allora, accanto al metodo ipotetico-deduttivo, si andrà affermando nella scienza anche un approccio probabilistico all'induzione.

Il concetto di «conoscenza probabile», in senso lato, è un concetto implicito in ogni concezione che riconosca alla conoscenza carattere ipotetico ed aperto; in senso stretto, invece, esso è connesso con *l'approccio bayesiano,* cioè con un metodo il cui scopo finale è, come vedremo in seguito, quello di pervenire all'attribuzione di un valore di probabilità alle ipotesi, laddove il procedimento ipotetico-deduttivo, anche nelle sue forme derivate, è volto alla formulazione di un giudizio di accettazione o di rifiuto delle ipotesi .

La concezione di Laplace, in parte anticipata da Jakob Bernoulli, si pone come compromesso tra il rigido determinismo meccanicistico e la consapevolezza scientifica che la natura è *conoscibile,* ma non *conosciuta.* La conoscenza, nella sua forma complessiva ed assoluta, può ben costituire il limite ideale a cui tende la scienza effettiva, nel suo continuo processo di estensione del sapere; ma la conoscenza storicamente costituita non può prescindere dalla consapevolezza dell'ignoranza di cui essa è ancora permeata. In questo contesto, che salda il determinismo della natura con il probabilismo della conoscenza, il calcolo delle probabilità sembra fornire lo strumento matematico più idoneo all'accrescimento del conoscere attraverso l'acquisizione di sempre nuovi fatti empirici.

Rinnovato favore il metodo ipotetico-deduttivo ritroverà poi, nel clima diffuso di nuovo newtonianesimo proprio del positivismo ottocentesco, il cui perseguimento di un ideale di oggettività e neutralità della scienza finirà per ridurre in gran parte quest'ultima entro gli angusti limiti di una funzione meramente descrittiva di fenomeni osservati, cioè ad una sorta di ripetizione tautologica del dato. Caratterizza questo tipo di positivismo il richiamo costante ai fatti ed all'esperienza, e ad essi soltanto, ed il rifiuto di qualsiasi elaborazione mentale che non sia semplicemente deduttiva, ciò che comporterà una diffusa diffidenza anche verso l'approccio bayesiano all'induzione.

1.4 Vari sensi dell'induzione

II termine *induzione* ha avuto, nella storia del pensiero, accezioni assai differenti. Tale molteplicità di divergenti significati si è talora negativamente riflessa sulla metodologia scientifica e sulla concezione epistemologica, originando equivoci e sterili polemiche ma anche incomprensioni circa la portata dei medesimi strumenti metodologici. Vedremo che ci sono, nell'induzione, almeno due aspetti che sarà utile tenere distinti.

Originariamente, e in senso generale, induzione significa, come si è detto, inferenza dal particolare all'universale. Tipico esempio di questa definizione è l'induzione per *enumerazione semplice,* ripudiata da Bacone come «qualcosa di puerile, che conclude senza necessità, perché resta esposta al pericolo di un'istanza che contraddica»[19].

Con l'affermarsi del metodo sperimentale, in effetti, il senso dell'induzione ha legato sempre più il processo di generalizzazione a metodi sistematici di sperimentazione, cioè alla interrogazione intelligente e *finalizzata* della natura, in contrapposizione al rilevamento casuale e passivo dei fatti, ritenuto caratteristico appunto dell'induzione per enumerazione semplice. Per quasi tutto il secolo XVIII, i sostenitori del metodo induttivo tendono a sottolineare sopratutto l'esigenza e l'importanza dell'esperimento, in polemica con la sillogistica aristotelica e con il cartesianesimo. Ma nei primi decenni del secolo successivo cominciano ad emergere differenti accentuazioni di significato del termine «induzione».

Nel 1826, ad esempio, R. Whately[20] denuncia la vaghezza del termine «induzione», che sarebbe inteso, talora, per designare il processo di «investigazione e di connessione dei fatti», talaltra, quello di «deduzione» da questo. Whately vedeva bene che i due momenti erano distinti: il primo riguarda la scoperta e l'assunzione di premesse generali; il secondo è puro ragionamento deduttivo. Tra i due sensi, Whately non mancava di prendere posizione a favore del primo.

Qualche anno più tardi W. Hamilton[21] distingueva tre accezioni nell'uso del termine «induzione»:

a) un processo oggettivo di investigazione di fatti particolari, di classificazione e di raccolta, come processo preparatorio all'inferenza vera e propria;

[19] F. Bacone, *Opere filosofiche,* Bari 1965, vol. I, p. 234.
[20] R. Whately, *Elements of Logic,* London 1826.
[21] W. Hamilton (1788-1856), *Lectures on Logic*, London 1866.

b) un'inferenza *materiale* dell'universale dal particolare, sostenuta o da analogie naturali o da presunzioni specifiche dell'oggetto di indagine;
c) un'inferenza *formale* dell'universale dal particolare, legittimata unicamente dalle leggi del pensiero.

Sebbene le inferenze materiale e formale risultino spesso confuse tra loro, per Hamilton, che era un logico e si interessava solo di processi logici, soltanto l'inferenza formale esprimerebbe un corretto processo logico e induttivo. Quanto al processo di investigazione indicato al primo punto, Hamilton ritiene che esso, in quanto puro processo di scoperta, non possa che restare al di fuori di una scienza rigorosa. Tuttavia, Hamilton richiama l'attenzione sulla necessità di tenere distinti il senso dell'induzione come 'scoperta' da quello dell'induzione come 'inferenza'.

Quanto questa distinzione sia stata disattesa mostra anche la polemica tra Herschel, Whewell e J.S. Mill, di cui abbiamo già parlato. Per Whewell, che propende per tesi razionalistiche ed aprioristiche di tipo kantiano, l'induzione è il processo di «invenzione», con il quale si attua una connessione vera tra fatti, mediante una nuova ed appropriata concettualizzazione[22]. Per Herschel e per Mill, invece, l'induzione è sempre una *inferenza* dall'esperienza, e pertanto l'essenza del processo induttivo sta nei principi di 'prova', non nell'invenzione. Nel sostenere la natura inferenziale dell'induzione, Mill tende a sottolineare come sia meno ambiguo definire l'induzione come inferenza «dal noto all'ignoto», piuttosto che come generalizzazione da casi particolari, perché ci sono enunciati generali che non sono vere inferenze, ma semplici *descrizioni* di fatti osservati. Spiegare l'induzione come un collegamento di fatti per mezzo di concetti appropriati, come fa Whewell, significa appunto per Mill «confondere le pure descrizioni di fatti osservati con l'inferenza da questi fatti». Esiste, è vero, una relazione effettiva tra collegamento di fatti ed induzione: «il collegamento non è sempre induzione, ma l'induzione è sempre collegamento»[23]. Così, la scoperta di Keplero delle orbite ellittiche dei pianeti non fu per Mill, in polemica con Whewell, un'induzione, ma una semplice descrizione. Una scoperta del genere non è un'induzione più di quanto lo sia lo stabilire che una terra è un'isola dopo averla completamente circumnavigata: il navigante semplicemente *vede* che quella terra *è* un'isola, non l'inferisce[24].

È in relazione ai criteri di 'prova', dunque, che le induzioni differiscono dalle descrizioni.

[22] W. Whewell, *The Philosophy of Inductive Sciences*, cit; *Philosophy of Discovery*, London 1860.

[23] J.S. Mill, *System of Logic*, trad. it. cit., p. 296.

[24] Ivi, pp. 288 sgg.

Si sono precisati, così, due primi sensi divergenti dell'induzione: induzione come 'scoperta' guidata solo dall'intuito e dalla genialità dello scienziato, ed induzione come 'inferenza', come metodo di prova, cioè, che identifica nell'esperimento la sua ragion d'essere.

Questi due sensi dell'induzione, fra loro inconciliabili, si affermano e si spartiscono equamente il campo per tutto il XIX secolo, mentre il panorama metodologico tende ad arricchirsi e a complicarsi con l'applicazione sempre più estesa del calcolo delle probabilità e con lo sviluppo della statistica, come disciplina soprattutto descrittiva ed analitica di dati, in un primo momento, ma successivamente anche inferenziale.

Sul versante positivistico si manifestano anche forti resistenze all'uso della probabilità. G. Boole[25], ad esempio, riconosce all'induzione carattere di *inferenza,* ma, muovendosi su un piano rigidamente oggettivistico, ne restringe il senso a quel tipo di inferenza probabilistica chiamata «inferenza diretta», un'inferenza, cioè, da popolazione nota ad un suo campione casuale (l'equivalente probabilistica della deduzione), perché egli ritiene che, solo sulla base di dati completamente noti per esperienza, sia lecito trarre inferenze.

A pochi anni di distanza J. Venn[26], anch'egli rigorosamente positivista, definirà invece l'induzione come procedimento di classificazione dei dati e di generalizzazione da essi, ma sosterrà che essa non ha niente a che fare né con i processi di prova, né con le inferenze da tali generalizzazioni. In tal modo il calcolo delle probabilità ed i processi d'inferenza, deduttiva o probabilistica, vengono nettamente distinti e separati dall'induzione. Quest'ultima, come semplice *scoperta* (ma sarebbe più esatto chiamarla «invenzione»), pare a Venn poter contare sull'avallo della stessa esperienza, la quale rivelerebbe, come *dato di fatto,* l'esistenza di leggi generali.

Mentre Boole e Venn perseguono in modo assai poco convincente il loro ideale di certezza oggettiva, W.S. Jevons[27] non teme di proclamare con decisione il carattere sempre ipotetico dell'induzione, che per lui si identifica con l'intero metodo ipotetico-deduttivo, comprendente tanto il momento della scoperta quanto quello della prova: «Così ci sono tre passi nel processo di induzione: 1) costruire un'ipotesi che abbia il carattere di legge generale; 2) dedurre conseguenze da questa legge; 3) osservare se le conseguenze concordano con i fatti particolari»[28].

[25] G. Boole (1815-1864), *An Investigation of the Laws of Thougth,* London 1854; trad. it. *Indagine sulle leggi del pensiero,* Torino 1976.

[26] J. Venn, *The Logic of Chance,* London 1866.

[27] W.S. Jevons (1835-1882), *The Principles of Science*, London 1873 e 1877; rist. New York 1958.

[28] Ivi, pp. 265 sgg.

A fine secolo K. Pearson[29], delineando una concezione della conoscenza in cui egli cerca di mediare alcune istanze provenienti dal kantismo con quelle di un empirismo intransigente, preciserà il suo senso dell'induzione, probabilmente influenzato da Whewell, come *scoperta* di una *formula* che permetta di riassumere brevemente l'intero campo di fatti osservati.

Al fascino di Whewell non sfugge neppure E. Mach:

> L'operazione psichica che fa ottenere nuove conoscenze, per lo più designata col termine inadatto di «induzione», non è un processo logico, anche se può contenere processi logici come membri intermedi e ausiliari. Il lavoro principale nel reperimento di nuove conoscenze spetta all'astrazione ed all'attività della fantasia. La circostanza – sottolineata dallo stesso Whewell – che qui il metodo può far poco, rende intelligibile anche il carattere misterioso che è tipico, come nota Whewell, delle scoperte cosiddette «induttive». [...] In mancanza di un metodo sufficiente, che faccia da guida alle scoperte scientifiche, tali scoperte appaiono, se sono felici, come un'opera artistica: lo hanno visto assai bene Johannes Müller, Liebig e altri[30].

La distinzione tra un senso dell'induzione quale scoperta ed uno quale inferenza, o prova, necessita tuttavia ancora di qualche chiarimento. Sembra caratterizzare il primo senso una concezione intuizionistica, quasi mistica, della «scoperta», affidata esclusivamente alla geniale inventiva dello scienziato.

Neppure la definizione dell'induzione come 'inferenza' appare tuttavia priva di ambiguità, perché tutt'altro che univoco è il senso di quell'*inferenza*. Essa oscilla infatti, come si è visto, dal senso di 'prova' empirica, come in Mill, a quello, decisamente snaturato, di inferenza puramente deduttiva, da generalizzazioni stabilite non si sa bene in qual modo, come in Boole e in Venn.

Se tuttavia consideriamo questi due sensi con riferimento al metodo ipotetico-deduttivo, sembra di poter identificare il senso dell'induzione come scoperta con la sola prima fase del processo, mentre il senso dell'induzione come inferenza o prova comprenderebbe tutte le altre. Ciò vale, in effetti, per la maggioranza degli studiosi dell'induzione. Fa tuttavia eccezione Mill, il quale, pur difendendo un senso inferenziale ed empirico dell'induzione, di contro ad un senso «mistico» ed inventivo, non intende affatto rinunciare con ciò al senso peculiare dell'induzione che è per lui quello della scoperta. Unico forse fra tanti, Mill ha inteso difendere un senso della scoperta non affidata all'intuizione geniale, ma guidata dal-

[29] K. Pearson (1857- 1936), *The Grammar of Science*, London 1892.

[30] E. Mach (1838-1916), *Erkenntnis und Irrtum*, Leipzig 1905; trad. it. *Conoscenza ed errore*, Torino 1982, pp. 312 sgg.

l'applicazione corretta di metodi induttivi. I suoi quattro metodi (della concordanza, della differenza, dei residui, delle variazioni concomitanti) sono propriamente metodi di prova per la *scoperta* di generalizzazioni causali. Si tratta, più specificatamente, di metodi di eliminazione per la scoperta di connessioni invariabili tra circostanze antecedenti ed effetti. Tali metodi hanno gradi differenti di valore probatorio, perché la certezza che un antecedente sia la causa dell'effetto studiato può essere raggiunta soltanto quando sia possibile arrivare a riprodurre artificialmente l'effetto (a ciò può arrivare in pratica soltanto il metodo della differenza). Altrimenti, permane sempre il dubbio che un antecedente invariabile possa non essere anche 'incondizionale', ma precedere l'effetto nello stesso modo in cui il giorno precede la notte, senza esserne, cioè, la causa[31].

Quanto al metodo ipotetico-deduttivo, che Mill chiama semplicemente *deduttivo,* esso svolge nella sua concezione un doppio ruolo: da un lato, costituisce il metodo di prova a cui è necessario ricorrere quando, per la complessità e per la compresenza di molte cause, appare opportuno procedere attraverso l'assunzione di ipotesi; dall'altro, costituisce il metodo peculiare dei livelli più alti della conoscenza, quando, ad esempio, dal piano più basso della generalizzazione empirica ci si eleva verso ipotesi esplicative di più ampia estensione e di maggior grado di astrazione.

Possiamo finalmente riassumere i risultati della nostra disamina indicando i seguenti quattro sensi del termine «induzione»:

a) induzione come scoperta, affidata interamente alla geniale inventiva ed all'intuito dello scienziato;
b) induzione come scoperta guidata da metodi di prova;
c) induzione come giustificazione (prova) di ipotesi già formulate;
d) induzione come inferenza probabilistica.

L'ultimo senso è quello connesso con l'idea di conoscenza probabile ed è il senso implicito nella concezione di Laplace e, oggi, nell'approccio bayesiano all'induzione.

1.5 Tendenze contemporanee

Nel XX secolo si assiste alla indissolubile saldatura tra induzione e probabilità, sotto il convergere di sviluppi di diversa provenienza.

Nella scienza, l'affermarsi di teorie indeterministiche mette in luce il ruolo sempre più preponderante delle leggi statistiche, tanto che queste hanno finito, se non per soppiantarle, almeno per relegare le ipotesi di forma universale ad una sorta di proprio caso limite.

[31] Cfr. J.S. Mill, *System of Logic*, trad. it. cit., pp. 334 sgg.

La presenza massiccia di leggi statistiche, d'altra parte, è un incentivo all'elaborazione di nuovi metodi statistici ed allo sviluppo della statistica come disciplina autonoma che, accanto ad una funzione 'descrittiva', che le era già peculiare, va ora assumendo anche una funzione inferenziale e induttiva.

Nel pensiero filosofico, il processo avviato da Neurath all'interno del neopositivismo, con l'attribuzione ad ogni enunciato empirico di un valore esclusivamente ipotetico, e sviluppato da Carnap con la cosiddetta «liberalizzazione dell'empirismo», significò, assieme alla rinuncia delle tematiche più tipiche del primo neopositivismo, quale la ricerca di un fondamento assoluto della conoscenza nell'esperienza e l'unificazione del linguaggio della scienza mediante la traduzione completa di ogni asserto scientifico in asserti osservativamente 'verificabili', la fine di ogni tentativo di «costruzione logica del mondo». Ma ha significato soprattutto, per quanto ci riguarda, l'assunzione del metodo scientifico, basato sulla 'conferma', come criterio di significanza conoscitiva e di unità della scienza. Lo stesso Carnap ha successivamente identificato la logica induttiva con la teoria della probabilità logica, la quale vuole essere, a sua volta, una *ricostruzione razionale* «del pensiero induttivo quale è abitualmente applicato nella vita quotidiana e nella scienza»[32].

Anche per Reichenbach, «lo studio dell'inferenza induttiva rientra nella teoria della probabilità, dato che i fatti empirici possono rendere una teoria probabile, ma mai assolutamente certa»[33]. Poiché la conoscenza procede per prove ed errori attraverso *assunzioni,* delle quali non conosciamo la verità ma solo la probabilità, la teoria delle probabilità permette, secondo Reichenbach, di dare una rigorosa espressione matematica al metodo per prove ed errori, guidando la nostra scelta verso quelle assunzioni che risultino «più probabili».

Che le nostre assunzioni siano probabili, del resto, è anche tutto ciò che si può esigere; pertanto, la teoria delle probabilità, permettendoci di valutarle, ci offre un metodo per *giustificare* le nostre assunzioni. Con ciò, Reichenbach perviene a denunciare una diffusa, quanto pericolosa, confusione nell'interpretazione dell'induzione: poiché lo scienziato che scopre una nuova teoria, pur guidato da congetture, non può dire con quale metodo egli sia pervenuto alla sua scoperta, certuni ne hanno tratto la conseguenza che l'induzione non sia affatto un'inferenza logica tra fatti e teorie, dimenticando che lo scienziato è tuttavia in grado di giustificare la sua teoria alla luce di prove empiriche, senza le quali egli neppure renderebbe pubblici i risultati della propria ricerca.

[32] R. Carnap, *The Logical Foundations of Probability*, 1950 e 1962, Chicago 1967, p. 576.

[33] H. Reichenbach, *The Rise of Scientifico Philosophy*, Los Angeles 1951; trad. it. *La nascita della filosofia scientifica*, Bologna 1961, p. 224.

È in questa pretesa di giustificazione che lo scienziato ricorre all'inferenza induttiva: «l'inferenza induttiva serve non per scoprire teorie, bensì per giustificarle in termini di dati d'osservazione»[34].

Perciò l'interpretazione 'mistica' dell'induzione ha origine, secondo Reichenbach, dalla confusione tra il *contesto della scoperta* ed il *contesto della giustificazione,* il primo effettivamente irriducibile, anche per Reichenbach, a criteri metodologici, il secondo invece soggetto alla dottrina dell'induzione, cioè alla teoria della probabilità. Anche Popper[35] è del parere che la scoperta non abbia un proprio metodo, ma che sia affidata alla genialità del ricercatore; perciò è possibile parlare soltanto dei metodi di controllo delle ipotesi. Tuttavia, per Popper, a differenza di Reichenbach, l'induzione non esiste ed il metodo conoscitivo valido poggia sul tentativo di falsificare le congetture che, man mano, vengono formulate.

La posizione di Popper, la cui importanza sul piano metodologico è stata decisamente sopravvalutata, si inquadra, relativamente al metodo, entro lo schema ipotetico-deduttivo: ad una congettura, che verrebbe formulata inizialmente senza nessuna giustificazione logica o empirica, dovrebbe seguire un rigoroso controllo empirico delle proposizioni da essa *deduttivamente* ricavate.

La novità di Popper, rispetto al metodo ipotetico-deduttivo vero e proprio, sta nell'aver sostituito il controllo di verifica dell'ipotesi con un procedimento mirante alla sua *falsificazione.* All'origine di ciò sta l'ovvia constatazione che nessun enunciato universale può mai essere verificato, mentre può essere falsificato da una sola istanza negativa. L'incertezza della inferenza induttiva può, cioè, essere soppiantata dalla certezza di un rigoroso ragionamento deduttivo, quel ragionamento noto fin dall'antichità come *modus tollens:* da $p \rightarrow q$ e $\neg q$ segue necessariamente $\neg p$. La proposta metodologica di Popper presta tuttavia il fianco ad almeno due obiezioni. La prima riguarda la falsifìcabilità di ipotesi statistiche qualsiasi; queste, infatti, per loro natura, non possono mai dirsi, a rigore, falsificate, poiché l'evento improbabile è pur sempre possibile. Naturalmente, le ipotesi statistiche possono essere rese sempre meno attendibili, mediante la ripetuta osservazione di eventi improbabili subordinatamente ad esse, dal momento che una congiunzione di eventi è sempre meno probabile di un accadimento singolo. Occorrerebbe, perciò, un criterio per valutare quante osservazioni 'improbabili' occorrerebbero per sostenere un eventuale giudizio di falsificazione. Tale metodo, però, sarebbe comunque, di nuovo, di natura induttiva e non deduttiva.

[34] Ivi, p. 223.

[35] K. Popper, *The Logic of Scientific Discovery,* London 1934; trad. it. *Logica della scoperta scientifica*, Torino 1970.

La seconda obiezione, connessa con la prima, è la seguente: il falsificazionismo popperiano, così come è stato enunciato dal suo autore, non è in grado di fornire criteri metodologici *effettivi* per programmare esperimenti volti alla falsificazione di ipotesi 'qualsiasi'. Neppure le ipotesi deterministiche, che, come si è detto, godono della caratteristica della falsificabilità, possono essere inserite in un programma sperimentale, mirato alla loro falsificazione, restando invece affidate all'accidentalità dell'eventuale osservazione confutante. Un'idea falsificazionistica, invero, è stata tradotta in termini di rigoroso metodo statistico da R.A. Fisher, ma tale metodo, come vedremo in seguito, si riferisce soltanto all'eliminazione di correlazioni accidentali e poco ha a che fare con il falsificazionismo popperiano, che è connesso con la questione della demarcazione tra scienza e non scienza, e che vuol riferirsi ad ogni tipo di ipotesi[36].

Il relegamento, pressoché unanimemente condiviso, della 'scoperta' al di fuori della possibilità di una indagine logica e metodologica, ha equivalso, di fatto, alla negazione di una duplicità di aspetti, o momenti, della metodologia induttiva, poiché ha portato ad assumere che ogni discorso 'sensato' sull'induzione, ogni inferenza induttiva e perfino il confronto tra gli stessi metodi induttivi, non possa rientrare che nel contesto di giustificazione delle ipotesi. Ne consegue che i metodi di inferenza induttiva vengono ad essere tutti collocati su uno stesso piano, quali possibili concorrenti, o rivali, nel processo di accreditamento di un'ipotesi. Così, procedimenti volti all'analisi ed alla valutazione dei dati empirici, in relazione ad un contesto di scoperta, come ad esempio la teoria della correlazione o l'analisi della varianza, vengono completamente ignorati, oppure fraintesi, come accade per i test di significatività.

Ma su ciò torneremo in seguito. Basterà, per ora, aver richiamato l'attenzione sul fatto che, in epoca a noi contemporanea, l'induzione è stata circoscritta al solo contesto della giustificazione delle ipotesi, escludendone del tutto la scoperta, come contesto non suscettibile di metodo. Questa sottovalutazione del contesto di scoperta nella metodologia della scienza empirica è, in gran parte, responsabile di molte inutili polemiche tra approcci differenti.

Ancora due parole sui vari tipi di inferenza induttiva. La centralità del concetto di probabilità nella metodologia scientifica odierna ha comportato l'estensione del significato dell'espressione «inferenza induttiva»

[36] D. Gillies, nel tentativo di dare concretezza al falsificazionismo popperiano, ha elaborato un metodo logico-statistico di falsificazione, ma anche questo si applica soltanto a quelle ipotesi che hanno la stessa struttura formale dell'ipotesi-nulla di Fisher e non a ipotesi qualsiasi. Cfr. D. Gillies, *An Objective Theory of Probability*, London 1973. La fondamentale differenza, sia sul piano epistemologico che su quello metodologico, tra l'impostazione di Fisher e quella di Popper, sfugge tuttavia a Gillies, il quale parla di un comune approccio Popper-Fisher.

rispetto a quello originario, praticamente limitato alla generalizzazione. Un notevole ampliamento di significato si era già verificato con Mill, e ancor più con Laplace, quando è stata posta nell'inferenza induttiva anche l'inferenza previsiva da fatti noti ad uno, o più, fatti ignoti. Oggi, il concetto di inferenza induttiva abbraccia ogni inferenza 'dal noto all'ignoto', per dirla con Mill, e la teoria delle probabilità si propone quale 'logica dell'incerto', per dirla con de Finetti.

L'inferenza *inversa* (che rappresenta il senso originario dell'induzione, da fatti noti alla generalizzazione, universale o statistica) non è che una delle inferenze induttive; accanto ad essa si può parlare di una inferenza *previsiva* (da fatti noti ad altri fatti non noti) e di un'inferenza *diretta* (da un universo noto ad un suo campione casuale).

1.6 Leggi e teorie

Nel riferirsi alle formulazioni scientifiche abbiamo finora preferito usare, per lo più, il termine «generalizzazione». Nella sua genericità tale termine è infatti onnicomprensivo perché tutti gli asserti propriamente scientifici pretendono di essere 'generali'. Ma ora è forse giunto il momento, a chiusura di questo capitolo introduttivo, di chiederci quali siano gli asserti generali che la scienza è interessata a formulare e se tutti possano esser considerati confermabili attraverso l'uso di metodi di inferenza induttiva.

Torneremo successivamente sull'argomento, cercando di trarre anche qualche conclusione; per il momento si tratta soltanto di sollecitare la riflessione su alcuni aspetti e problemi.

I filosofi e i logici della scienza si sono generalmente soffermati ad esaminare una sola forma di enunciato, quella della generalizzazione empirica universale, il cui esempio più abusato è l'enunciato «tutti i corvi sono neri». Solo in tempi assai più recenti, essi hanno esteso la loro attenzione anche agli asserti generali di forma statistica. Le 'leggi di natura' sono generalmente identificate con generalizzazioni universali, di contenuto empirico, mentre il termine «ipotesi» sta in genere ad indicare l'incertezza circa la loro effettiva validità. L'atteggiamento radicalmente positivistico del pensiero filosofico-scientifico, dalla metà dell'Ottocento fino ai nostri giorni, ha influito in direzione di una generale indifferenziazione tra gli asserti scientifici stessi, poiché, in definitiva, ogni enunciato conoscitivamente sensato doveva essere una *descrizione,* vera o falsa, di fenomeni empirici, sebbene gli enunciati universali ponessero pesanti interrogativi circa la loro verificabilità e, quindi, circa la loro natura.

Il pensiero scientifico ha visto la sua prima origine in problemi suggeriti dall'osservazione di eventi di esperienza comune e, in questi eventi, esso ha introdotto un ordine sistematico attraverso la classificazione e la generalizzazione empirica. Con generalizzazione empirica si intendono proprio asserti del tipo «tutti i corvi sono neri», cioè una mera correlazione di eventi, o di proprietà, osservati costantemente insieme, o secondo una regolarità statistica. Questo tipo di asserti ha caratterizzato il primo stadio della conoscenza e costituisce la forma più elementare e più primitiva di 'leggi' empiriche: una riscontrata uniformità, descrivendo un aspetto della realtà, acquistava immediatamente il valore di una 'regola' cui le singole istanze dovevano sottostare.

La legge scientifica al suo primo livello non è, pertanto, che una descrizione generalizzata di esperienze. Ma nel procedere, la ricerca scientifica è andata elevandosi dal piano empirico più immediato: ha formulato leggi sempre più astrattive, e non più mere generalizzazioni empiriche, e leggi esplicative di leggi, introducendo, tra l'altro, concetti non denotanti entità direttamente osservabili (concetti teorici) e costruendo strutture formali per tali concetti.

La scienza teorica si presenta, a questo punto, come un insieme di sistemi di stratificazioni logiche sovrapposte di enunciati generali e di leggi: dalle generalizzazioni empiriche, che si trovano al piano più basso, a diretto contatto con i fenomeni empirici, si passa a leggi quali le tre leggi di Keplero sul movimento dei pianeti o la legge della caduta dei gravi, e quindi ad altre leggi ancor più generali quale quella di Boyle circa il rapporto tra pressione, temperatura e volume dei gas, per arrivare, attraverso livelli intermedi, a leggi come quelle del moto di Newton, o quella di Maxwell, o i principi dell'elettromagnetismo, e così via.

Nel suo desiderio di una sempre maggiore comprensione dei fenomeni, la scienza si è volta a cercare 'spiegazioni' di quelle regolarità già espresse nelle leggi empiriche, giungendo ad elaborare 'teorie' che permettessero di raccogliere e di spiegare una gran quantità di leggi empiriche qualitativamente differenti, unificando campi di fenomeni molto dissimili tra loro[37].

Questa visione gerarchica di fatti, generalizzazioni empiriche, leggi e teorie fa parte ormai di una tradizione, nata dalle analisi dell'empirismo logico (Carnap, Hempel, ecc.), che trova ampio consenso. Ciò ha portato a dotare le teorie scientifiche di una struttura complessa che non può essere ridotta alla forma degli enunciati esprimenti leggi empiriche. Nasce, perciò, il rilevante problema se anche le teorie scientifiche siano provate ed accettate per via induttiva, come le altre descrizioni del mondo, oppure se il processo conoscitivo della scienza percorra anche strade diverse, non riconducibili alla metodologia induttiva e probabilistica.

[37] La concezione empiristica della scienza, nel perseguire l'istanza di una radicale eliminazione della metafisica dal contesto scientifico, ha per lungo tempo negato una funzione esplicativa che non fosse interamente contenuta nella descrizione generalizzata dei fenomeni stessi: vale a dire che questo corvo è nero *perché* tutti i corvi sono neri!

Di fronte alla problematicità di domande del tipo «qual è il grado di probabilità della teoria della relatività?», nasce il dubbio che occorra stabilire un limite per l'applicabilità dei metodi di inferenza induttiva. Tale limite equivarrebbe all'abbattimento di una costrizione, troppo spesso imposta alla ricerca scientifica, e radicata nella stessa nozione di razionalità. Probabilmente, è giunto il momento di riconoscere alla scienza un compito 'costitutivo' di linguaggi e di sistemi teorici, la cui progettazione non debba essere limitata a priori da rigidi canoni di adeguamento ad un piano pregiudiziale di esperienza empirica, il cui senso lo stesso sviluppo della scienza ha ormai reso problematico. Ciò comporta una ridefinizione del significato della razionalità scientifica e della stessa conoscenza.

Detto ciò, resta comunque il compito di indagare tra i diversi tipi di enunciati generali che sembrano rientrare nel campo di applicazione della metodologia induttiva; ed anche qui si deve ammettere che non sempre è facile determinare dove finisca una generalizzazione empirica e dove inizi una legge.

«Tutti gli uomini sono mortali», ad esempio, è una generalizzazione empirica o una legge? Se è una generalizzazione, lo è nello stesso modo in cui lo è l'enunciato «tutte le palline di quest'urna sono bianche»? La questione sembra riposare sul senso ambiguo di quel «tutti» che, se riferito ad una popolazione finita, sembra identificarsi con una estensione, una quantità (ad esempio una frequenza o una proporzione: come dire che il 100% degli individui ha la tal proprietà), mentre, se riferito ad un universo infinito, sembra sottolineare l'impossibilità di eccezioni nella connessione tra i due predicati (o classi). Hintikka, ad esempio, fa valere nella sua logica induttiva, proprio l'idea di una completa indipendenza delle leggi dall'ampiezza della popolazione di riferimento[38].

Vi è dunque una differenza qualitativa di cui tener conto, anche tra enunciati generali formalmente simili, e tale differenza sembra valere, non solo tra enunciati di forma universale (quelli, cioè, con il quantificatore «tutti»), ma anche tra quelli di forma statistica. Infatti, anche tra questi ultimi è possibile distinguere enunciati tendenti ad evidenziare l'esistenza di una correlazione statistica, con implicito riferimento ad un universo

[38] La teoria delle generalizzazioni induttive di Hintikka si ripromette, tra l'altro, di risolvere lo 'scandalo' della zero-conferma, cioè del valore di probabilità uguale a zero che tutti gli enunciati universali in linguaggi infiniti ricevono entro la logica induttiva carnapiana. Per far ciò, Hintikka pone alla base della sua logica induttiva i *costituenti*, ossia tipi di mondi possibili che esprimono solo relazioni tra predicati, ma sono indipendenti dal numero degli individui dell'universo di discorso. Cfr. J. Hintikka, *Towards a Theory of Inductive Generalization,* in Y. Bar-Hillel, *Logic, Methodology and Philosophy of Science,* Amsterdam 1965; *A two-dimensional Continuum of Inductive Methods,* in J. Hintikka, P. Suppes (a cura di), *Aspects of inductive Logic*, Amsterdam 1966; *Induction by Enumeration and Induction by Elimination*, in I. Lakatos (a cura di), *The problem of Inductive Logic,* Amsterdam 1968.

infinito (ad esempio «il fumo è cancerogeno»), da enunciati riferentisi, invece, al valore di un parametro di una popolazione (una frequenza, una media ecc.; ad esempio «l'x% degli italiani è biondo», ma anche «tutti (=100%) i ragazzi americani leggono "Topolino"»).

Probabilmente le caratteristiche intrinseche delle ipotesi formulate debbono essere tenute presenti quando si tratta di scegliere il metodo di indagine più adeguato. Ad esempio, le ipotesi puntuali nel continuo hanno sempre un valore di probabilità uguale a zero; tuttavia, una generalizzazione empirica, vertente su un parametro quantitativo, potrebbe essere formulata come ipotesi su un intervallo di valori, e quindi affrontata con metodi di stima, anche nel caso di una popolazione estremamente grande o tendente all'infinito.

Analogamente, le caratteristiche strutturali di una teoria scientifica dovranno essere prese in considerazione per poter affrontare la questione se l'accettabilità di una teoria possa essere decisa con metodi induttivi.

La questione è complessa e richiede di essere vagliata in tutte le sue implicazioni. Il punto di vista qui assunto vuole tuttavia adombrare l'idea che sia oggi impensabile che l'elaborazione e lo sviluppo della conoscenza sia riducibile entro gli angusti confini della metodologia induttiva, sebbene questa costituisca uno strumento indispensabile e prezioso per la ricerca scientifica, così come per settori di interesse più pratico e circoscritto.

Capitolo 2
Probabilità

2.1 Origini del calcolo delle probabilità

Il calcolo delle probabilità ebbe origine, come è noto, in riferimento ai giochi d'azzardo, verso la metà del Seicento. I. Thodunter[1] riferisce tuttavia di valutazioni delle probabilità di lanci con tre dadi già in un *Commentario alla Divina Commedia* del 1477, e Ch. Gouraud[2] ha rintracciato i primi accenni a calcoli combinatori addirittura in un poema latino del Basso Impero, intitolato *De vetula*.

Anche Galileo[3] scrisse un'opera sui giochi dei dadi; tuttavia si è soliti far risalire l'origine del calcolo delle probabilità a certe questioni di scommessa, poste dal Cavaliere de Mère a Pascal, e da questi discusse con Fermat[4]. La più antica trattazione sulle combinazioni pare sia dovuta, appunto, a B. Pascal.

Tutta la seconda metà del XVII secolo, tuttavia, vede un fiorire di trattatelli sull'argomento, tra i quali va ricordata anche la *Dissertatio de arte combinatoria* (1666) di Leibniz. Merita di essere ricordato, come il più antico trattato in lingua inglese sull'argomento, anche il

[1] I. Todhunter, *A History of the Mathematical Theory of Probabilità,* Cambridge 1865.

[2] Ch. Gouraud, *Histoire du calcul des probabilitès depuis ses origines jusqu'à nos jours*, Paris 1848.

[3] G. Galilei, *Considerazioni sopra il giuoco dei dadi,* data incerta, ovviamente anteriore al 1642.

[4] Restano alcune lettere del 1654. Il problema in questione, chiamato il «problema dei punti», è così enunciato: due giocatori necessitano ciascuno di un dato numero di punti per vincere. Se essi si lasciano senza terminare il giuoco, come dovrebbero essere divise le poste? La questione corrisponde a chiedersi quale sia la probabilità di ciascuno dei giocatori, ad un dato momento del giuoco, di vincere la partita.

Maria Grazia Sandrini, *Filosofia dei metodi induttivi e logica della ricerca*, ISBN 978-88-6453-034-5 (print) ISBN 978-88-6453-037-6 (online)

Discourse of Combinations, Alternatives, and Aliquot Parts (aggiunto come *Appendice* alla sua *Algebra*), di John Wallis[5], forse ispirato dalla regola delle combinazioni esposta nell'*Aritmetica memorativa* (1567) di W. Duckley.

La prima opera sistematica sulla probabilità in giochi di dadi può essere considerata l'opera di Christiaan Huygens, *De ratiocinis in ludo aleae,* forse del 1657, che Jakob Bernoulli pose come prima parte del suo famoso trattato, l'*Ars conjectandi,* rimasto purtroppo incompiuto nella sua quarta parte, a causa della morte dell'autore (1705), e pubblicato postumo, nel 1713, a cura del nipote Nicola.

Nei primi anni del secolo XVIII si collocano ancora due opere importanti, l'*Essai d'analyse sur les jeux d'hazar*d (1708) di Pierre-Rémond de Montmort, ed il noto trattato di Abraham De Moivre, *Tbe Doctrine of Chance,* la cui prima edizione, del 1718, recava una dedica a Newton, allora presidente della Royal Society, e di cui una seconda edizione, ampliata, apparve nel 1738, con la dedica a Lord Carpenter. Una terza edizione, postuma, è del 1756.

Ancora da segnalare è l'opera più tarda di Thomas Simpson, *Tbe Nature and Laws of Chance,* del 1740, nonché l'opera costante dei Bernoulli, famiglia di matematici illustri, fra i quali anche Nicola e Daniele dettero contributi importanti al calcolo delle probabilità. Di quest'ultimo è l'opera *Specimen tbeoriae novae de mensura sortis,* pubblicato nel 1730[6].

Le prime estensioni dell'applicazione del calcolo delle probabilità, al di fuori del campo del gioco d'azzardo, furono rivolte al campo «morale», cioè a quel settore dell'attività umana che non sembrava subordinabile a quell'idea di certezza assoluta perseguita nelle scienze fisiche e matematiche. Certamente influente era tuttora la distinzione cartesiana tra una *evidentia mathematíca* ed una *evidentia moralis,* una scienza intesa come cognizione certa ed evidente, fondata su metodi rigorosamente deduttivi, ed il campo incerto della vita pratica, cioè delle cose civili, giuridiche, economiche. Lo stesso Jakob Bernoulli dedicò la quarta parte dell'*Ars conjectandi* all'uso ed all'applicazione della dottrina delle probabilità alle «cose civili, morali ed economiche». Va anche ricordato, a questo proposito, che data già dal XVII secolo l'uso della compilazione di statistiche,

[5] J. Wallis (1616-1703) fu uno dei fondatori della «Royal Society of London». Scrisse sui più disparati argomenti scientifici: la sua opera più vasta, *Aritmetica Infinitorum* (Oxford 1655) aprì la strada al calcolo infinitesimale di Newton (compare qui per la prima volta il simbolo ∞); la sua *Algebra* (London 1685 e 1693) è ritenuta il primo serio tentativo, in Inghilterra, di scrivere una storia della matematica.

[6] Daniele Bernoulli è noto specialmente per la sua teoria dell'aspettativa morale e per il cosiddetto problema di Pietroburgo, sul quale si sono accaniti a lungo matematici insigni.

in specie delle tavole di mortalità, usate soprattutto per stimare i premi delle assicurazioni sulla vita[7].

Molti sono gli scritti volti al tentativo di applicare il calcolo delle probabilità alla valutazione del grado di credibilità delle testimonianze nei processi giuridici. È di Nicola Bernoulli un trattato dal titolo *Specimena artis conjectandi ad quaestiones juris applicatae,* del 1733; ma già nel 1699 Craig aveva scritto *Theologiae christianae principia mathematica,* in cui egli si occupava delle probabilità di verità delle testimonianze; suo è forse anche il saggio, dello stesso anno, *A Calculation of tbe Credibility of Human Testimony*[8].

Così, il calcolo delle probabilità si andava affermando come guida all'azione, come strumento per il calcolo matematico del grado di credenza in situazioni di incertezza morale.

Nella prima edizione (1718) del suo trattato, *Tbe Doctrine of Chances,* A. De Moivre riconosce ad Huygens il merito di essere stato il primo ad aver fornito una qualche sistematizzazione ad un certo numero di problemi relativi al giuoco d'azzardo, ponendo, così, le prime basi del calcolo delle probabilità. Il trattato di Huygens, *De Ratiociniis in ludo aleae,* consiste in 14 Proposizioni, in ciascuna delle quali è affrontato e risolto un problema di gioco, in ordine di difficoltà crescente. Alla fine del trattato sono presentati cinque problemi irrisolti, le cui soluzioni saranno poi fornite da J. Bernoulli.

Anche il trattato di De Moivre, sebbene assai più ampio, segue lo stesso schema: dopo una breve introduzione, nella quale vengono esposte le regole e le nozioni basilari del calcolo delle probabilità, il trattato si sviluppa attraverso 53 problemi di gioco, di ciascuno dei quali è offerta la soluzione e la dimostrazione. Questi tipi di problemi richiedono, per la loro soluzione, il ricorso alla teoria delle permutazioni e delle combinazioni, che difatti è introdotta, ma non in modo sistematico.

Un approccio più sistematico e più generale presenta invece l'*Ars Conjectandi* di Jakob Bernoulli. Il trattato si divide in quattro parti: la prima riproduce, commentandolo, il trattato di Huygens; la seconda è dedicata alla teoria delle permutazioni e delle combinazioni, sistematicamente esposta; la terza consiste di soluzioni a vari problemi di giochi d'azzardo; la quarta parte, infine, incompiuta, si propone di estendere la

[7] J. Graunt registrò le morti a Londra dal 1592 al 1603; J. de Witt, *La valeur des rentes viagères en raison de ventes libres ou remborsables,* 1671; E. Halley, *An Estimate of the degrees of the Mortality of man kind, drawn from curious Tables of the birth and funerals at the city of Breslaw,* 1693. Numerosi studi statistici per le assicurazioni furono pubblicati in Olanda.

[8] A titolo di curiosità, vale la pena di ricordare anche il saggio di J. Arbuthnot, *An argument for Divine Providence, taken from the constant Regularity observed in the Births of both sexes,* 1710-12.

teoria stessa ad applicazioni in campo morale ed economico. In quest'ultima è contenuta anche la dimostrazione del famosissimo teorema, noto, appunto, come teorema di Bernoulli.

Può destare sorpresa il fatto che, nella prima edizione del suo trattato, uscita ben cinque anni dopo la pubblicazione dell'opera di Bernoulli, De Moivre non abbia fatto alcun riferimento agli importanti risultati ivi raggiunti. In realtà, la prima edizione di tale opera non è altro che la traduzione inglese di un precedente studio scritto in latino, *De mensura sortis,* apparsa nel 1711, e quindi antecedente alla pubblicazione del trattato di Bernoulli. Nella *Prefazione*, tuttavia, De Moivre inserì l'invito a Nicola Bernoulli, affinché proseguisse l'opera incompiuta dello zio; compito al quale, dice De Moivre, era stato chiamato lui stesso, ma per il quale dichiara di sentirsi inadeguato. Nella seconda edizione del trattato, apparsa a vent'anni di distanza, l'opera di J. Bernoulli è invece ben presente: De Moivre aggiungerà una intera sezione dedicata al tentativo di risolvere l'inverso del teorema di Bernoulli (lo stesso compito affrontato successivamente da Bayes), del quale presenterà anche una estensione che, avvalendosi di un teorema sviluppato nel frattempo da Stirling, ne semplifica molto la dimostrazione.

L'ambito del gioco d'azzardo, a cui costantemente si faceva riferimento, correlava strettamente il concetto classico di probabilità alle situazioni di scommessa ed a nozioni quali *aspettativa, odds*[9], e simili.

Fondamentalmente, la nozione di probabilità si presenta, in questi autori, piuttosto unitaria, anche se non del tutto uniforme.

De Moivre dà della probabilità una definizione puramente matematica, che è, di fatto, universalmente condivisa: «la probabilità di un evento è maggiore o minore a seconda del numero dei casi nei quali esso può accadere, rapportato al numero totale dei casi nei quali esso può accadere o non accadere»[10].

Altri, tuttavia, vogliono anche precisare ciò che questo rapporto matematico esprime. Così Jakob Bernoulli, che afferma: «La probabilità è il grado di certezza, e da questa differisce come la parte dal tutto»[11]. Bernoulli precisa anche che la certezza, di cui egli parla, è una certezza soggettiva; infatti, egli dice, la certezza delle cose può essere valutata sia da un punto di vista oggettivo, o in sé, sia da un punto di vista soggettivo,

[9] L'*aspettativa* corrisponde alla posta da pagare in una situazione di equa scommessa. Gli *odds* indicano il rapporto di scommessa. Essi sono proporzionali alla probabilità. Ad esempio, se un evento ha la probabilità pari a 0,9 di verificarsi, gli odds sono 9 a 1 a favore del suo accadere; se l'evento si verificherà, la vincita sarà di 1, 10, 100, ..., a seconda che si sia pagato 10, 100, 1000,...

[10] A. De Moivre (1667-1754), *The Doctrine of Chance,* 1° ed. , London 1718, p. 1.

[11] J. Bernoulli (1654-1705), *Ars Conjectandi*, Basel 1713, parte IV, p. 211.

ovvero con riferimento a noi. Ora, non c'è dubbio per Bernoulli che «tutte le cose che sono o saranno sotto il sole, passate, presenti o future, in sé e oggettivamente hanno sempre somma certezza»[12]. Ma, considerata in ordine a noi, la certezza delle cose non è sempre la stessa, ma varia in più e in meno. Le cose delle quali abbiamo esperienza sensibile e cognizione razionale, e della cui esistenza non è possibile perciò dubitare, godono di somma e assoluta certezza. Ma ve ne sono altre che ottengono una misura di certezza commisurata con le probabilità, maggiori o minori, a favore della loro esistenza, passata, presente o futura che sia.

Alla distinzione tra una certezza oggettiva ed una soggettiva, si collega in Bernoulli la distinzione tra lo *scire,* o *intelligere,* e il *conjicere,* o *opinari,* cioè tra una conoscenza piena, oggettiva e assoluta delle cose, ed un congetturare, pesando numero e qualità degli argomenti pro e contro. Il calcolo delle probabilità è appunto l'arte del congetturare, cioè del ragionare correttamente circa probabilità. Così, per Bernoulli, come del resto anche per De Moivre, la teoria delle probabilità finisce per assumere un valore che trascende l'ambito del giuoco d'azzardo.

Nella prefazione alla prima edizione del suo trattato, De Moivre afferma che i risultati da lui stesso ottenuti, studiando alcuni problemi di gioco, meritano di essere considerati, «non come questione relativa soltanto al gioco, ma come contenenti alcune speculazioni generali non indegne di essere prese in considerazione da chi ami la verità». Ed aggiunge che il suo trattato può essere utile a vari scopi, fra i quali solo il primo, e ovvio, consiste nel soddisfare la curiosità di chiunque desideri conoscere su quali fondamenti poggi il gioco d'azzardo: la dottrina delle probabilità può servire, congiuntamente alle altre parti della matematica, quale «conveniente introduzione all'arte del ragionare». Infatti, essa insegna «a distinguere la verità da ciò che pare rassomigliarle»; insegna, cioè, ad evitare gli errori, cui facilmente potremmo essere indotti dall'apparente semplicità di alcuni problemi, e, in particolare, a diffidare dall'affidarsi semplicemente al naturale buon senso. La dottrina delle probabilità può altresì aiutare a non cadere in quel tipo di superstizione consistente nel credere nella fortuna, cioè nel credere che esista una qualità umana che, se posseduta, farà conseguire sempre successi, o, almeno, più successi che insuccessi. Infine, la dottrina delle probabilità può utilmente essere usata per distinguere tra eventi prodotti dal caso ed eventi prodotti secondo un piano definito, essendo in grado di provare, attraverso un graduale aumento della probabilità, che «dove c'è uniformità, ordine e costanza, là risiede anche un piano prestabilito».

Quest'ultima considerazione appare di particolare rilievo: essa si accorda con il tentativo di estendere la teoria della probabilità al di fuori

[12] Ivi, p. 210.

del circoscritto ambito di origine, quello del gioco, verso il più complesso e ampio campo dell'esperienza; estensione che trova la sua prima importante espressione matematica proprio nel teorema di Bernoulli.

Il punto di partenza della riflessione di Jakob Bernoulli sta in una constatazione semplice e ovvia: per calcolare le probabilità è necessario conoscere il numero dei casi, sia quello dei casi a favore, sia quello totale, perché il valore numerico della probabilità non è altro che questo rapporto. Per formare congetture attorno a qualsiasi cosa, egli dice, nient'altro è richiesto se non che siano determinati accuratamente i numeri dei casi in questione. Ora, vi sono situazioni, come quella del gioco d'azzardo, nelle quali il numero dei casi è sempre noto, o comunque facilmente calcolabile: così è nei giochi con dadi, con carte, e in qualsiasi altro gioco noto. Ma come è possibile calcolare, ad esempio, il numero delle malattie che possono condurre a morte un uomo, oppure il numero delle mutazioni quotidiane dell'aria, in modo da poter prevedere che tempo farà tra un mese o tra un anno? In breve, nelle cose che riguardano la vita e la natura, tante e varie sono spesso le cause latenti ed interagenti, da risultare impossibile valutare il numero dei casi che vi concorrono. Vi è però un'altra via – conclude Bernoulli -, per la quale sembra possibile soddisfare al nostro problema, e cioè che «ciò che non è dato scoprire a priori sarà almeno consentito investigare a posteriori, cioè dall'evento osservato più volte in esempi simili»[13]. Se, ad esempio, si saranno osservati trecento uomini della stessa età e complessione fisica di quelle attuali di Tizio, e si sarà osservato che duecento di essi sono morti entro il successivo decennio, si potrà concludere che vi è un numero doppio di casi, a sostegno dell'ipotesi che anche Tizio morirà entro i prossimi dieci anni, rispetto a quello dei casi a favore dell'ipotesi che egli non morirà entro tale periodo di tempo.

Questo modo empirico di determinare il numero dei casi attraverso l'osservazione, del resto, – osserva Bernoulli – non è né nuovo né insolito: tale è anzi costantemente il comportamento degli uomini nella loro vita quotidiana. La dimostrazione della correttezza di tale comportamento, però, non è affatto banale: si tratta di dimostrare se, aumentando il numero delle osservazioni, sia possibile accrescere la probabilità di conoscere l'effettiva proporzione esistente tra i casi relativi ad un dato tipo di eventi, fino a superare un qualsivoglia predeterminato grado di certezza (eventualmente fino a raggiungere la certezza morale), oppure se si dia un qualche grado di certezza che non sia mai possibile superare, per quanto si moltiplichi il numero delle osservazioni.

Si supponga dunque che in un'urna vi siano 3000 sassolini bianchi e 2000 neri, ma che si voglia stimare la loro proporzione attraverso

[13] Ivi, pp. 223 sgg.

l'osservazione, estraendo un sassolino alla volta e riponendolo quindi nuovamente nell'urna, in modo da non alterare mai la loro proporzione. Ci si chiede se sia possibile proseguire questa operazione un numero (per quanto grande) di volte, tale che sia alla fine probabile, ad un certo grado voluto (ad esempio,$1-\varepsilon$), che la proporzione osservata tra i sassolini bianchi e neri estratti rifletta l'effettiva proporzione dell'urna, e non un'altra diversa. Se non sarà così, conclude Bernoulli, bisognerà riconoscere l'impossibilità di indagare il numero dei casi attraverso gli esperimenti; ma, se sarà così, avremo trovato un modo per stimare a posteriori numero e proporzione dei casi. Ciò permetterà di guidare le nostre congetture in qualsiasi materia *contingente* della vita civile, dove il «moralmente certo» sta per l'assolutamente certo, non meno scientificamente che nel giuoco dei dadi.

Bernoulli dimostrerà quindi, avvalendosi delle leggi dell'espansione del binomio, che, posto sia $r/(r+s)$ l'effettiva proporzione dei casi relativi ad un dato evento, possono essere compiuti tanti esperimenti (quantitativamente determinabili) da rendere probabile, al grado voluto, che la frequenza relativa osservata cada tra $(r+1)/(r + s)$ e $(r -1)/(r + s)$.

Come esempio, Bernoulli assume r=30 e s=20. Egli trova quindi che per poter scommettere 1000 contro 1 (cioè con probabilità pari a 0,999), che la frequenza osservata cadrà tra 31/50 e 29/50, saranno sufficienti 25.550 esperimenti; ne occorreranno invece 31.258 per scommettere 10.000 contro 1 (cioè con probabilità pari a 0,9999), ed infine saranno necessari 36.966 esperimenti per poter scommettere 100.000 contro 1 (cioè con probabilità pari a 0,99999). Consegue da ciò che, «se si continuassero le osservazioni di tutti gli eventi per tutta l'eternità (sfumando infine la probabilità nella perfetta certezza), tutte le cose del mondo sarebbero colte secondo rapporti certi e secondo una legge costante di avvicendamento»[14].

Laplace darà il massimo rilievo a questo teorema, da lui considerato di grandissima importanza, ed osserverà che «ne deriva una conseguenza che deve essere considerata come legge generale, e cioè che i rapporti degli effetti della natura sono press'a poco costanti, quando siano considerati in gran numero». Da esso, «consegue anche che, in una serie indefinitamente prolungata di eventi, l'azione delle cause regolari e costanti deve alla lunga imporsi sull'azione delle cause irregolari»[15].

Immaginiamo, quindi, un'urna che contenga palle bianche e nere, e supponiamo di estrarre ogni volta una palla e di riporla nell'urna per procedere ad una nuova estrazione. Il rapporto del numero delle palle bianche estratte, rispetto a quello delle palle nere, estratte, sarà gene-

[14] Ivi, p. 239.
[15] P.S. Laplace, *Essai philosophique sur les probabilités,* Paris 1814, p. 42.

ralmente piuttosto irregolare nelle prime estrazioni; tuttavia, afferma Laplace, «le cause variabili di questa irregolarità producono degli effetti alternativamente favorevoli e contrari allo scorrere regolare degli eventi e, annullandosi reciprocamente nell'insieme d'un gran numero di estrazioni, lasciano sempre più scorgere il rapporto delle palle bianche rispetto alle palle nere contenute nell'urna». Pertanto, «la probabilità che il rapporto tra il numero delle palle bianche estratte e il numero totale delle palle uscite non si allontani, oltre un intervallo dato, dalla probabilità di estrarre una palla bianca in ciascuna estrazione si avvicina indefinitivamente alla certezza attraverso la moltiplicazione indefinita degli eventi, per quanto piccolo sia quest'intervallo»[16]

Al di là dell'importanza, universalmente riconosciuta, di questo teorema, l'opera di Bernoulli ha una portata forse sottovalutata. Bernoulli appare il primo, infatti, ad aver intuito consapevolmente la funzione induttiva del calcolo delle probabilità e ad aver tentato di fornire una base teorica al tentativo di estenderne l'uso al di fuori dell'ambito dei giochi d'azzardo.

Uomo del Seicento, egli appare tuttavia per molti aspetti più vicino a Laplace che a Cartesio. La concezione deterministica del suo tempo si lega in lui ad un senso per così dire relativistico della conoscenza, che permette di far posto al 'probabile' come ad una categoria scientifica del sapere. Se gli eventi del mondo sono, in sé, assolutamente determinati e necessari, la conoscenza umana di essi è tale da lasciare ampi spazi all'ignoranza: perciò, categorie quali «certo», «necessario», «contingente» non sono categorie assolute, ma categorie relative al grado di conoscenza raggiunto dall'uomo. Ad esempio, *contingente* significa, per Bernoulli, sia ciò che è libero, cioè ad arbitrio di una creatura razionale, sia ciò che è fortuito e casuale: in ogni caso, è ciò che può non essere o che avrebbe potuto non essere. Tuttavia, il contingente non esclude sempre il necessario. Molti eventi sono chiamati contingenti solo perché non si è in grado di precisarne le cause; anch'essi, però, sono determinati dalle loro cause prossime. Sembra, anzi, di poter arguire che per Bernoulli niente è in realtà non determinato e non necessario, perché ogni evento ha delle cause, per le quali esso è quello che è, e non potrebbe essere diversamente. Tuttavia, l'uscita di una carta, oppure una tempesta, sono chiamati eventi contingenti, mentre le eclissi sono annoverate tra gli eventi necessari; per Bernoulli, non vi è altra ragione di ciò, se non la scarsa conoscenza delle cause dei primi. Anche le eclissi, prima che lo studio della geometria e della fisica fosse così avanzato da permetterne la previsione, erano annoverate tra gli eventi contingenti. Il 'contingente' è, per Bernoulli, un dominio storicamente determinato: «ne segue che può essere considerato

[16] Ivi, p. 41.

contingente in un tempo ciò che in un altro tempo, dopo la conoscenza delle sue cause, diventa necessario»[17].

Pertanto, sembra non esservi, in definitiva, una gran differenza tra l'ambito del gioco e quello degli eventi del mondo e della vita. In entrambi vi è sempre assoluto determinismo, da un lato, e gradi differenti di ignoranza umana, dall'altro.

Se c'è una differenza tra le applicazioni del calcolo delle probabilità al gioco e quelle ai fatti del mondo, essa non risiede nella natura degli eventi, ma nella maggiore o minore completezza delle nostre informazioni e nel tipo di domande che a queste si connettono. Per primo, Bernoulli ha il merito di aver colto la differenza esistente tra una situazione, in cui tutto sia noto dell'universo di riferimento, ed una situazione, in cui di esso non si abbiano che le informazioni pervenuteci da singole istanze: in terminologia odierna, potremmo dire che egli ha avvertito la differenza che corre tra problemi di inferenza induttiva *diretta* e problemi di inferenza *inversa*, presentendo altresì che solo la soluzione del secondo tipo di questioni avrebbe fatto del calcolo delle probabilità uno strumento utile per finalità genuinamente conoscitive. Ciò porta a ricollegarci con una precedente osservazione: il teorema di Bernoulli, un caposaldo dell'inferenza statistica induttiva diretta, nasce tuttavia dal tentativo di trovare una risposta a problemi di inferenza inversa. La lezione, che possiamo poi trarre dal famoso teorema, pare confortare quel senso comune dell'induzione per cui si ritiene sia possibile apprendere dall'esperienza. Esso dimostra infatti come una lunga ripetizione di osservazioni analoghe accresca la probabilità che le nostre osservazioni rispecchino l'autentica struttura del reale, e come tale probabilità possa convergere alla certezza, se le osservazioni proseguono abbastanza a lungo. Sfortunatamente, tale teorema non risulta applicabile a questioni di inferenza inversa, perché il numero delle osservazioni, necessarie al raggiungimento del grado di certezza desiderato, non può essere determinato se non conoscendo il numero e la proporzione effettivi dei casi, vale a dire proprio ciò che, nell'inferenza inversa, è oggetto di indagine.

Nel 1764, comparve infine, nel volume relativo all'anno precedente dei «Philosophical Transactions of the Royal Society»[18], il saggio, destinato ad essere al centro di numerose polemiche, del rev. Thomas Bayes, *An Essay towards Solving a Problem in tbe Dottrine of Chances*, in cui è affrontato e risolto il problema dell'inferenza inversa. Con le applicazioni di tale teorema alle scienze fisiche e naturali, compiute successivamente da Laplace, si può datare l'entrata ufficiale del calcolo delle probabilità nella metodologia induttiva della scienza.

[17] Ivi, pp. 212 sgg.
[18] Vol. 53, pp. 370-418.

Il saggio fu pubblicato postumo[19], a cura di Richard Price, il quale vi ha premesso una breve presentazione, nella quale, oltre a manifestare una entusiastica ammirazione per l'opera dell'amico scomparso, osserva come la soluzione del problema della probabilità delle cause sia necessaria per fondare ogni nostro ragionamento relativo ai fatti passati ed alle previsioni sul futuro. Sebbene il senso comune sia sufficiente a suggerirci che le nostre previsioni saranno tanto più sicure quanto maggiore sarà il numero delle osservazioni a loro sostegno, tuttavia solo la soluzione del problema affrontato da Bayes, – egli afferma –, rende possibile determinare rigorosamente a quale grado esperimenti ripetuti confermino una certa conclusione. L'importanza della dimostrazione di Bayes, tuttavia, andrebbe ancora oltre, a suo parere, fino al massimo degli scopi perseguibili: essa offrirebbe, nientemeno, che un valido supporto agli argomenti a favore dell'esistenza di Dio. La regola di Bayes, infatti, permette, afferma ancora Price, di mostrare quali ragioni abbiamo per credere che gli eventi siano regolati da leggi stabili, e, pertanto, che la struttura del mondo debba discendere da una Causa intelligente.

Il problema dell'inferenza inversa è da Bayes affrontato come il problema di trovare «la probabilità che un evento accada, in date circostanze, sulla supposizione che non si conosca niente di esso se non che, nelle stesse circostanze, esso è accaduto un certo numero di volte e mancato un certo altro numero di volte».

La formula binomiale, $\binom{p+q}{q} \cdot x^p \cdot (1-x)^q$, che è alla base della dimostrazione del teorema di Bernoulli, permette di calcolare la probabilità che un dato evento accada *p* volte e fallisca *q* volte in *p+q* osservazioni, purché sia noto il valore *x* della probabilità di una singola istanza dell'evento, o, ciò che è lo stesso, l'effettiva proporzione dei casi, nella totalità dell'universo considerato. Ma, non conoscendo noi la probabilità *x*, cioè ignorando tutto, circa l'universo cui il nostro evento appartiene (potremmo immaginare che si tratti di estrazioni di palline da un'urna chiusa, la cui composizione ci sia del tutto ignota), non abbiamo neppure il modo di valutare quale peso sia da attribuire alle nostre **p + q** osservazioni.

La soluzione avanzata da Bayes è semplice, quanto geniale: se potessimo stimare la probabilità che il valore *x* cada entro un intervallo dato, avremmo trovato una regola per dare un peso alle nostre osservazioni: tutto il problema si ridurrebbe poi ad un semplice calcolo di probabilità composte.

La questione è dunque, per Bayes, quella di trovare una regola per stimare «la probabilità (*chance*) che la probabilità (*probability*) relativa all'accadere di un evento perfettamente sconosciuto cada tra due qualsiasi

[19] Sembra che Bayes (1702-1761) non abbia pubblicato in vita i suoi risultati perché dubbioso circa la validità del suo metodo.

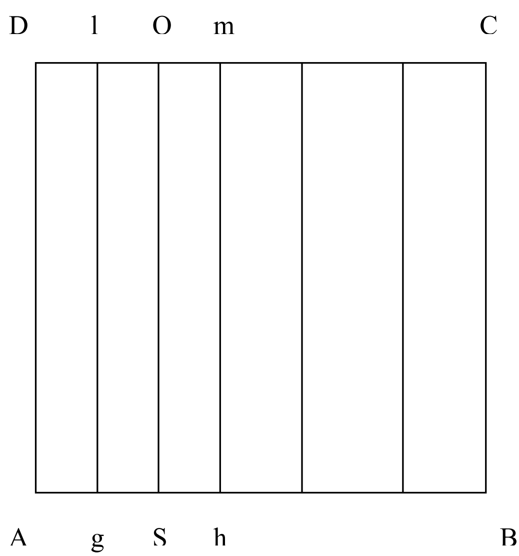
prefissati gradi di probabilità (*probability*), antecedentemente a qualsiasi esperimento».

Per risolvere questo problema, Bayes immagina che sia dato un piano *ABCD,* perfettamente liscio e regolare, tale che non vi siano motivi per cui una palla da biliardo, lanciatavi sopra a caso, debba fermarsi in un punto, anziché in un altro qualsiasi del piano, e tale anche che essa debba comunque restarvi sopra. Egli postula quindi che sia tracciata casualmente, sul piano, una parallela *OS* ad *AD.* Egli chiama «evento M» l'arresto di una palla, lanciata sul piano *ABCD,* in un qualsiasi punto compreso tra *OS* e *AD.*

Come fare, ora, per calcolare la probabilità dell'evento M?

Ai fini della soluzione vanno distinti due diversi quesiti: il primo consiste nel chiedersi quale sia la probabilità, *prima* di tracciare la linea OS, che questa cada entro due date linee, anch'esse parallele ad *AD,* diciamo *lg* ed *mh* (Fig. 3); il secondo consisterà nel calcolare la probabilità dell'evento M, sulla base della precedente valutazione.

Figura 3



Semplificato, il ragionamento di Bayes è il seguente: dividiamo il piano *ABCD* in un certo numero di rettangoli uguali i cui lati più lunghi siano tutti paralleli ad *AD.* Ora, il rettangolo *ghlm,* entro cui cade la linea *OS,* copre sul piano *ABCD* una certa porzione facilmente calcolabile. Il quoziente, tra l'area di tale porzione e l'area dell'intero piano, può essere ragionevolmente assunto, secondo Bayes, come misura della probabilità che la linea OS cada tra *lg* ed *mh*. Analogamente, la probabilità dell'evento M, una volta tracciata la linea *OS,* è espressa dal quoziente tra l'area di *ASOD* e quella di *ABCD*.

Pertanto, sulla base del solo postulato che il piano *ABCD* sia tale, che nessuno dei suoi punti risulti privilegiato rispetto agli altri (e che quindi ogni porzione in cui il piano sia suddiviso abbia a priori una uguale probabilità di essere prescelto), Bayes ritiene di poter dedurre i due fondamentali lemmi che gli permettono di affrontare e risolvere il problema dell'inferenza inversa.

Lemma 1 – La probabilità che il punto S cada tra due qualsiasi punti dati sulla linea AB è il quoziente tra la misura della distanza tra i due punti e la lunghezza dell'intera linea AB.

Lemma 2 – Una volta tracciata la linea OS, la probabilità dell'evento M è il quoziente tra AS e AB. Prima che la linea OS sia tracciata, la probabilità dell'evento M avrà un valore compreso tra il quoziente Ag/AB e il quoziente Ah/AB.

2.2 I principi del calcolo delle probabilità e la concezione classica

Il calcolo delle probabilità si basa su pochi, fondamentali, principi, comuni a tutte le «scuole» probabilistiche. Tali principi possono essere sinteticamente esposti come segue (dove $\wedge, \vee, \neg$ sono i simboli logici rispettivamente per la congiunzione, la disgiunzione e la negazione):

P_1: $0 \leq p(A) \geq 1$.

La probabilità di un evento A è compresa tra 0 e 1, estremi inclusi. Se A è un evento certo, p(A) = 1; se A è un evento impossibile, p(A) = 0.

P_2: Probabilità composte:

a) se A e B sono eventi indipendenti, allora

$$p(A \wedge B) = p(A) \cdot p\,(B)$$

b) se A e B non sono indipendenti, allora

$$p(A \wedge B) = p(A) \cdot p\,(B/A)$$

P_3: Probabilità totali:

a) se A e B sono tra loro incompatibili, allora

$$p(A \vee B) = p(A) + p(B)$$

b) se A_1, A_2, ... A_n sono eventi incompatibili, allora

$$p(A_1 \vee A_2 \vee .... \vee A_n) = \sum_{i=1}^{n} p(A_i)$$

Discendono da questi tre principi le seguenti conseguenze:

C_1 – Vale in generale che $p(A) = p(A/B) + p(A/\neg B)$
C_2 – Da P_2b deriva direttamente il seguente principio di divisione:

$$p(A / B) = \frac{p(A \wedge B)}{p(B)}$$

C_3 – Dai precedenti C_1, C_2 e da P_3 è derivabile il teorema di Bayes:

$$p(A / B) = \frac{p(A) \cdot p(B / A)}{\sum_i p(A_i) \cdot p(B / A_i)}$$

Nel suo *Essai,* Pierre Simon de Laplace, (1749-1827), poneva, come primo principio, la definizione matematica della probabilità che, come si è visto, è il *rapporto fra il numero dei casi favorevoli e quello di tutti i casi possibili.*

Lo faceva seguire da un secondo principio, nel quale egli si soffermava sulle eventuali difficoltà nella determinazione dei casi equipossibili, ai quali è necessario, comunque, ridurre ogni problema di probabilità, affinché ne sia possibile la soluzione:

> [...] ciò suppone i diversi casi egualmente possibili: se non lo sono, si determineranno in primo luogo le loro rispettive possibilità la cui giusta stima è uno dei punti più delicati della teoria delle probabilità. Allora la probabilità sarà la somma delle possibilità di ogni caso favorevole[20].

Ricordiamo che già J. Bernoulli aveva richiamato l'attenzione proprio su questo punto, cioè sulla necessità di determinare il numero dei casi: «ad conjecturas de re qualibet rite formandas aliud nil requiratur quam ut tum numeri horum casuum accurate determinentur...»; ed aggiungeva: «pono autem omnes casus aeque possibiles esse seu pari facilitate evenire posse», consapevole, tuttavia, della delicatezza del compito e delle difficoltà ad esso inerenti[21].

[20] P.S. Laplace, *Essai philosophique*, cit., p. 7.

[21] J. Bernoulli, *Ars Conjectandi,* cit., pp. 219 e 223 sgg. «Per congetturare rettamente intorno ad una qualsiasi cosa non è richiesto nient'altro che siano determinati accuratamente i numeri di questi casi»; «assumo inoltre che tutti i casi siano equipossibili ovvero che possano verificarsi con la stessa facilità».

A questi due principi, Laplace ne faceva seguire un terzo ed un quarto che sono il nostro P_2, *a* e *b*.

Come quinto principio Laplace poneva il nostro C_2 .

Infine come sesto e settimo poneva rispettivamente la regola di Bayes ed il principio per la probabilità di un evento futuro, da cui egli derivava la nota «regola di successione».

Il sesto principio, che esprime la regola di Bayes, recita come segue:

> Una qualsiasi delle cause alle quali si può attribuire un evento è tanto più verosimile quanto più è probabile che, supposta l'esistenza di tale causa, l'evento abbia luogo. La probabilità dell'esistenza di una qualunque di queste cause è dunque una frazione, il cui numeratore è la probabilità dell'evento risultante dalla causa, ed il cui denominatore è la somma delle probabilità simili relative a tutte le cause. Se queste diverse cause, considerate a priori, hanno un diverso grado di probabilità, bisogna adoperare, in luogo delle probabilità dell'evento risultante da ciascuna causa, il prodotto di questa probabilità per quella della causa stessa[22].

La formula di Bayes è, secondo Laplace, il principio fondamentale di quella parte dell'analisi dei casi, che consiste nel risalire dagli eventi alle cause. Essa spiega il motivo per cui si attribuiscono eventi regolari e costanti ad una causa particolare; per esempio, perché, davanti all'uscita di «testa» cento volte consecutive, nasca il sospetto che la moneta sia truccata. Dal momento che tale evento, «deve essere l'effetto o di una causa regolare o del caso, la prima supposizione è più probabile della seconda»[23].

Alla concezione di Laplace sono state rivolte varie critiche, che riguardano sostanzialmente tre aspetti: la definizione della probabilità, il principio di indifferenza e la regola di successione. Si tratta di critiche che si sono tramandate nel tempo, per una sorta di forza di inerzia, ma che sono, nel complesso, abbastanza inconsistenti.

Alla definizione classica della probabilità è stata rivolta l'accusa di circolarità, in quanto rinvierebbe alla nozione di possibilità, che ad alcuni è sembrata sinonimo di quella; si è osservato, in sostanza, che, per sapere che cosa fosse la probabilità avremmo già dovuto conoscere il significato della 'equipossibilità', e quindi della probabilità stessa.

Questa critica di circolarità non sembra condivisibile. Infatti, i concetti di «possibilità» e di «probabilità» non possono in nessun caso essere considerati equivalenti, né i due termini sinonimi. Chiunque consideri, ad esempio, un dado, non avrà difficoltà, penso, a convenire che ciascuna

[22] P.S. Laplace, *Essai philosophique*, cit., p. 10.

[23] Ivi, p. 11.

delle sue facce rappresenta un possibile risultato di un lancio, e si spingerà, credo, fino a convenire che non sembra esservi motivo per dubitare che i sei risultati possibili non siano anche «equipossibili», a meno che il dado non sia truccato. L'identificazione delle sei facce come l'insieme esaustivo dei risultati possibili, ed anzi equipossibili, di un lancio del dado, non è tuttavia sufficiente ad offrire un criterio per calcolare valori numerici di probabilità di lanci del dado. Per poter esprimere una qualche valutazione probabilistica quantitativa, infatti, dobbiamo ancora sapere che uso fare di queste informazioni, e precisamente in quale relazione i casi equipossibili debbano essere messi tra loro: a ciò risponde appunto la definizione della probabilità, che, nel caso di Laplace, appare piuttosto come un criterio applicativo del calcolo, che non una vera e propria definizione. Del resto, come già abbiamo notato, la definizione matematica della probabilità, come quoziente fra casi favorevoli e totalità dei casi, è accolta da ogni scuola probabilistica ed anche la frequenza relativa non è altro che l'espressione di questo quoziente.

Probabilità e possibilità, insomma, non sembrano essere termini sinonimi, neppure nel più approssimativo linguaggio comune; dalla determinazione dei casi egualmente possibili non si ricava automaticamente alcun valore di probabilità; possiamo dire soltanto che sembrerebbe alquanto controintuitivo che valori differenti di probabilità fossero attribuiti a casi considerati equipossibili; ma non si va oltre. Come osserva de Finetti, «la probabilità è un qualche cosa che può venire distribuito sul campo delle possibilità», mentre queste ultime sono determinate sulla base delle informazioni disponibili[24].

Se la definizione di Laplace non intende affermare nulla circa il «significato» della nozione di probabilità, riguardo alla 'natura' di tale concetto, tuttavia, dovremo invece ricordare che, per Laplace, come per Bernoulli, e generalmente per tutti i probabilisti classici, esso sta ad esprimere un grado di certezza, relativa sia alla nostra ignoranza sia alla nostra conoscenza, e suscettibile di sfociare nella certezza piena, con l'aumentare delle nostre conoscenze. Il mondo, non dimentichiamolo, è qui concepito deterministicamente ordinato in ogni suo più piccolo evento, cosicché il concetto di probabilità non avrebbe senso alcuno per una mente onnisciente.

> Dobbiamo dunque considerare lo stato presente dell'universo come l'effetto del suo stato anteriore e come la causa del suo stato futuro. Una intelligenza che, per un dato istante, conoscesse tutte le forze da cui la natura è animata, e la situazione rispettiva degli esseri che la compongono, se inoltre fosse così vasta da sottomettere questi dati all'analisi,

[24] B. de Finetti, *Teoria delle probabilità*, Einaudi, Torino 1970, 2 voll., p. 36.

> abbraccerebbe nella stessa formula i movimenti dei più grandi corpi dell'universo e quelli del più leggero atomo: nulla sarebbe incerto per essa, e l'avvenire, come il passato, sarebbe presente ai suoi occhi[25].

Una seconda critica riguarda il cosiddetto «principio di indifferenza», il principio, cioè, che vuole che le ipotesi alternative siano tutte considerate come equiprobabili, in mancanza di motivi o di evidenza a favore di qualcuna di esse.

Intendo riferirmi, in particolare, a quelle critiche che non sono ispirate ad un rigido empirismo, perché questo nega in assoluto la legittimità di fare ricorso a criteri che permettano di congetturare in assenza di dati di osservazione. Perciò la critica empiristica colpisce pregiudizialmente il principio di indifferenza così come, in generale, ogni criterio di valutazione *a priori*.

Le critiche, cui alludo, sono invece quelle che tendono a mostrare come il principio di indifferenza sia fonte di conseguenze inconsistenti se non addirittura assurde. Due esempi chiariranno meglio il tipo di imputazioni rivolte al principio in questione.

1. Abbiamo un'urna della quale non si sa nulla circa il possibile colore delle palline in essa contenute. Allora, l'alternativa *bianco/non-bianco* formerà un'esaustiva coppia di ipotesi a ciascuna delle quali il principio di indifferenza attribuisce un valore di probabilità uguale ad ½. Ma anche le coppie rosso/non-rosso, nero/non-nero, ecc., possono essere considerate ragionevoli ipotesi alternative e pertanto anche a ciascuna di esse verrà attribuito un valore di probabilità uguale ad ½, in contraddizione con il principio delle probabilità totali.
2. Assumiamo di non conoscere assolutamente nulla circa l'ampiezza territoriale e numerica della popolazione di certe nazioni europee, diciamo l'Inghilterra, la Francia e l'Olanda. Concludiamo, per il principio di indifferenza, che un uomo preso a caso tra gli abitanti di questi tre paesi avrebbe uguale probabilità di essere inglese o francese o olandese.

L'argomentazione dei due esempi, che sono ispirati a quelli di J.M. Keynes[26], è tuttavia fallace. Essi rivelano di essere basati su un ragionamento alquanto confuso: viene qui fatto carico, al principio di indifferenza, della formulazione vaga e indeterminata del problema, come nel primo esempio, o della carenza di informazione, come nel secondo. Il primo

[25] Ivi, p. 2.
[26] Cfr. J.M. Keynes, *A treatise of Probability,* (1921), New York 1962, cap. IV.

esempio mostra di confondere l'applicazione delle regole del calcolo con la determinazione del suo campo di applicazione, la quale deve precedere l'attribuzione di valori di probabilità: possiamo essere interessati all'ipotesi di due, o più, specifici colori; in ogni caso la distribuzione di probabilità si riferirà sempre e soltanto all'insieme delle ipotesi predeterminate. Se volessimo ipotizzare tre colori, anziché due, ad ogni ipotesi spetterebbe, sempre per il principio di indifferenza, un valore di probabilità uguale ad 1/3; e tale valore diminuirebbe ove volessimo considerare un numero maggiore di colori. In tutto ciò non vi è nulla di paradossale: l'insieme dei predicati che vogliamo assumere nell'ipotesi non deriva dal principio di indifferenza; né, del resto, neppure dall'esperienza, perché, quand'anche avessimo estratto dall'urna un certo numero di palline bianche, tale evidenza non sarebbe ancora determinante, da sola, per la scelta dell'alternativa *bianco/non-bianco* a preferenza dell'intera famiglia dei colori!

Nel secondo esempio, al principio di indifferenza vengono addebitate le conseguenze di una totale carenza di informazioni; ma le attribuzioni di uno stesso valore di probabilità, alle tre possibilità considerate, non possono apparire assurde che allorquando siano giudicate alla luce delle informazioni rilevanti che di fatto possediamo! Se ci poniamo, invece, dal punto di vista di una effettiva totale assenza di informazioni, la paradossalità scompare; ad ogni modo, in assenza di informazioni, la questione non può porsi che in termini di scelta tra l'adozione di un criterio a priori, come il principio d'indifferenza, oppure il completo silenzio.

In breve, si può accettare o rifiutare l'uso del principio di indifferenza, come di altri basati sulla simmetria, o si può, più ragionevolmente, vagliare di volta in volta l'opportunità di ricorrervi o di soprassedere; non si deve tuttavia dimenticare che il problema, la cui soluzione si intende affidare al calcolo delle probabilità, deve essere chiaramente formulato e delimitato *prima* dell'applicazione delle regole del calcolo!

2.3 La regola di successione di Laplace

Il terzo bersaglio critico della teoria di Laplace è la regola di successione, ovvero la formula $(n + 1) / (m + 2)$ che permetterebbe di calcolare la probabilità di un determinato evento futuro, dopo che si siano osservati n precedenti casi tutti simili (in tal caso $m = n$), oppure un numero di n casi analoghi su m casi osservati. Ad esempio, qual è la probabilità di estrarre una nuova pallina bianca da un'urna dalla quale abbiamo già estratto n palline, tutte bianche?

La formula di successione discende dal settimo principio esposto da Laplace nell'*Essai,* che così recita:

> la probabilità di un avvenimento futuro è la somma dei prodotti della probabilità di ogni causa, tratta dall'avvenimento osservato, per

la probabilità che, esistendo tale causa, l'avvenimento futuro abbia luogo[27].

Sia *n* un numero di eventi osservati, tutti uguali (ad es., estrazioni di palline bianche da un'urna non nota), sia *e* l'evento futuro (l'estrazione di una ulteriore pallina bianca dalla stessa urna) e sia H_i l'insieme delle ipotesi alternative possibili (le varie possibili composizioni dell'urna). Allora:

$$\text{(E)} \quad p(e/n) = \sum_i p(H_i/n) \cdot p(e/H_i)$$

Come si vede, questa formula è identica al denominatore della formula di Bayes, salvo che al posto delle probabilità a priori delle ipotesi, troviamo le probabilità delle stesse, *dopo* le *n* estrazioni compiute. La formula ci dice qual è la probabilità di un evento, dopo averne osservati *n* simili, indipendentemente da ogni ipotesi, o, ciò che è lo stesso, subordinatamente ad ogni ipotesi possibile.

Il settimo principio deriva visibilmente dai principi del calcolo e non può essere messo in discussione senza mettere in discussione anche quelli.

Laplace ricava quindi la regola di successione dall'applicazione del teorema di Bayes e del principio (E), ora esposto, al problema seguente: un'urna racchiude *infiniti* biglietti bianchi e neri, in un rapporto ignoto; se ne estraggono *(p+q)*, di cui *p* bianchi e *q* neri; si domanda la probabilità che, estraendo un nuovo biglietto dalla stessa urna, esso sarà bianco.

La soluzione del problema richiede il ricorso al calcolo integrale.

Se indichiamo con *x* la proporzione ignota tra biglietti bianchi e il numero totale dei biglietti nell'urna, la probabilità di estrarre *p* biglietti bianchi e *q* neri è espressa dalla formula $x^p \cdot (1-x)^q$. È allora possibile calcolare la probabilità che *x* sia il vero rapporto tra i biglietti bianchi e la totalità dei biglietti nell'urna (con l'integrale calcolata tra 0 e 1):

$$p(x/(p+q) = \frac{x^p (1-x)^q dx}{\int x^p (1-x)^q dx}$$

La formula ci permette ora di calcolare la probabilità, che indichiamo con E, di estrarre un nuovo biglietto bianco dall'urna:

$$E = \frac{\int x^{p+1} (1-x)^q dx}{\int x^p (1-x)^q dx}$$

[27] P.S. Laplace, *Essai philosophique*, cit., p. 17.

Laplace mostra, quindi, come sia ricavabile da questa formula la semplicissima espressione

$$E = \frac{p+1}{p+q+2}$$

Laddove le estrazioni fatte siano molto numerose, e quindi p e q molto grandi, è possibile assumere, in virtù del teorema di Bernoulli, che la frequenza osservata corrisponda alla proporzione effettiva di biglietti bianchi e neri nell'urna, e quindi che essa esprima anche il valore della probabilità di una nuova estrazione di un biglietto bianco.

La regola di successione, intesa valere in un dominio infinito ed ignoto di eventi simili, ben si accorda con l'idea della regolarità degli effetti della natura, specie là dove essi siano considerati in gran numero.

L'apparente paradossalità della regola di successione si manifesta quando essa venga applicata ad un piccolo numero di osservazioni casuali. Ad esempio, dopo soltanto due estrazioni di palline bianche da un'urna, la probabilità di estrarne una successiva bianca sarebbe, in accordo alla formula di successione, uguale a 3/4, un valore davvero eccessivamente alto!

È facile costruire situazioni, nelle quali l'applicazione della formula di successione porti a risultati addirittura contraddittori con gli assiomi del calcolo: immaginiamo, ad esempio, di aver ottenuto 10 teste in 10 lanci di moneta. Ci chiediamo quale sia la probabilità di ottenere ancora una testa. Per la regola di successione, avremmo che $p(T_{11}) = 11/12= 0,91$; ma, per gli assiomi del calcolo, la probabilità di avere una undicesima testa è uguale a $(1/2)^{11}=0,00048$!

Invero, si può replicare a questo esempio che non sembra affatto che la formula di successione possa essere applicata a situazioni di questo tipo, dove il numero delle osservazioni è piccolo e la probabilità di ogni lancio è nota. Tuttavia, si può ribattere che niente vieta di supporre che il reale, seppure ignoto, rapporto tra biglietti bianchi e neri, in un'urna contenente infiniti biglietti, sia uguale ad 1/2, cosicché, dopo averne estratti 10 bianchi, la probabilità di un prossimo biglietto bianco si troverebbe a divergere, dall'ideale valore effettivo, esattamente come sopra, anche se noi non saremmo in grado di accorgercene. Così, in conclusione, non si vedrebbe di quale utilità possa essere una formula capace di sviare in tal misura la previsione di un nuovo evento!

Il fatto è che, applicata a casi bernoulliani, ovvero a modelli di estrazione con ripetizione (vale a dire con reimbussolamento di volta in volta), e con riferimento a popolazioni finite, la regola di successione si mostra suscettibile di critiche assai fondate, malgrado che non sembri attaccabile sotto il profilo matematico. D'altra parte, Laplace sembra riferirsi ad eventi tratti da un universo infinito e sorretti da un numero elevato di osservazioni, più o meno come accade per gli eventi naturali. Certo è che la formula non sembra portare a risultati credibili, se applicata a piccoli numeri, mentre, in relazione a grandissimi numeri, essa concorda sostanzialmente con il teorema di Bernoulli.

Potremmo fermarci a questa conclusione; ma possiamo anche chiederci se non sia possibile tentare un'interpretazione più profonda della regola di successione, tale da conciliare la sua apparente paradossalità con il senso generale della concezione di Laplace. Possiamo chiederci, ad esempio, se i risultati insoddisfacenti, come quelli sopra indicati, possano dipendere dal modello di applicazione, piuttosto che dalla regola stessa, e se non sia proprio il modello a dover esser posto al centro dell'indagine. In verità, un modello casuale, come quello bernoulliano, non sembra in armonia con il determinismo di Laplace!

Seguendo questo tipo di ragionamento, è facile arrivare a dimostrare che la regola di successione può essere facilmente e direttamente ricavata dal teorema di Bayes, quando questo sia applicato ad un modello non bernoulliano, indifferentemente dal fatto che si abbia a che fare con una popolazione finita o infinita. Consideriamo un'urna contenente $n+1$ palline, dalla quale siano state estratte, *senza rimetterle nell'urna*, esattamente n palline bianche.

Allora, chiedersi quale sia la probabilità che la prossima pallina sia bianca equivale a chiedersi quale sia la probabilità che l'urna contenga *tutte* palline bianche: si vuol conoscere, cioè, la probabilità di un evento che si colloca come *ultimo* in una serie della quale conosciamo tutti gli eventi precedenti, e per il quale non si prospettano a priori che due possibilità: la pallina sarà bianca, oppure no.

Sia quindi H_l l'ipotesi che tutte le palline siano bianche e sia H_2 l'ipotesi che tutte le palline siano bianche ad eccezione di quella rimasta nell'urna. Con il teorema di Bayes possiamo calcolare la probabilità di H_1 sulla base di n, la quale coincide con la probabilità che la prossima e ultima pallina sia bianca. Osserveremo che il risultato concorda con la formula di successione[28]:

$$p(H_1/n) = \frac{1}{1+\dfrac{n!}{(n+1)!}} = \frac{1}{\dfrac{(n+1)!+n!}{(n+1)!}} = \frac{(n+1)!}{n!+(n+1)!} = \frac{n!(n+1)}{n!(n+1+1)} = \frac{n+1}{n+2}$$

La formula si ricava, nello stesso modo, anche nel caso che le palline estratte non siano tutte bianche. Siano m le palline estratte senza ripetizione da un'urna che ne contiene esattamente $m+1$; di esse, n siano bianche ed $(m-n)$ nere. H_1 sia l'ipotesi che anche l'ultima pallina sia bianca; H_2 che

[28] *n!* indica un numero fattoriale ed equivale al numero stesso moltiplicato per tutti i numeri che lo precedono fino ad 1. Esempio: 5! = 5·4·3·2·1.

l'ultima pallina non sia bianca. Indichiamo con E le estrazioni effettuate. Allora:

$$p(H_1/E)=\frac{\dfrac{(n+1)!}{(m+1)!}}{\dfrac{(n+1)!}{(m+1)!}+\dfrac{n!(m-n+1)!}{(m+1)!}}=$$

$$=\frac{(n+1)!}{(m+1)!}\cdot\frac{(m+1)!}{(n+1)!+n!(m-n+1)!}=\frac{(n+1)n!}{n!(n+1+m-n+1)}=\frac{n+1}{m+2}$$

Il risultato ottenuto sembra particolarmente interessante, perché il modello dell'urna, dalla quale le palline vengono estratte senza reimbussolamento, si presta, assai meglio del modello casuale bernoulliano, quale metafora degli eventi naturali nel mondo deterministico di Laplace. Ogni evento, infatti, in una concezione deterministica, occupa un posto prestabilito nella catena causale degli eventi; esso non si dà casualmente e tanto meno entra in combinazioni possibili con gli eventi che lo hanno preceduto o con quelli che lo seguiranno. Pertanto il «prossimo» evento, non ancora osservato, di una catena di eventi simili, già osservati, viene sempre a porsi come *l'ultimo* di una serie di eventi, della quale non è necessario al momento prospettarsi l'ulteriore sviluppo: così, in un'ottica deterministica, la questione relativa alla probabilità di un prossimo evento naturale di un certo tipo, può essere posta, anziché con riferimento ad una serie infinita di eventi, come la questione della probabilità di un'urna finita nella quale sia rimasta una sola pallina da estrarre!

2.4 L'assiomatizzazione del calcolo delle probabilità

Assiomi

Nei primi decenni del nostro secolo furono proposte svariate assiomatizzazioni del calcolo delle probabilità, fra le quali quella proposta da A.N. Kolmogorov ottenne il più esteso riconoscimento perché sviluppava rigorosamente, a partire dagli assiomi, l'intero calcolo[29].

Oggi, l'approccio assiomatico è forse il più diffuso nei testi matematici di calcolo delle probabilità. In esso, il concetto di probabilità è introdotto come nozione primitiva, implicitamente definita dagli assiomi che ne governano l'uso. I principi poco sopra elencati possono essere sintetizzati

[29] A.N. Kolmogorov, *Grundbegriffe der Wahrscheinlichkeitsrechnung*, Berlin 1933. Un sistema assiomatico analogo era stato proposto un anno prima da H. Reichenbach, *Axiomatik der Wahrscheinlichkeitsrechnung*, «Math. Zeits.», 34, 1932. Anche L. Bohlmann (1901) e S. Bernstein (1917) avevano proposto sistemi di assiomi molto simili a quello del matematico russo.

nei seguenti quattro assiomi che costituiscono il nucleo essenziale di ogni sistema di calcolo delle probabilità:

1. Gli eventi sono sottoinsiemi di uno spazio **Ω** e formano una sottoclasse additiva **Δ.**
2. Ad ogni E∈**Δ** è assegnato un numero reale non negativo p(*E*), detto probabilità di *E*.
3. $P(\mathbf{\Omega}) = 1$
4. Se $A \cap B = \emptyset$, allora $p(A \cup B) = p(A) + p(B)$.

L'assioma 4 esprime la legge delle probabilità totali nella sua forma più semplice, nella quale compaiono soltanto due eventi incompatibili. Passando a condizioni più generali, possono aversi i due seguenti asserti:

a) per un insieme finito di eventi a due a due incompatibili:

$$p(A_1 \cup \ldots\ldots \cup A_n) = \sum_{i=1}^{n} p(A_i)$$

b) per un insieme numerabile di eventi a due a due incompatibili:

$$p(A_1 \cup \ldots\ldots \cup A_\infty) = \sum_{1}^{\infty} p(A_i)$$

La formula *a*) esprime l'addittività semplice, o finita; la formula *b*) l'addittività completa o numerabile, o σ-addittività. Non tutti concordano nell'ammissione dell'addittività numerabile. Chi l'accetta, aggiunge al gruppo di assiomi sopra indicati un quinto assioma dal quale è derivabile l'addittività completa. A favore dell'ammissione di quest'ultima può essere addotta una trattazione più comoda e unitaria che deriva dalla riduzione della teoria delle probabilità alla teoria della misura. Ad ogni modo, la decisione circa l'ammissione della σ-addittività è largamente questione di opinioni.

L'assioma 3 richiede qualche precisazione. Esso dice che l'evento certo **Ω** ha probabilità 1, ma non esclude che vi siano anche altri eventi con tale valore di probabilità (eventi quasi certi), e, di conseguenza, non è escluso che vi siano eventi diversi da **Ø** con probabilità nulla (eventi quasi impossibili). La necessità di eventi quasi certi e quasi impossibili può evincersi considerando una successione di lanci di moneta. Su *n* lanci abbiamo 2^n risultati possibili, che costituiscono lo spazio **Ω** e ai quali deve essere attribuita una uguale probabilità. Se ora consideriamo una successione infinita di lanci, che dà luogo ad un insieme più che numerabile di successioni di risultati possibili, ognuna di queste successioni, essendo infinita, dovrà avere probabilità nulla, altrimenti si avrebbero eventi composti (ad esempio una disgiunzione di un certo numero di successioni) con probabilità maggiore dell'unità. Pertanto, se sostituissimo l'assioma 3 con un altro che affermasse che *solo* $\mathbf{\Omega}$ ha probabilità 1, la teoria che ne conseguirebbe non potrebbe essere applicata al modello di successioni infinite di lanci di moneta.

Teoremi notevoli

Tra i risultati di particolare rilievo della teoria delle probabilità ci soffermiamo brevemente sulla distribuzione binomiale e sul teorema di Bernoulli, rimandando ovviamente per una trattazione completa e matematicamente rigorosa del calcolo della probabilità ad uno dei tanti testi esistenti.

Distribuzione binomiale. Dati n eventi E_i $(i= 1, 2, ..., n)$ indipendenti, tutti con probabilità p, la probabilità di avere esattamente r successi tra di essi è data dalla formula

$$\mathrm{p}(\mathrm{E}_\mathrm{r}) = \frac{n!}{r!(n-r)!} \cdot p^r (1-p)^{n-r}$$

Tale distribuzione, detta binomiale perché deriva dallo sviluppo della potenza del binomio, è chiamata anche *modello bernoulliano o delle prove ripetute*. Lo studio di tale modello risale alle origini del calcolo delle probabilità e si connette con quello dei lanci di moneta o di dado. In effetti, il modello costituito da lanci ripetuti di una moneta, oltre ad essere il più antico, è forse anche quello maggiormente dotato di immediatezza. L'andamento di una successione di lanci di moneta, chiaramente espresso dal triangolo di Tartaglia (o di Pascal), mostra come, essendo la moneta simmetrica, avrà sempre maggiore probabilità quella successione di risultati la cui distribuzione più si avvicina alla probabilità $p = 1/2$.

L'estensione di tale modello è costituita da ciò che è chiamata «estrazione bernoulliana» da un'urna nota (estrazione di una pallina alla volta, con reimbussolamento): condizione indispensabile di tale distribuzione è l'indipendenza delle estrazioni, che devono avvenire tutte in condizioni immutate.

La distribuzione binomiale costituisce una delle basi dell'*inferenza diretta,* cioè l'inferenza da una popolazione nota ad un suo campione casuale. Segue da essa che la probabilità di una ipotesi statistica relativa ad un campione ha il suo massimo quando la frequenza della caratteristica che ci interessa, nel campione, è uguale a quella effettiva nella popolazione.

Esempio: si facciano 9 estrazioni bernoulliane di palline da un'urna contenente una proporzione di 2/3 di palline bianche. Si vuol sapere qual è la probabilità di avere nel campione 5 palline bianche e 4 non-bianche.

Soluzione: $p(e) = \frac{9!}{5!(9-5)!} \cdot (2/3)^5 \cdot (1/3)^4 = 0{,}2$

Triangolo di Tartaglia

$(a+b)$	1 1	$2^1=2$
$(a+b)^2$	1 2 1	$2^2=4$
$(a+b)^3$	1 3 3 1	$2^3=8$
$(a+b)^4$	1 4 6 4 1	$2^4=16$
$(a+b)^5$	1 5 10 10 5 1	$2^5=32$
$(a+b)^6$	1 6 15 20 15 6 1	$2^6=64$
$(a+b)^7$	1 7 21 35 35 21 7 1	$2^7=128$
$(a+b)^8$	1 8 28 56 70 56 28 8 1	$2^8=256$
$(a+b)^9$	1 9 36 84 126 126 84 36 9 1	$2^9=512$

La figura espone i primi nove righi del triangolo di Tartaglia, che corrisponde alla distribuzione binomiale quando $p=1/2$. I numeri del triangolo corrispondono ai coefficienti dei fattori letterali nello sviluppo del binomio corrispondente (a sinistra). La somma dei coefficienti è uguale a 2^n (a destra).

Teorema di Bernoulli o legge dei grandi numeri. Questo teorema, di cui si è già parlato, dimostra che la frequenza relativa di successi, in un campione di ampiezza n molto grande, tende al valore p della probabilità di ogni singolo evento, e che, per $n \to \infty$, tende ad 1 la probabilità che la differenza tra la frequenza relativa di successi rilevata nel campione e la frequenza relativa nella popolazione sia minore di un valore piccolo a piacere.

Anche questo teorema rappresenta una base fondamentale dell'inferenza diretta perché dà indicazioni sulle probabilità connesse con un campionamento da popolazione nota. Come abbiamo precedentemente osservato, Bernoulli arrivò a questo teorema nel tentativo di risolvere il problema dell'inferenza inversa.

Discende da questo teorema ciò che è comunemente noto come *legge empirica del caso,* che sostiene che la frequenza nel campione tende alla frequenza nella popolazione, per n molto grande. La differenza tra il teorema e questa 'legge' consiste nel fatto che, mentre il primo asserisce qualcosa circa una 'probabilità', la seconda fa un'asserzione sull'andamento dei fatti, suggerendo un'aspettativa certa là dove questa è solo probabile. Tuttavia, la legge empirica del caso assolve ad un'utile funzione come ponte tra livello teorico-matematico e livello pratico, a patto che non si dimentichi che il contesto in cui ci si trova è pur sempre quello dell'incerto.

2.5 Concezioni della probabilità

Kolmogorov credeva che l'assiomatizzazione del calcolo ponesse fine alle accese polemiche tra i sostenitori delle diverse concezioni della probabilità. In realtà così non è stato, né poteva esserlo. I sistemi assiomatici hanno il merito di sviluppare nel modo più rigoroso tutti i teoremi del calcolo delle probabilità, a partire da un piccolissimo numero di assiomi, che condensano, come si è detto, i principi classici del calcolo e che formano il nucleo essenziale ed indiscutibile di ogni approccio al calcolo delle probabilità. Tuttavia, un sistema assiomatico permette soltanto di sviluppare le relazioni che legano le probabilità tra loro, di far scaturire probabilità da probabilità, ma non contiene criteri applicativi per valutare le probabilità degli eventi semplici, né, in generale, per la ripartizione dello spazio $\boldsymbol{\Omega}$ in eventi incompatibili ed equiprobabili.

Consideriamo, ad esempio, l'assioma quattro: esso dice che se due eventi sono tra loro incompatibili, la probabilità che si verifichi l'uno o l'altro dei due è data dalla somma delle probabilità di ciascuno dei due eventi. La probabilità di ottenere «un asso oppure un tre», in un lancio di dado, è perciò uguale alla somma della probabilità dell'evento-asso e della probabilità dell'evento-tre. Fin qui è tutto chiaro; ma il sistema di assiomi non offre suggerimenti in relazione al modo in cui vadano calcolate queste due probabilità separatamente prese, quella dell'evento-asso e quella dell'evento-tre. Ecco perché nessun sistema assiomatico può diventare applicativo se non viene integrato da qualche criterio per la valutazione di tali probabilità di base. In altre parole, il sistema di assiomi non è *categorico,* cioè non è tale da permettere una determinazione univoca dei valori di probabilità. La definizione della nozione di «probabilità» sopperisce, quindi, nel fornirci un criterio per l'applicazione del calcolo agli eventi elementari.

Sembra ovvio, e, per così dire, quasi istintivo, affermare che la probabilità di ottenere un «tre» in un lancio di dado sia uguale ad 1/6; eppure, con questa asserzione, noi stiamo di fatto applicando la definizione di Laplace; e non tutti potrebbero concordare con questa scelta, come già sappiamo. Ma c'è ancora qualcos'altro da osservare.

Chiediamoci, ad esempio, quale sia la probabilità che, in due lanci di una moneta, si abbia almeno una «testa». Indichiamo «testa» con T e «croce» con C. Sono possibili i tre seguenti risultati: T=0; T=1; T=2. Di questi casi, due sono favorevoli; perciò la probabilità richiesta sarebbe uguale a 2/3. Potremmo però indicare i risultati dei due lanci di moneta in maniera diversa, cioè nel modo seguente: TT; TC; CT; CC. Qui i risultati favorevoli sono tre su quattro, dunque 3/4 sarebbe il valore della probabilità desiderata.

Quale risposta dobbiamo ritenere come giusta, 2/3 oppure 3/4? A priori nessuna argomentazione teorica sembra in grado di giustificare una scelta piuttosto che l'altra.

Una questione analoga è posta dalle tre differenti statistiche di Maxwell-Boltzmann, Bose-Einstein e Fermi-Dirac, con riferimento ad n particelle e k celle.

La statistica di Maxwell-Boltzmann attribuisce equipossibilità a tutte le possibili distribuzioni, distinguibili e no, delle *n* particelle nelle *k* celle, il cui numero è perciò k^n (ciò equivale a considerare le particelle come oggetti singolarmente individuabili); questa statistica, che si riduce alla distribuzione binomiale quando $k=2$, caratterizza un mondo entropico. La statistica di Bose-Einstein si basa invece sull'equipossibilità di tutte le possibili distribuzioni distinguibili delle *n* particelle nelle *k* celle, il cui numero è allora $(n+k-1)!/n!(k-1)!$ (poiché le particelle sono considerate individualmente indistinguibili, vengono prese in considerazione soltanto le distribuzioni statistiche delle particelle nelle celle). La statistica di Fermi-Dirac, infine, distribuisce uniformemente i pesi iniziali su tutte le distribuzioni distinguibili delle *n* particelle nelle *k* celle, tali che in ogni cella ci sia *al massimo* una particella, il cui numero (per $k \geq n$) è pertanto $k!/n!(k-n)!$

È evidente che la decisione circa ciò che debba essere assunto come equipossibile è in larga misura soggettiva e può essere fondata su criteri tra loro assai differenti, anche in relazione alle caratteristiche dell'oggetto di studio e di analisi. Si tratta, in sostanza, di scegliere il modello che sembri più adeguato a rappresentare il campo di oggetti sotto indagine. Tuttavia, il calcolo delle probabilità può essere applicato soltanto dopo una tale scelta preliminare ed in accordo con le regole implicite nella concezione della probabilità adottata.

2.5.1 La concezione logicista

La concezione logicista della probabilità si presenta, per molti aspetti, come una prosecuzione di quella classica, della quale abbiamo già parlato.

Suggerita da Keynes[30], è stata sviluppata in modo sistematico e su basi rigorosamente logiche da R. Carnap, il quale, nei *Foundations*[31], definisce la probabilità come una relazione logica tra enunciati di un linguaggio adeguatamente predeterminato. Si tratta, più esattamente, di una relazione di implicazione parziale tra *campi* (*ranges*) di enunciati, cui la logica induttiva ha il compito di fornire una misura quantitativa.

Se la logica deduttiva regola i casi di implicazione totale (L-verità) o di totale esclusione tra due enunciati (L-falsità, o contraddizione), la logica induttiva è intesa quale sviluppo del sistema di regole idonee a misurare quantitativamente l'estensione delle intersezioni tra campi di enunciati. Si tratta, per Carnap, del compito di trovare un adeguato e rigoroso *explicatum* per il concetto di probabilità, un compito, anche in questo caso, di *ricostruzione razionale*.

[30] J.M. Keynes, *A Treatise on Probability,* cit.
[31] R. Carnap, The *Logical Foundations of Probability,* cit.

Per illustrare i concetti fondamentali di questa concezione, prendiamo un linguaggio semplicissimo L_3, con tre individui (*a, b, c*) ed un solo predicato *P*. Con riferimento a tale linguaggio, possiamo allora determinare 2^3 *descrizioni di stato,* ovvero tutti i «mondi possibili» descrivibili in questo linguaggio, che formano un insieme esaustivo di enunciati fra loro incompatibili.

$Pa \wedge Pb \wedge Pc$

$Pa \wedge Pb \wedge \neg Pc$

$Pa \wedge \neg Pb \wedge Pc$

$\neg Pa \wedge Pb \wedge Pc$

$Pa \wedge \neg Pb \wedge \neg Pc$

$\neg Pa \wedge Pb \wedge \neg Pc$

$\neg Pa \wedge \neg Pb \wedge Pc$

$\neg Pa \wedge \neg Pb \wedge \neg Pc$

Osserviamo che queste otto descrizioni di stato possono essere distribuite in quattro *descrizioni di struttura,* cioè in *partizioni* contraddistinte soltanto dal *numero* degli individui (e non da *quali*) aventi la proprietà *P*, ovvero: $3P$; $2P \wedge 1\neg P$; $1P \wedge 2\neg P$; $3\neg P$. Queste costituiscono tutte le descrizioni statistiche del nostro linguaggio.

Stabilite certe condizioni generali di adeguatezza, di valore più euristico che formale, Carnap passa a definire una prima classe di funzioni di misura e di conferma, che egli chiama *funzioni regolari*, con le quali è possibile esprimere i principi fondamentali del calcolo classico delle probabilità, ivi compresi il teorema di Bayes ed il teorema di Bernoulli.

Una funzione di misura è detta *regolare* se soddisfa le seguenti condizioni:

a) assegna numeri reali positivi alle descrizioni di stato di ogni linguaggio finito L_N;
b) la somma dei valori di misura di tutte le descrizioni di stato di L_N è uguale ad 1.

La definizione è quindi estesa agli *enunciati* di L_N nel modo seguente:

- per ogni enunciato L-falso j, $m(j) = 0$;
- per ogni enunciato non L-falso j, $m(j) =_{df}$ la somma dei valori di misura delle descrizioni di stato in cui j vale, cioè $m(j) = m(R_j)$, (dove R_j indica il *campo* di j).

Sulla base delle *m*-funzioni regolari è quindi definita la funzione regolare di conferma *(c-funzione regolare)* per ogni linguaggio finito L_N:

$$c(h,e) = \frac{m(e \wedge h)}{m(e)}, \text{ per } m(e) \neq 0$$

Se $m(e) = 0$, allora $c(h,e)$ non ha nessun valore.

Il successivo passo, verso la definizione di un *explicatum* adeguato per la probabilità, porta Carnap ad operare una limitazione dell'insieme delle funzioni possibili ed alla definizione delle *funzioni simmetriche,* con le quali è possibile trattare e risolvere le questioni di inferenza diretta. Per questa inferenza, il concetto fondamentale è quello di *descrizione di stato*: viene qui attribuita una uguale *misura* a tutte le descrizioni di stato tra loro *isomorfe,* cioè quelle comprese in una medesima descrizione di struttura[32].

Le funzioni simmetriche sono sufficienti per trattare tutte le questioni di inferenza diretta poiché, in questo tipo di inferenza, la popolazione è sempre nota, e, di conseguenza, è nota anche la descrizione di struttura entro cui dovranno essere trattate le inferenze in questione.

Per affrontare i problemi d'*inferenza inversa*, Carnap ritiene opportuno operare una ulteriore restrizione del campo delle funzioni, individuando quella c^*-*funzione* di conferma basata sull'attribuzione di una distribuzione uniforme alle *descrizioni di struttura,* privilegiando così una scelta analoga a quella della statistica di Bose-Einstein.

Al suo concetto di probabilità logica, o meglio di *conferma*, Carnap attribuisce tre significati; essa può rappresentare: 1) la misura del supporto fornito da una evidenza ad una ipotesi; 2) la misura del quoziente di equa scommessa; 3) la stima della frequenza relativa.

Nelle opere più tarde, Carnap, avvicinandosi alla posizione soggettivistica, tende a considerare il suo concetto di conferma anche come funzione di credibilità razionale, ossia come quel grado di credenza, oggettivo perché fondato sulla relazione logica di implicazione parziale, che tutti coloro che condividono il suo metodo induttivo si troverebbero a condividere, in relazione a dati enunciati *h* ed *e* di un linguaggio determinato. La probabilità logica sarebbe, per così dire, il limite ideale di quella credenza personale, espressa dalla posizione soggettivistica[33].

La logica induttiva di Carnap non offre, in verità, un metodo effettivo ed autonomo per il calcolo delle probabilità in situazioni concrete, sebbene contenga una indicazione originale e preziosa per la valuta-

[32] La distribuzione uniforme su tutte le descrizioni di stato, non solo quelle isomorfe, corrisponde alla distribuzione operata nella statistica di Maxwell-Boltzmann.

[33] Cfr. R. Carnap, *Inductive Logic and Rational Decisions,* 1969; trad. it. in R. Carnap, *Analiticità, significanza, induzione*, Il Mulino, Bologna 1971. Per una esposizione più dettagliata e per la discussione di alcuni aspetti della concezione carnapiana, si veda M.G. Sandrini, *Probabilità e induzione. Carnap e la conferma come concetto semantico,* Franco Angeli, Milano 1991.

zione delle probabilità nell'inferenza inversa. Piuttosto, la concezione carnapiana, costituendo, come ho già detto, una esplicazione ed una ricostruzione razionale del concetto, e del calcolo, della probabilità, offre un inestimabile strumento di chiarificazione concettuale di indiscutibile valore teoretico, dovuto anche al fatto che essa si estende a coprire tutte le definizioni proposte per tale concetto. In tal modo essa contribuisce anche ad individuare, separare e chiarire numerose questioni spesso tra loro confuse.

2.5.2 La concezione frequentistica

La definizione frequentistica della probabilità è connessa con lo sviluppo della statistica, cioè con lo studio dei fenomeni di massa, ma si andò affermando in particolar modo, sia pur in modo alquanto vago, nel clima positivistico della seconda metà del secolo XIX, quando i criteri aprioristici della concezione classica vennero sottoposti a critica radicale e non sempre giustificata.

L'assunto fondamentale dei sostenitori di tale concezione è anzitutto un assunto empiristico: nessun asserto conoscitivo può avere senso se non è basato sull'esperienza e controllato dai fatti; perciò la probabilità non può esprimere né un grado di credenza, soggettivo e psicologico, né può fare appello a criteri *a priori* di simmetria o simili, ma deve consistere in una asserzione fattuale circa eventi osservati, deve esprimere cioè una frequenza relativa osservata.

Ne consegue che soltanto l'osservazione empirica delle frequenze effettive delle 'teste' e delle 'croci' permetterà di compiere valutazioni di probabilità circa serie di lanci di una moneta. A rigore, dovrebbe trattarsi sempre, si badi bene, di una moneta ben definita e concreta, perché le probabilità sui lanci futuri riflettono le frequenze relative già osservate in lanci precedenti della stessa moneta. Se presa in senso stretto, tale concezione verrebbe così a limitare in modo non marginale le applicazioni del calcolo delle probabilità; non si può non osservare, infatti, che, se nessuna moneta 'ideale' può sostituirsi alle frequenze effettivamente rilevate nei lanci (ciò equivarrebbe infatti a far uso di un criterio di simmetria), neppure nessuna moneta concreta può proporsi per *tutte* le monete concrete, perché, di fronte ad ogni nuova moneta 'non esperita', non sembra possibile esimerci dal dubbio che essa sia, ad esempio, 'truccata', con la conseguenza di trovarci sempre nell'impossibilità di fare uso delle frequenze relative rilevate in lanci di monete forse solo apparentemente simili. L'appello rigoroso all'esperienza sembra perciò ridurre e limitare l'estensione dell'esperienza stessa e di ciò che da essa potremmo imparare!

D'altra parte, nessun numero finito di lanci sembra sufficiente a determinare il valore della probabilità: ogni successione finita di essi tenderà presumibilmente a proporre un valore differente. Per ovviare a tale difficoltà, i frequentisti definiscono la probabilità come *il limite* della frequenza relativa di successi, al tendere del numero delle prove all'infinito.

La concezione frequentistica della probabilità non è neppure immune da difficoltà e critiche interne, connesse talvolta con il concetto stesso di *limite*, non sempre ben definito. Al fine di rendere consistente la sua teoria, R. von Mises[34], uno dei fondatori di tale concezione, la restringe alla teoria di classi di fenomeni di massa e di eventi ricorrenti, di modo che la probabilità sia asseribile solo in riferimento ad un *collettivo,* che soddisfi le due condizioni seguenti:

a) le frequenze relative dell'evento considerato devono tendere ad un limite;
b) ogni successione ottenuta come sottoinsieme del collettivo totale deve essere 'irregolare' nei risultati delle prove (deve essere cioè garantita la casualità della scelta della successione), di modo che il limite delle frequenze rimanga sempre immutato.

F. Waismann ha duramente criticato la teoria di von Mises, mettendo in evidenza, fra l'altro, l'incompatibilità del concetto di «serie convergente ad un limite» con quello di irregolarità[35]. In effetti, i concetti di limite e di collettivo sono in Mises troppo vaghi per non dare adito a contraddizioni. Numerosi tentativi sono stati fatti, da parte dei sostenitori della concezione frequentistica, per ovviare agli inconvenienti della teoria di von Mises[36].

H. Reichenbach, pur accettando la definizione della probabilità di Mises, identifica il concetto di limite della frequenza con il concetto usuale dell'analisi matematica, mentre rinuncia al concetto di collettivo, ed ammette come classe di riferimento ogni tipo di successione.

L'identificazione della probabilità con la frequenza relativa a lungo andare non permette, a rigore, neppure di parlare di probabilità di eventi singoli, cosicché non solo enunciati come «questo malato ha la probabilità *r* di morire» non avrebbero senso, ma neppure le singole scommesse. Asserire che la probabilità di 'testa' in lanci di moneta è uguale ad 1/2 significa asserire che, in lanci ripetuti della moneta, verrà testa nel cinquanta per cento dei casi; un lancio singolo, viceversa, non può essere misurato in gradi.

[34] Cfr. R. von Mises, *Wahscheinlichlkeit, Statistik und Wahrheit,* Wien, 1928, 2° ed. rielaborata 1936.

[35] Cfr. F. Waismann, *Logische Analyse des Wahrscheinlichkeitsbegriffs,* «Erkenntnis», vol. 1, 1930-31.

[36] Sul concetto di collettivo cfr. anche B. de Finetti, *Statistica e probabilità nella concezione di R. von Miises*, in B. de Finetti, *Logica dell'incerto,* Il Saggiatore, Milano 1989, pp. 196 sgg.

Al riguardo, Reichenbach, consapevole della ragionevolezza pratica delle asserzioni probabilistiche circa eventi unici, tanto nella vita quanto nella scienza, ne tenta una interpretazione in termini di frequenza: le asserzioni di probabilità circa previsioni singole sarebbero da interpretarsi come formulazioni ellittiche, che, mentre sembrano affermare la probabilità di un singolo accadimento, si riferirebbero in realtà alla frequenza relativa di accadimenti simili in una lunga serie: è evidente, egli afferma, che il grado di probabilità non ha nulla a che fare con le singole asserzioni; tuttavia, esso serve come indicazione per la scelta di un'*assunzione*. Che significato possiamo dunque dare, seguendo Reichenbach, all'affermazione che il signor A, affetto da grave malattia, ha il 75% di probabilità di morirne? Il signor A, ovviamente, vivrà o morirà, né l'evento è suscettibile di ripetizione. La frequenza relativa, osservata in casi analoghi, nulla può dirci a proposito della vita o della morte del signor A. Eppure, se dovessimo scommettere sulla sua vita o sulla sua morte, sarebbe più ragionevole, secondo Reichenbach, scommettere sulla sua morte, perché, in una serie di ripetute scommesse analoghe, avremmo maggiori probabilità di successo! Idealmente, dunque, il caso singolo, irripetibile, è interpretato come appartenente ad una serie di eventi analoghi, dei quali possiamo conoscere la frequenza relativa.

In ciò consiste, per Reichenbach, l'acquisizione di un buon abito induttivo, ed in ciò si giustifica, a suo parere, «il trasferimento di significato dal generale al particolare» che si compie quando parliamo di probabilità di eventi singoli: le assunzioni compiute con tale criterio ci porteranno a conseguire nella vita, mediamente, più vincite che perdite. Dal momento che la conoscenza non può pretendere alla certezza, un buon abito induttivo, e cioè un accorto uso della probabilità frequentistica, appare a Reichenbach il miglior comportamento che l'uomo possa assumere nella vita: se nulla può essere asserito con certezza, l'esperienza fornisce tuttavia delle indicazioni, esprimibili in frequenze relative di eventi osservati, che possono essere assunte provvisoriamente come valide per le previsioni sul futuro. Introducendo il concetto di «assunzione» (*posit*), tuttavia, Reichenbach si è già allontanato da quell'empirismo che generalmente è invocato a sostegno della concezione della probabilità come frequenza relativa e troverebbe probabilmente molti oppositori proprio all'interno dei sostenitori della sua stessa concezione.

Altri, come Fisher e numerosi statistici, preferiscono introdurre il concetto frequentistico della probabilità come termine primitivo in un sistema di assiomi.

2.5.3 La concezione soggettivistica

La concezione soggettivistica, che fa capo principalmente alla scuola di Bruno de Finetti, definisce la probabilità come grado di credenza soggettiva: «grado di fiducia di un dato soggetto, in un dato istante e con un dato insieme di informazioni, riguardo al verificarsi di un dato evento»[37]. Nella sua concezione, l'elemento soggettivo è considerato come essenziale e tale da non poter essere eliminato, «senza al tempo stesso eliminare la stessa ragion d'essere del concetto di probabilità»[38]

Le difficoltà di tradurre in valutazioni quantitative queste opinioni soggettive possono essere operativamente risolte ponendosi idealmente in una situazione di *equa scommessa* e chiedendosi quale prezzo massimo *p* si sarebbe disposti a pagare per ricevere un guadagno *s* nel caso si verificasse l'evento *e*.

Per illustrare efficacemente questa situazione possiamo ricorrere al seguente esempio, cui è stato dato il nome di «gioco di de Finetti». Si supponga che un amico abbia appena sostenuto una prova di concorso e ritenga di averla svolta in modo eccellente. Allora possiamo proporgli il seguente gioco: ti sono offerte due possibilità: o estrarre una palla da un'urna contenente 98 palle rosse e due nere, e, se ne estrai una rossa, ti diamo 10000 euro, oppure aspettare il risultato dell'esame, dopo di che ti daremo 10000 euro se riceverai il massimo dei voti. Cosa fai: estrai, oppure aspetti? Se l'amico è davvero convinto di meritare il massimo dei voti, presumibilmente deciderà di aspettare. Il tal caso il suo grado di fiducia soggettiva è superiore al 98 per cento. Se invece sceglie di estrarre la palla, gli si proponga di nuovo: adesso nell'urna ci sono 80 palle rosse e venti nere. Restando ferme le condizioni di cui sopra, estrai o aspetti? Se l'amico decide di aspettare, sappiamo che la sua fiducia soggettiva nella massima votazione dell'esame è superiore all'ottanta per cento ma inferiore al 98. Scegliamo allora un valore intermedio, ad esempio 90, e ripetiamo la domanda. Proseguiamo così finché l'amico non dirà «per me ora è indifferente estrarre o aspettare». Il valore a cui si fermerà darà la misura della sua probabilità soggettiva nel raggiungimento del massimo risultato della sua prova[39].

La condizione di *equità* della scommessa richiede che si sia sempre disposti a scambiare le parti con l'avversario. In tal caso, diviene naturale assumere *p/s* come misura del proprio grado di fiducia circa il verificarsi di *e*[40].

La richiesta di equità, che serve ad assicurare che non sia garantito a priori un guadagno, o una perdita, introduce nel calcolo la condizione

[37] B. de Finetti, *Teoria delle probabilità,* cit., p. 6.

[38] B. de Finetti, *La previsione: le sue leggi logiche, le sue fonti soggettive,* in B. de Finetti, *La logica dell'incerto,* cit., p. 135.

[39] Cfr. A.D. Aczel, *Chance. Dai giochi d'azzardo agli affari (di cuore)*, trad. it. R. Cortina, Milano 2005, pp. 32-34.

[40] Contro le interpretazioni comportamentistiche della credenza, i soggettivisti osservano, giustamente a mio avviso, che *ritenere* equa una scommessa non implica essere anche effettivamente disposti a scommettervi.

di *coerenza*, «che costituisce il solo principio da cui è possibile dedurre l'intero calcolo delle probabilità». Tale calcolo si presenta come «l'insieme di quelle regole cui devono essere assoggettate le valutazioni soggettive della probabilità di vari eventi da parte di un dato individuo se si desidera che non vi sia al loro interno una contraddizione fondamentale»[41].

La regola di coerenza può essere considerata equivalente ai tre seguenti principi:

1. $p(A) \geq 0$
2. la probabilità di un evento certo è = 1
3. $p(A \vee B) = p(A) + p(B)$

Ora, siccome tutto il calcolo delle probabilità può anch'esso essere derivato da questi tre soli principi, è possibile concludere che, se i gradi di fiducia assegnati da un soggetto sono coerenti, allora essi debbono combinarsi secondo le regole del calcolo delle probabilità, che mostra, con ciò, di essere l'unica logica possibile dell'incerto. Dal principio di coerenza discendono anche i teoremi delle probabilità totali, con cui viene introdotta l'addittività finita.

Vi è quindi un senso in cui la nozione di coerenza deve essere considerata *oggettiva*: «le condizioni di coerenza dovranno escludere la possibilità di certe conseguenze la cui inaccettabilità appaia esprimibile e riconoscibile per chiunque, indipendentemente da ogni sua opinione e/o da ogni suo giudizio di maggiore o minore 'ragionevolezza' circa opinioni altrui»[42]

Il significato soggettivo della probabilità è, per i soggettivisti, «un fatto oggettivo e indiscutibile»[43], che è buona norma accettare, anziché cercare di nascondere, come fanno i sostenitori di altre concezioni, ed in particolare i frequentisti. A questo proposito, de Finetti osserva:

> [...] non si può trovare la chiave del giardino incantato dove, tra l'erba voglio e gli arbusti di bacchette magiche, tra gli alberi carichi di monadi e di noumeni, sbocciano i fiori di «Probabilitas realis». Quei fiori favolosi che, infilati all'occhiello, fungerebbero da surrogati di opinioni, dispenserebbero dal fabbricarsele con la testa, renderebbero superfluo tale ingombro sopra il collo[44].

Agli occhi di de Finetti può applicarsi ad ogni tentativo di definizione 'oggettivistica' della probabilità una efficace immagine presa a

[41] B. de Finetti, *La previsione,* in B. de Finetti, *La logica dell'incerto*, cit., p. 78.

[42] B. de Finetti, *Teoria delle Probabilità*, cit., p. 103.

[43] Ivi, p. 256.

[44] Ivi, p. 393.

prestito da L.J.Savage: «si pretenderebbe di confezionare un'omelette probabilistica rifiutandosi di rompere uova probabilistiche. E i casi sono due: o il risultato non è un'omelette, oppure uno ci ha messo le uova di nascosto o senza avvedersene»[45]. L'uovo probabilistico, vale a dire l'elemento soggettivo, appare del resto evidente a de Finetti, non appena dal piano astratto si passi a quello concreto dell'applicazione del calcolo delle probabilità. Allora, tanto il criterio basato sulla simmetria, quanto quello basato sulla frequenza, rinviano alla soggettività della scelta dei requisiti atti a formare classi di 'eventi simili'. Se ci chiediamo quale sia la probabilità che esca un 'cinque' in un lancio di dado, diremo che essa è pari ad 1/6, se attribuiamo ad ogni faccia del dado la stessa possibilità di uscire. Ma è *necessario* fare così? Per quanto considerazioni di simmetria spingano in tal senso, o per quanti lanci di dado si siano osservati, resta pur sempre *soggettiva* la decisione di attenersi ad un dato criterio.

Supponiamo che un tale voglia sottoscrivere una polizza di assicurazione sulla vita. Poiché il premio annuo è valutato sulla probabilità che egli ha di morire entro un certo periodo, come dovrebbe essere valutata tale probabilità? Potremmo consultare le statistiche dei decessi relativi a individui 'simili' a lui sotto certi rispetti: dello stesso paese (o regione, o città, o quartiere), della stessa età (dello stesso anno, o nati in uno stesso intervallo di tempo), dello stesso sesso, dello stesso stato civile, della stessa professione (o reddito, o titolo di studio), delle stesse abitudini di vita (sport, alimentazione), dello stesso nome, dello stesso segno zodiacale, e così via. Ma dove terminano i requisiti atti a costituire la 'somiglianza' che ci interessa? Ad ogni diverso raggruppamento di requisiti faranno sicuramente riscontro differenti frequenze; dove trovare un criterio oggettivo di 'somiglianza' tra eventi?

Perciò, neppure la concezione frequentistica può vantarsi, a ben guardare, di essere immune da valutazioni soggettive ed arbitrarie. Del resto, «nessuna connessione tra probabilità e frequenza può avere un carattere empirico, poiché la frequenza osservata, quale che sia, è sempre compatibile con tutte le possibili opinioni circa la sua probabilità»[46]. Frequenza e probabilità, ben lungi dall'identificarsi, sono invece concetti differenti, e, se da un lato è indispensabile tener conto del peso dell'esperienza, dall'altro non si può negare che, seppure l'opinione *può* tener conto della frequenza, essa non si identifica mai con questa.

La risposta, per de Finetti, non può essere che una: «sono oggettive le circostanze su cui ci basiamo, ma la probabilità non è una loro conseguenza automatica bensì una valutazione in cui è soggettiva la scelta di

[45] Ivi, p. 259.
[46] B. de Finetti, *La previsione,* in *La logica dell'incerto*, cit., p. 97.

quali tra le circostanze oggettive note (ed in parte supposte) debbono o meno influenzare il giudizio»[47]. Dunque, dal momento che questo significato soggettivo della probabilità è «un fatto oggettivo e indiscutibile»[48], meglio è riconoscere subito che la probabilità non è altro che il grado di fiducia di un dato soggetto, in un dato istante, e con un dato insieme di informazioni, davanti ad una situazione di incertezza.

Non si tratta quindi, per de Finetti, di negare l'eventuale dipendenza della probabilità da osservazioni e previsioni di frequenza, ma piuttosto di dare soddisfazione all'esigenza di una migliore analisi di tali relazioni nella loro complessità, per non escludere tutti gli altri elementi che, oltre alla frequenza, possono e debbono influire sul nostro giudizio.

Ai suoi critici, de Finetti osserva:

> [...] non è che di tutte le circostanze oggettive di cui è formata l'informazione si voglia non tener conto per basarsi soltanto su sensazioni soggettive; al contrario, ci si deve preoccupare di tener conto di tutte le circostanze oggettive note – e in particolare di eventuali «simmetrie» che inducano a giudizi di equiprobabilità, e di osservazioni di frequenze che inducano a prevederne una certa stabilità – vagliandole però approfonditamente e responsabilmente anziché riducendole a rudimentali e presuntuose ricette[49].

La probabilità, per il soggettivismo, è sempre riferita ad uno stato di informazione soggettiva, o, se si preferisce, al grado soggettivo di ignoranza. Essa «non è una cosa in sé che si possa conoscere o ignorare; essa esiste in quanto serve ad esprimere proprio, da parte di ciascuno, ciò che egli sceglie nel suo dato stato d'ignoranza»[50]. Le asserzioni probabilistiche, pertanto, non esprimono fatti, ma opinioni, anche se esse assumono fatti come elementi di giudizio.

Sebbene nell'opinione individuale entrino anche complessi fattori psicologici ed emotivi, difficilmente ponderabili, non tutte le attribuzioni di probabilità saranno ammissibili per il soggettivista. Al contrario, ogni esercizio di valutazione e di previsione impegna l'individuo in un ragionamento, che consiste nel vagliare il pro e il contro di ogni scelta, e richiede che si considerino «ponderatamente tutte le alternative possibili per ripartire fra di esse nel modo che pare più appropriato le proprie

[47] B. de Finetti, *L'adozione della concezione soggettivistica come condizione necessaria e sufficiente per dissipare secolari pseudoproblemi,* in *Fondamenti del calcolo delle probabilità. Atti della Tavola Rotonda tenuta a Poppi nei giorni 11-12 giugno 1966*, Scuola di Statistica dell'Univ. di Firenze, p. 67.

[48] B. de Finetti, *Teoria delle probabilità*, cit., p. 256.

[49] B. de Finetti, *L'adozione della concezione….., cit.,*. in *Fondamenti…*, cit., p. 67.

[50] B. de Finetti, *Teoria delle probabilità,* cit.,, p. 103.

aspettative, le proprie sensazioni di probabilità»[51]. Affinamento della propria sensibilità probabilistica e rigorosa osservanza delle condizioni di coerenza devono guidare il soggettivista.

Il rispetto delle condizioni di coerenza, in particolare, costituisce un limite alla possibilità di valutazioni 'stravaganti' e rende di fatto impossibile una valutazione «stravagante» isolata, giacché ogni attribuzione di probabilità implica sempre l'attribuzione di una distribuzione di probabilità ad un intero insieme di ipotesi alternative, e tale distribuzione deve risultare coerente.

Sebbene i richiami ad una pacata ragionevolezza siano meno evidenti negli scritti giovanili, nei quali prevale un tono forse eccessivamente individualistico, non sembrano tuttavia giustificate le critiche miranti a fare della concezione di de Finetti una porta aperta all'arbitrio e all'irrazionalità. Il soggettivismo, almeno così come è inteso dal suo più eminente rappresentante, non intende affatto incoraggiare giudizi avventati, irrazionali o puramente emotivi; il 'soggettivo' è qui inteso, non come ciò che è opposto al 'ragionevole', bensì come ciò che implica l'intervento del giudizio individuale: tale giudizio può e deve essere quanto più ponderato possibile.

Come per Laplace, anche per i soggettivisti il senso della probabilità si lega al senso dell'ignoranza dell'uomo. In un mondo privo di uomini non esisterebbe la probabilità perché non esisterebbero né conoscenza né ignoranza, ma solo «fatti»; in un mondo di esseri «onniscienti», la probabilità non esisterebbe perché già tutto sarebbe previsto e prevedibile. È in relazione all'uomo, al suo stato di incertezza e di dubbio, ma anche alla sua volontà di dominio sul mondo, mediante la previsione e la teorizzazione, che acquista senso parlare di probabilità. Così la logica del probabile non è la logica dei fatti, ma la logica dell'opinione basata sui fatti.

[51] *Ivi*, p. 88.

Capitolo 3
L'approccio bayesiano

3.1 Il teorema di Bayes come atteggiamento induttivo

L'approccio bayesiano si caratterizza per il ricorso al teorema di Bayes, e alle sue implicazioni matematiche, in ogni tipo di questione induttiva, sia inversa che previsiva. In tale impostazione, come dice de Finetti, «il teorema di Bayes costituisce la chiave di volta e il concetto informatore di ogni attività costruttiva del pensiero»[1].

Del teorema di Bayes abbiamo già parlato in relazione alla concezione classica delle probabilità, ma può essere utile richiamarne la formula:

$$p(H/E) = \frac{p(H).p(E/H)}{\sum_i p(H_i).p(E/H_i)}$$

In essa, *p(H) è* detta probabilità *a priori, o* iniziale, dell'ipotesi, cioè prima di ogni informazione empirica, mentre *p(H/E),* detta probabilità *a posteriori* o finale, esprime la probabilità della stessa ipotesi dopo che sia stato osservato e considerato l'evento E.

Con $p(E/H)$ e con $p(E/H_i)$ sono invece indicate rispettivamente le probabilità dell'evento E, subordinatamente all'ipotesi che ci interessa ed a tutte le ipotesi concorrenti: normalmente ci si riferisce ad esse con il termine inglese *likelihood,* talora tradotto con il termine italiano «verosimiglianza». Useremo preferibilmente il termine inglese.

La probabilità a priori ed i *likelihoods* concorrono nella determinazione della probabilità finale *p(H/E)*. Quando la distribuzione *a priori* è uniforme, la formula si riduce evidentemente al solo quoziente dei *like-*

[1] B. de Finetti, *Teoria delle probabilità,* cit. , p. 563.

Maria Grazia Sandrini, *Filosofia dei metodi induttivi e logica della ricerca*, ISBN 978-88-6453-034-5 (print) ISBN 978-88-6453-037-6 (online)

lihoods. Ricordiamo anche che, per brevità, il denominatore è talvolta indicato semplicemente con *p(E)*, dal momento che sta ad indicare la probabilità di E, indipendentemente da qualsiasi ipotesi. Ciò non significa, ovviamente, che *p(E)* possa essere calcolato senza far riferimento ad un insieme ben determinato di ipotesi possibili: il valore di probabilità relativo alla disgiunzione di tutti i $p(E/H_i)$ possibili esprime appunto tale indipendenza da una qualsiasi ipotesi specifica.

La formula è, in un certo senso, un differente modo di scrivere il teorema delle probabilità composte e consente di affrontare il problema dell'inferenza inversa, ossia di calcolare la probabilità di una ipotesi generale (un'ipotesi riguardante una popolazione), sulla base delle informazioni disponibili, e di osservare di quanto muti questa stessa probabilità alla luce di nuove informazioni. Ma essa permette anche di affrontare le questioni di inferenza previsiva, ossia di valutare la probabilità che si verifichi uno o più eventi di un certo tipo, in base a precedenti dati osservativi analoghi. Definiamo così, una volta per tutte, *l'inferenza inversa* come inferenza da campioni noti alla popolazione ignota e l'*inferenza previsiva* come inferenza da campioni noti ad altri campioni non noti (o singoli eventi) di una stessa popolazione sconosciuta.

Se da un'urna abbiamo estratto *s* palline bianche ed *r* nere, qual è la probabilità di estrarre una nuova pallina bianca? Se la composizione dell'urna è conosciuta, la domanda si caratterizza come questione di inferenza diretta e trova una facile ed immediata risposta; ma se la composizione dell'urna è ignota, ci troviamo alle prese con una questione di inferenza inversa. Abbiamo visto che a questo problema Laplace dava soluzione mediante la regola di successione, qualora il numero delle osservazioni fosse sufficientemente grande.

Bayes vide con chiarezza che questo tipo di problemi poteva ridursi ad un semplice problema di probabilità composta, se fosse stato possibile assumere una qualche distribuzione iniziale di probabilità per l'insieme delle possibili ripartizioni delle palline nell'urna.

Tuttavia, sulla liceità di tale assunzione *a priori,* egli stesso manifestava qualche perplessità. Il teorema di Bayes, considerato da molti lo strumento in grado di indicare come si possa imparare dall'esperienza, ha trovato serie opposizioni alla sua applicazione, connesse quasi esclusivamente con le probabilità iniziali. I logici ed i probabilisti a Bayes posteriori si divisero su tale questione, sebbene nessuno abbia mai messo in discussione la validità matematica del teorema di Bayes. Agli occhi dei positivisti del XIX secolo il teorema apparve per lo più di ridotta applicabilità, limitandolo o all'ambito dei giuochi d'azzardo o a quei casi in cui i valori delle probabilità iniziali fossero empiricamente noti. Delle critiche oggettivistiche al teorema di Bayes parleremo più diffusamente nel prossimo capitolo.

L'importanza storica del teorema sta nel fatto che le ipotesi considerate possono essere interpretate come le possibili 'cause' degli eventi osservati.

Le condizioni di esaustività ed incompatibilità delle ipotesi assicurerebbe che, se E si verifica, una delle ipotesi dell'insieme considerato ne deve essere la causa.

Le ipotesi H_i devono perciò formare un insieme esaustivo di ipotesi a due a due incompatibili, ossia una *partizione*; di conseguenza, in accordo ai principi del calcolo delle probabilità, la somma di tutte le probabilità $p(H_i)$ deve essere uguale all'unità.

Osserviamo, inoltre, che, affinché si possa procedere al calcolo, ciascuna ipotesi deve essere tale da consentire di determinare il valore quantitativo del corrispondente *likelihood,* cioè della probabilità dell'evento subordinatamente ad essa. Il *likelihood* ha il ruolo di trasformare le probabilità iniziali delle ipotesi in probabilità finali.

È facile verificare che, dato un insieme esaustivo di ipotesi alternative e con distribuzione iniziale uniforme, avrà sempre probabilità finale più alta l'ipotesi sotto la quale l'evento osservato avrà ricevuto il più alto valore di *likelihood*.

Sulla base della sola osservazione della formula, noteremo anche che l'ipotesi riceve dall'evento E un grado di probabilità tanto più alto quanto più basso sia $p(E)$. Inoltre, un'ipotesi H sarà falsificata tanto nel caso banale in cui $H \rightarrow \neg E$, e quindi $p(E/H)=0$, ed E *si* verifica, quanto nel caso, altrettanto banale, in cui $H \rightarrow E$, e quindi $p(E/H) = 1$, ma E non si verifica.

Qualche esempio servirà a prendere maggiore dimestichezza con il teorema.

Esempio I: sia data un'urna, della quale si sappia solo che contiene N= 5 palline. Sono eseguite 3 estrazioni con reimbussolamento e si ottengono 2 palline bianche ed 1 non-bianca. Qual è la composizione più probabile dell'urna?

Le ipotesi possibili sono le seguenti, essendo state escluse dall'evidenza sia l'ipotesi che l'urna contenga tutte palline bianche sia quella che non ne contenga nessuna:

H_1 : $4B \wedge 1\neg B$
H_2 : $3B \wedge 2\neg B$
H_3 : $2B \wedge 3\neg B$
H_4 : $1B \wedge 4\neg B$

Se ora assumiamo che le ipotesi abbiano tutte una uguale probabilità iniziale, la formula si riduce al quoziente dei *likelihoods,* cioè:

$$p(H/E) = \frac{p(E/H)}{\sum_i p(E/H_i)}$$

Poiché: $p(E/H_1) = (4/5)^2 \cdot 1/5$; $p(E/H_2) = (3/5)^2 \cdot 2/5$; $p(E/H_3) = (2/5)^2 \cdot 3/5$; $p(E/H_4) = (1/5)^2 \cdot 4/5$, sostituiti i relativi valori nella formula, con pochi semplici calcoli otteniamo: $p(H_1/E) = 0{,}32$; $p(H_2/E) = 0{,}36$; $p(H_3/E) = 0{,}24$; $p(H_4/E) = 0{,}08$.

L'ipotesi più probabile è quindi proprio quella sotto la quale il campione estratto riceve il più alto valore di *likelihood,* come può facilmente esser verificato.

Esempio II: sulla base dei dati precedenti possiamo ora chiederci qual sia la probabilità di estrarre dalla stessa urna una ulteriore pallina bianca.

Poiché, come abbiamo visto in precedenza (cap.II), la formula è in questo caso la seguente,

$$p(E/n) = \sum_i p(H_i/n).p(E/H_i),$$

otteniamo i seguenti valori:

$$p(E/n) = 0{,}32.(4/5) + 0{,}36.(3/5) + 0{,}24.(2/5) + 0{,}08.(1/5) = 0{,}58.$$

Osserviamo, al proposito, che, in generale, l'inferenza previsiva, relativa sia ad una sola istanza sia ad un più numeroso campione, richiede sempre che sia presa in considerazione la composizione della popolazione di origine. Quando questa è nota, l'inferenza previsiva non è altro che inferenza diretta; quando non è nota, ma le possibili conformazioni della popolazione possono essere determinate, come nell'esempio ora discusso, il valore di probabilità richiesto è facilmente calcolabile; il calcolo si presenta più complesso nel caso di una popolazione ignota e infinita.

L'approccio bayesiano, sebbene si sia sviluppato all'interno della concezione classica e trovi oggi sostenitori soprattutto nella corrente soggettivistica, non è necessariamente connesso con tali impostazioni. Anche la concezione logicista è sostanzialmente bayesiana, ma non mancano sostenitori addirittura fra i frequentisti, alcuni dei quali hanno condiviso tale approccio all'inferenza inversa, pur conservando il loro atteggiamento empiristico di base. Per questo motivo, le varie concezioni della probabilità possono essere considerate anche come criteri per la valutazione delle probabilità a priori nelle applicazioni del teorema di Bayes: criteri basati sulle simmetrie, sul principio di indifferenza, sulle frequenze empiricamente rilevate, sui gradi di fiducia, o infine sulle regole della c*-funzione di conferma carnapiana, esprimono, tutti, differenti criteri per attribuire un valore quantitativo alle probabilità iniziali, nella formula di Bayes. È anche evidente che, sotto questo profilo, la posizione soggettivistica si impone per una particolare semplicità, concretezza ed atteggiamento pragmatico: dal momento che la probabilità è un grado di credenza, basterà fare una specie di esame interiore per esprimere la propria opinione, in relazione ad un insieme di ipotesi o di eventi, sebbene la concezione soggettivistica non escluda il ricorso anche a considerazioni di simmetria o di frequenza.

Tra i sostenitori della concezione frequentistica, H. Reichenbach merita un posto di primo piano per l'importanza che egli attribuisce alla formula di Bayes, da lui assunta addirittura quale espressione matematica del ragionamento induttivo e del modo in cui si attuerebbe il progresso della scienza, che, attraverso un processo per prove ed errori, si evolverebbe verso acquisizioni man mano sempre più probabili. Come nei romanzi polizieschi, vi sono alcuni dati ed alcune spiegazioni possibili; l'investigatore cerca di determinare la spiegazione più probabile. Le sue considerazioni seguono certe regole di probabilità: utilizzando tutte le informazioni, egli cerca di giungere a delle conclusioni che successivamente controlla mediante ulteriori osservazioni: ogni controllo, basato su nuovi dati, aumenta o diminuisce la probabilità della spiegazione prescelta. Naturalmente, non si possono attingere risultati quantitativamente precisi, se il materiale a disposizione consente soltanto valutazioni generiche di probabilità. Tuttavia, la formula di Bayes potrebbe, a parere di Reichenbach, essere applicata, almeno in senso qualitativo, in ogni tipo di induzione, perché ogni inferenza induttiva segue il processo di ragionamento che essa esprime in termini numerici.

In accordo con le esigenze dell'empirismo, Reichenbach ritiene che, in ogni caso, debba essere l'esperienza stessa a fornire, in termini di frequenze relative osservate, le probabilità iniziali delle ipotesi. Il processo conoscitivo si configura quindi, per lui, come una catena di ripetute applicazioni della formula di Bayes a partire dall'esperienza. Si tratta, idealmente, di un processo di lenta e progressiva approssimazione alla verità, ammesso che una verità ci sia. Tuttavia, l'esperienza è filtrata, in Reichenbach, da un soggettivo e consapevole atto di *assunzione* (*posit*), quello con cui le frequenze relative osservate sono assunte valere per il futuro, e quindi valere anche come probabilità iniziali, che stempera il suo empirismo e avvicina notevolmente la sua posizione a quella di de Finetti.

3.2 Soggettività ed intersoggettività nell'approccio bayesiano

Reichenbach e de Finetti rappresentano, per un certo verso, i due opposti esiti di uno stesso atteggiamento empiristico di fondo, che li porta a condividere la critica verso ogni atteggiamento metafisico e che li spinge, entrambi, verso forme di pragmatismo più o meno accentuato. Ma, mentre Reichenbach, mantenendo ferma l'illusione, forse in lui stesso contraddittoria, dell'oggettività del dato empirico, identificato nella frequenza relativa, si piega agli esiti più spiccatamente pragmatistici quasi soltanto di fronte al problema della giustificazione dell'induzione, de Finetti trae immediatamente dall'empirismo stesso la lezione di un soggettivismo radicale della conoscenza, preferendo appoggiare tale soggettivismo ad un pragmatismo che vada ad integrare, ma non a contraddire, quel radicale empirismo di fondo. Si comprende, allora, perché il modello privilegiato ed indiscusso del soggettivismo sia quello della scommessa; laddove, invece, molte critiche sono state mosse alla pretesa soggettivistica di fondare anche il ragionamento conoscitivo.

La conoscenza – si obietta – deve basarsi su fatti e non su opinioni. Come sarebbe possibile altrimenti garantirle quella oggettività che sembra essere la sua prerogativa? Ed inoltre, non viene addirittura negata, dal soggettivismo, la possibilità di un sapere comune, intersoggettivo, dal momento che le opinioni sono tante quante le teste che le elaborano? Pertanto, il soggettivismo è stato visto da molti come un atteggiamento antiscientifico.

Eppure, nel processo di 'apprendimento dall'esperienza' guidato dalla regola di Bayes, l'opinione soggettiva iniziale si evolve e si trasforma in modo sorprendentemente 'oggettivistico', come lo stesso de Finetti ha dimostrato in un suo ben noto teorema.

Per comprendere come ciò sia possibile, occorre pensare al processo conoscitivo come ad un progressivo accrescimento di informazioni e conseguente modifica delle opinioni precedentemente espresse: il processo conoscitivo si caratterizza dunque, tanto per de Finetti, quanto per Reichenbach, come una lunga catena inferenziale ottenuta mediante l'applicazione ripetuta del teorema di Bayes, con le probabilità finali, di volta in volta ottenute, funzionanti come probabilità iniziali nella successiva applicazione del teorema, al sopraggiungere di nuove informazioni.

È da notare che in tal modo l'approccio soggettivistico limita di fatto l'intervento della libera espressione individuale (ma pur sempre *coerente*!) alla sola prima (in senso temporale) valutazione delle probabilità iniziali. Ovvero, è sempre ammissibile, per il soggettivista, valutare ogni volta direttamente (cioè senza far ricorso al teorema di Bayes) le probabilità, ma le condizioni di coerenza impongono che il risultato debba essere esattamente lo stesso, sia che si calcolino le nuove probabilità (finali) di una ipotesi mediante il teorema di Bayes, sia che si calcolino quelle stesse probabilità direttamente, in modo soggettivo: le opinioni iniziali devono trasformarsi, al crescere delle informazioni, sempre in accordo alla regola di Bayes!

È così possibile dimostrare che un processo abbastanza lungo di ripetute applicazioni del teorema, a seguito di nuove informazioni, porta ad una convergenza dei valori delle probabilità finali anche nel caso di opinioni soggettive iniziali tra loro differenti in modo rilevante (ma purché, ovviamente, sempre maggiori di zero), espresse da individui diversi. Questo risultato fu dimostrato da de Finetti nel suo noto *teorema di rappresentazione,* che garantisce, sotto la condizione di *scambiabilità* degli eventi, una convergenza tra le valutazioni soggettive, dopo un numero di osservazioni che sarà tanto maggiore quanto maggiore è la differenza iniziale delle valutazioni medesime.

Così, se il soggettivismo non garantisce *a priori* l'intersoggettività, nemmeno la esclude: al crescere dell'esperienza tende a diventare sempre più irrilevante la distribuzione soggettiva iniziale, mentre acquista sempre maggior peso il fattore oggettivo contenuto nelle informazioni. Ciò è particolarmente visibile in relazione ai *grandi* e ai *piccoli* campioni: mentre resta, nel caso dei secondi, «la forte dipendenza delle conclusioni dall'opinione iniziale, nella cui valutazione diversi individui possono trovarsi lontani»[2], il peso della valutazione soggettiva iniziale diventa praticamente nullo nel caso di campioni di grande ampiezza.

Questa circostanza non pretende naturalmente di conferire significato oggettivo alla probabilità, ma può servire ad accreditare una immagine oggettiva, o quanto meno intersoggettiva, della conoscenza. Come per Laplace, anche per de Finetti il senso della probabilità si lega al senso dell'ignoranza dell'uomo. In un mondo 'oggettivo', ma privo di uomini, non esisterebbe la probabilità perché non esisterebbero né conoscenza né ignoranza, ma solo 'fatti', ed i fatti non sono né probabili né improbabili: semplicemente, 'sono'. Nello stesso mondo, popolato da uomini onniscienti, la probabilità non esisterebbe perché tutto sarebbe prevedibile. È perciò solo in relazione ad uno stato di dubbio, di incertezza, di ignoranza, che acquista senso parlare di probabilità: perciò la logica del probabile non è la logica dei fatti, ma *la logica dell'opinione basata sui fatti,* ed il teorema di Bayes ne dà la regola di trasformazione. «Il "ragionamento induttivo" [dice de Finetti] è tutto qui. Si suol dire che esso indica come si possa "imparare dall'esperienza", ed è ben detto, ma purché rimanga ben chiaro che l'esperienza non insegna a creare dal nulla un'opinione bensì soltanto ad adattare alla nuova situazione un'opinione preesistente»[3].

La concezione di Reichenbach, viceversa, sembra procedere piuttosto all'inverso: sebbene la probabilità sia sempre e solo questione di esperienza, cioè una frequenza osservata, non è possibile tuttavia prescindere da *assunzioni,* le quali non possono essere che soggettive, per quanto empiricamente fondate. Il processo conoscitivo si configura, per entrambi, come ripetizioni dell'applicazione della formula di Bayes, ma mentre per de Finetti si tratta di vedere quanto cambi la credenza alla luce dell'aumentare delle informazioni, per Reichenbach sembra trattarsi piuttosto di valutare quanto cambi l'esperienza, una volta che sia stata filtrata attraverso le assunzioni soggettive. Così, soggettività e oggettività si ripudiano e si richiamano reciprocamente nel difficile cammino della

[2] B. de Finetti,, *Teoria delle probabilità,,* cit., p. 621..
[3] Ivi*,* p. 568.

conoscenza, mostrando ancora una volta, ove ce ne fosse bisogno, quanto il terreno di un empirismo coerente a se stesso sia irto di difficoltà. A ben ragione, de Finetti afferma, parlando di Reichenbach: «in definitiva, fra la concezione del R. e la mia si ha piena concordanza nella posizione filosofica e scientifica attribuita alla teoria delle probabilità, discordanza nei metodi per realizzarla». E aggiunge: «la sua teoria potrebbe essere integralmente svolta dando alla probabilità il significato soggettivo, e verrebbe allora a coincidere sostanzialmente colla teoria delle probabilità quale appare dal mio punto di vista»[4].

3.3 Questioni di applicabilità del teorema di Bayes

Sebbene una linea di opposizione all'approccio bayesiano nasca, come si è accennato e come vedremo meglio nel prossimo capitolo, quale reazione oggettivistica alla definizione del concetto di probabilità come grado di credenza personale ed alla 'arbitrarietà' delle probabilità a priori, non sembrano esservi, in realtà, fondati motivi per limitare o interdire l'uso di valutazioni soggettive di probabilità a priori. Anzi, il ricorso ad esse sembra essere spesso utile, se non addirittura indispensabile. Del resto, i soggettivisti sembrano aver buon giuoco nel replicare che nessun approccio metodologico riesce a prescindere da assunzioni soggettive più o meno mascherate e che, sotto questo profilo, né la probabilità frequentistica né i metodi statistici 'oggettivisti', di cui parleremo, sono indenni da pesanti postulazioni arbitrarie.

Contrariamente dunque a quanto sostenuto dalla critica positivistica, non il ricorso alle probabilità a priori soggettive costituisce, a mio parere, un limite alle possibilità di applicazione del teorema di Bayes, quanto piuttosto aspetti e problemi che sembrano scaturire dalla pretesa di fare di tale formula uno strumento onnicomprensivo, identificandola con l'intero e complesso procedimento di crescita e di sviluppo della conoscenza, e sottomettendole, quindi, ogni tipo di questione che appaia rientrare nel campo dell'incertezza e dell'induzione.

Come abbiamo già detto, affinché il teorema di Bayes sia applicabile, è necessario che siano soddisfatte due condizioni fondamentali:

1. Le ipotesi devono formare un insieme esaustivo ed esclusivo, di modo che la somma delle loro probabilità iniziali sia uguale ad 1. Le alternative possibili devono cioè formare una partizione.
2. Ciascuna ipotesi deve essere tale che ne possano esser derivati valori numerici per i *likelihoods* degli eventi considerati.

[4] B. de Finetti, *Punti di vista: Hans Reichenbach*, (1941), in *La logica dell'incerto,* cit. , pp. 244 e 247.

Relativamente alla prima condizione, si può osservare che la partizione può essere finita, infinita o addirittura più che numerabile, a seconda del numero dei casi che si intenda considerare: ad esempio, una serie infinita di lanci di moneta darà luogo ad una partizione più che numerabile (cioè: 2^∞) di risultati possibili. Se ci riferiamo ad un'urna contenente *n* palline, delle quali ignoriamo la composizione relativamente al colore 'bianco', avremo $n+1$ composizioni possibili dell'urna (n, $n-1$, $n-2$,..., $n-n+1$, $n-n$ bianche). Essendo tali urne nient'altro che le alternative possibili a priori, il loro numero aumenterà al crescere di *n,* e sarà infinito per *n* infinito.

La determinazione dell'insieme di ipotesi alternative possibili è necessaria al fine di operare una distribuzione coerente dei valori iniziali di probabilità delle ipotesi, ma è pur vero che, per i soggettivisti, si può soddisfare a tale condizione di coerenza in vari modi. Per esempio, si possono trascurare alcune alternative possibili, attribuendo loro un valore iniziale di probabilità uguale a zero, oppure raggrupparne insieme un certo numero. In tal modo, il numero delle ipotesi da considerare può essere limitato a piacere, facilitando l'applicazione del teorema anche a casi non completamente determinati o infiniti: «la scelta di una classe di eventi è in sé arbitraria [...]. Di più, nulla vieta a priori di raggruppare l'evento che ci interessa con altri eventi di natura qualunque»[5].

Nella logica induttiva di Carnap, la determinazione delle ipotesi è connessa con la struttura del linguaggio, la cui scelta e costituzione precede e determina la formulazione stessa di ogni problema induttivo. Ne consegue che, mentre per i soggettivisti può essere talora di scarsa importanza l'estensione dell'universo in oggetto, per Carnap non è mai possibile prescindere dalla considerazione di tale estensione, che coincide appunto con l'assunzione di un determinato linguaggio: ciò inciderà, non solo sulla determinazione del problema e sul numero delle ipotesi alternative in gioco (che non possono essere ridotte a piacere, ma eventualmente solo raggruppate, una volta che il linguaggio sia stato assunto), ma anche sui valori delle probabilità *a priori*; ne è un esempio la *zero-conferma* di tutte le generalizzazioni universali in linguaggi infiniti[6].

In conclusione, il primo, indispensabile passo sembra dover consistere nella chiara determinazione dell'insieme delle ipotesi alternative. Laddove ciò non sia possibile, ovvero dove tale specificazione appaia troppo artificiosa per essere ragionevole, la formula di Bayes appare di difficile applicazione.

La specificazione dell'insieme delle ipotesi alternative rappresenta, in genere, «forse il più serio ostacolo alla generale applicazione di questo

[5] B. de Finetti, *La previsione,* in *La logica dell'incerto,* cit., p. 94.

[6] Cfr. R. Carnap, *Logical Foundations of Probability*, cit., pp. 570 sgg., e la discussione sull'argomento in M.G. Sandrini, *Probabilità e induzione,* cit., cap. 5.

approccio»[7]. Tale difficoltà è particolarmente evidente in campi, come quello della ricerca scientifica, dove non è sempre possibile determinare con rigore tutte le alternative possibili, o perché queste sono esprimibili soltanto qualitativamente, o perché la fase della ricerca non è ancora in grado di postularle. Tuttavia il calcolo delle probabilità esige che tale insieme di ipotesi sia determinato.

La questione della determinazione dell'insieme delle ipotesi da considerare investe anche la valutazione numerica dei *likelihoods* connessi con tali ipotesi, poiché tale valutazione è possibile soltanto quando si abbia a che fare con ipotesi parametriche ben determinate. Come possiamo, ad esempio, dire quale sia la probabilità che, in un campione di N post-infartuati trattati con antiaritmici, s di essi muoiano per aritmia, sotto l'ipotesi generica che «i farmaci antiaritmici migliorano la prognosi del post-infarto»?

La difficoltà di dare una valutazione quantitativa al *likelihood* di un evento appare particolarmente evidente quando le ipotesi siano vaghe o quando si sia interessati ad una singola ipotesi h, senza essere però in grado di collocarla in un insieme di ipotesi alternative.

In certi testi sull'argomento, questa difficoltà viene talvolta nascosta dietro l'uso di un apparato simbolico, indicando, ad esempio, con h l'ipotesi cui si è interessati e con $\neg h$ una non meglio specificata ipotesi alternativa.

Un esempio, ripreso da Howson ed Urbach[8], può servire a chiarire quanto ora affermato: l'esempio riguarda un esperimento su farmaci, nel quale sono considerati due gruppi di pazienti, uno trattato con un farmaco, l'altro con un placebo; è indicata con il simbolo h l'ipotesi che le differenze riscontrate, dopo un congruo periodo di trattamento, fra lo stato di salute dei pazienti appartenenti al primo gruppo e quello dei pazienti appartenenti al secondo, siano dovute all'effetto del farmaco. Sono indicati con e i risultati riscontrati alla fine del trattamento. Si vorrebbe calcolare, attraverso il teorema di Bayes, la probabilità dell'ipotesi h, dopo il trattamento. Come dovrebbe comportarsi il bayesiano a questo punto? I dati disponibili sono sufficienti per l'applicazione della formula di Bayes? In particolare, qual è l'alternativa all'ipotesi considerata? Gli autori da cui traggo l'esempio se la cavano inserendo, nella formula, il simbolo $\neg h$, per indicare un'ipotesi alternativa, che è lasciata però del tutto indeterminata (non è detto se $\neg h$ significhi che il farmaco non funziona oppure che le differenze riscontrate siano da attribuire al caso, o che altro). Gli Autori menzionati tralasciano di spiegare, inoltre, come possano essere calcolati,

[7] Cfr. G. McPherson, *Statistics in Scientific Investigation,* New York, 1990, p. 62.

[8] Cfr. C. Howson, P. Urbach, *Scientific Reasoning. The Bayesian Approach*, La Salle 1989, pp. 252 sgg.

con riferimento ad ipotesi così indeterminate, i valori di $p(e/h)$ e $p(e/\neg h)$: negligenza giustificata, a mio parere, dal momento che non si vede proprio con quale criterio tali valori potrebbero essere valutati!

In conclusione, l'ambito di applicazione, nel quale il teorema di Bayes appare ragionevolmente applicabile e di proficua utilità, sembra essere quello in cui si abbia a che fare con un insieme di ipotesi parametriche (tra le quali devono essere incluse anche quelle ipotesi di forma universale che altro non sono che ipotesi statistiche esprimenti una proporzione al 100%; ad esempio, «tutte le palline di quell'urna sono bianche»).

Si osservi, per inciso, che le questioni relative alla conformazione di una popolazione possono essere espresse in due modi differenti, per quanto connessi: ci si può chiedere «quale sia la probabilità che l'x% della popolazione abbia la caratteristica M», oppure ci si può domandare «quale sia la composizione della popolazione più attendibile (*stima*) relativamente alla caratteristica M». È evidente che, in base al teorema di Bayes, l'ipotesi più probabile esprime anche la stima migliore, essendo l'insieme di tutte le ipotesi alternative equivalente all'insieme dei possibili valori del parametro ignoto che vogliamo stimare.

In conclusione, il teorema di Bayes sembra applicarsi con successo a tutte le ipotesi quantitative (quelle stesse cui si applicano, come vedremo, i test di ipotesi ed i metodi di stima), con riferimento alle quali esso è in grado di offrire un supporto *giustificativo* per l'accettazione di una ipotesi.

3.4. Approccio bayesiano e induttivismo

Le concezioni della probabilità, per de Finetti, si dividono in due categorie:

> [...] quelle che potremmo chiamare *totalitarie* e cioè che intendono abbracciare tutto il significato in cui nel linguaggio corrente si usa il termine «probabile» [...], e quelle che si potrebbero dire *frammentarie,* in quanto tendono a ridurre il campo della teoria delle probabilità a particolari studi ed applicazioni: la irriducibilità concettuale fra queste due categorie è ben più profonda [...], perché, nella prima, la teoria delle probabilità contiene in sé anche il problema della giustificazione delle forme di ragionamento cui conduce e sulle quali si basano tutte le scienze applicative (non tautologiche) ed assurge perciò al rango di logica formale [...] come strumento primordiale del pensiero e come premessa teoretica alla stessa possibilità di qualunque scienza induttiva[9].

[9] B. de Finetti, *La logica dell'incerto*, cit., p. 238.

Non vi è dubbio che tanto de Finetti, quanto Reichenbach, debbano essere ascritti in questa categoria!

Di Reichenbach si è già, in gran parte, detto: per lui la formula di Bayes rispecchia la natura stessa dell'inferenza induttiva, entro la quale la convalida inferenziale delle teorie scientifiche non è che un caso particolare[10].

Perciò, per Reichenbach, si può, anzi si deve, parlare di probabilità delle teorie scientifiche, che devono essere scelte tra le diverse possibili spiegazioni dei dati empirici a disposizione. Una valutazione quantitativa a priori della probabilità di una teoria scientifica potrebbe essere fatta, secondo lui, in base alla frequenza delle teorie scientifiche risultate vere o false in una successione illimitata[11]. Questa proposta appare aperta a molteplici critiche demolitrici (e non solo a quelle riguardanti la probabilità come limite di frequenza relativa, mossegli da de Finetti)[12], sulle quali non mi sembra il caso di soffermarsi, e comunque troppo debole per poter essere presa in seria considerazione.

Il senso di questo esasperato induttivismo (userò questo termine per indicare la riduzione di tutta la conoscenza a processo induttivo) va considerato sia alla luce di una concezione descrittivistica della conoscenza, per la quale anche le teorie rappresentano ipotesi circa stati del mondo, sia alla luce della netta scissione, operata da Reichenbach, tra contesto della scoperta e contesto della giustificazione, e della conseguente riduzione di tutta la logica inferenziale induttiva al contesto giustificativo. Poiché il contesto della scoperta sembra a Reichenbach sfuggire alla presa della logica, alla quale non può essere richiesto di svolgere «la funzione creativa del genio», il metodo del progresso scientifico, nei limiti in cui esso può essere sottoposto ad analisi logica, entro l'orizzonte di una conoscenza solo probabile, verrebbe ad identificarsi con il metodo con il quale lo scienziato 'giustifica', a sé e agli altri, l'accettazione di una teoria: la regola di Bayes, appunto.

Anche per de Finetti, «il campo dell'induzione si estende in ogni ambito e ad ogni livello», ma egli non sembra condividere del tutto il punto di vista di Reichenbach. Particolarmente istruttivo gli pare, infatti, il riflettere sul

> [...] processo con cui nuove concezioni scientifiche vengono formulate in base a intuizioni suggerite da qualche circostanza osservata, e poi discusse spesso con alterne vicende in base a nuove risultanze volta a volta meglio spiegabili con questa o quella teoria. In sostanza si tratta sempre di un'analisi della situazione delle conoscenze fatta in base al

[10] Cfr. H. Reichenbach, *La nascita della filosofia scientifica*, cit., p. 225.

[11] Cfr. H. Reichenbach, *Experience and prediction,* Chicago University Press, 1938, p. 397.

[12] Cfr. B. de Finetti, *La logica dell'incerto,* cit., pp. 237-247.

> teorema di Bayes, salvo che, in situazioni così larghe e vaghe, sarebbe impossibile pensarne un'applicazione non meramente qualitativa[13].

Sarebbe di estremo interesse, egli aggiunge in una sorta di inciso, un'analisi approfondita del modo in cui si evolve il pensiero scientifico. A suo parere, esso dovrebbe consistere nella «storia di una serie incessante di capovolgimenti di concezioni, maturati o intuiti grazie alla riflessione su risultati imprevisti da parte di menti solitarie, accolti in genere con ostilità, incomprensione e diffidenza finché con l'accumularsi delle prove a favore e il perfezionamento delle formulazioni teoriche non pervengono a imporsi»[14]. Altro aspetto, più strettamente connesso con la teoria delle probabilità, «sarebbe quello di vagliare tali processi di scoperta ed accettazione alla luce delle basi probabilistiche del ragionamento induttivo». Siamo, comunque, a suo parere, in situazioni troppo vaghe per permettere una valutazione delle probabilità in gioco; tuttavia, egli non ha dubbi circa il «carattere ineluttabilmente incerto e ipotetico di tutte le "verità" scientifiche e circa la necessità di basare pertanto ogni ragionamento sulla logica probabilistica». Il discorso, però, gli sembra esaurirsi «con l'illustrazione e l'approfondimento di questa tesi, senza affrontare o menzionare la possibilità di applicarla ai grandi problemi di sintesi, ai problemi di scelta fra teorie contrastanti»[15].

Contesto della scoperta e contesto di giustificazione, livello di pura generalizzazione empirica e livello metalinguistico di spiegazione teorica, induzioni all'interno di una teoria accettata e scontro di teorie rivali, momenti di scienza 'normale' e momenti di rivoluzione scientifica (per dirla con Kuhn), tutto sembra mescolarsi e combinarsi in modo caotico in questa visione induttivistica della conoscenza e del processo scientifico, evocata dalle parole di de Finetti. Tuttavia, malgrado il fascino che evidentemente esercita su di lui questa sorta di totalitarismo bayesiano, de Finetti esprime una più pacata cautela nelle pagine immediatamente successive, dove sembra prendere corpo una implicita separazione tra il campo teorico, e idealizzato, del ragionamento induttivo e quello effettivo, nel quale l'applicazione della teoria delle probabilità risulti davvero adeguata e possibile:

> [...] il campo in cui il ragionamento induttivo si presta più specificamente all'applicazione del calcolo delle probabilità sotto forma di sviluppi tecnici è quello più modesto concernente problemi particolari nell'ambito di un'impostazione già accettata. E a questo tipo di problemi ci limiteremo[16].

13 B. de Finetti, *Teoria delle probabilità,* cit. vol. II, p. 559.
14 Ivi, p. 560. Evidente l'influsso di Kuhn!
15 Ivi, p. 561.
16 Ivi, p. 562.

Il discorso concernente la funzione determinante della teoria delle probabilità nel processo scientifico rimanda ad una concezione epistemologica, ampiamente condivisa nella tradizione, per la quale tutto il progresso scientifico si svolgerebbe per induzione progressiva (in definitiva per induzione per enumerazione), partendo dall'osservazione ed estendendo man mano le leggi trovate a campi sempre più vasti di fenomeni; in breve, un processo di accumulazione progressiva della conoscenza: le nuove teorie inglobano le precedenti come propria parte, spiegandole o estendendone il campo di validità. In questa concezione sembra esservi poco spazio per l'invenzione creativa, che non sia quella di trovare nuove formule per sintetizzare, ad un livello più ampio e più fecondo, i fenomeni osservati. In ogni caso, tutta la conoscenza empirica si baserebbe e si svilupperebbe esclusivamente per ragionamento induttivo. È appunto ciò che ho indicato con il termine induttivismo.

In generale, sul piano filosofico, l'impostazione induttivistica sembra spesso rinviare ad una concezione ingenuamente realistica della conoscenza, spesso tacitamente condivisa, anche se con una certa incoerenza, dall'empirismo: una conoscenza concepita come *corpus* di descrizioni del mondo e di leggi quanto più possibile 'probabili' (dove «probabile» indica un grado di approssimazione alla verità), almeno in relazione allo sviluppo scientifico della propria epoca. Questo atteggiamento epistemologico non è in genere molto sensibile alle differenze strutturali all'interno della sistematizzazione scientifica: tutte le asserzioni della scienza riguardano aspetti del mondo, ne sono una descrizione probabile, possibilmente probabile ad un grado elevato; lo scopo della scienza altro non è che un'approssimazione continua alla verità.

Va tenuto presente, del resto, che mentre l'articolazione teorica della scienza contemporanea si è sviluppata velocemente a partire da tempi relativamente recenti, questa tradizione filosofica ha le sue radici in tempi assai più lontani. Accade così che generalizzazioni empiriche, leggi e teorie siano trattate spesso come espressioni sinonime, almeno in relazione al metodo con cui dovrebbero essere confermate e accettate: ne consegue, allora, che abbia senso parlare della probabilità della teoria della relatività, o della meccanica di Newton, esattamente come lo ha il parlare della probabilità che l'altezza media degli italiani maschi superi i 170 cm, o che la produzione della FIAT nel primo trimestre dell'anno 2010 sia sotto controllo!

È questa l'eredità raccolta da autori come Reichenbach, Feigl, Salmon, ed altri[17], ed è a questa immagine della conoscenza che reagiscono, non a torto, e ciascuno a suo modo, tanto Popper quanto Kuhn.

[17] Ispirata da un'idea induttivistica è forse anche la logica di Hintikka, che persegue il tentativo di trovare un criterio che permetta di attribuire alle leggi di natura un valore di probabilità *a priori* diverso da zero, di modo che sia possibile sottoporre al teorema di Bayes la questione dell'accettazione delle medesime. Vero è che, in ogni

Lo sfondo empiristico di partenza, che il soggettivismo può solo velare, ma non nascondere, condiziona anche l'analisi epistemologica di de Finetti: se l'obiettivo della conoscenza è quello di pervenire a descrizioni del mondo sostenute dai fatti, anche le teorie, nella misura in cui esprimano conoscenza, e cioè ipotesi circa stati del mondo, dovrebbero essere sottoposte al calcolo delle probabilità, e la loro accettazione dovrebbe dipendere dal grado di questa. La consapevolezza dei limiti conoscitivi umani porta, poi, a riconoscere che la logica che guida la conoscenza non può essere che la logica dell'incerto, ovvero il calcolo delle probabilità; ma non incrina, esattamente come nella concezione di Laplace, la fiducia in uno stato di cose del mondo, del tutto autonomo e indipendente dall'uomo.

Alcuni, come de Finetti, pur attratti da un atteggiamento di tipo 'totalitario', mostrano tuttavia una certa cautela, riconoscendo la complessità delle teorie e la difficoltà di attribuir loro gradi quantitativi di probabilità. Nessuno, però, sembra porsi, come passo preliminare, le due seguenti questioni: *se* abbia senso parlare di probabilità delle teorie scientifiche e *se* esse possano soddisfare le condizioni richieste dal calcolo delle probabilità. Eppure, contro l'approccio induttivistico alle teorie non mancano, né sul piano logico-epistemologico, né su quello filosofico, argomenti di indiscutibile forza. Basti ricordare la tesi della sottodeterminazione empirica delle teorie, che dovrebbe bastare, da sola, a sollevare simili interrogativi.

3.5 Teorema di Bayes e teorie rivali

Tralasciamo, per ora, la prima delle due questioni ora indicate, per affrontare quella, più 'tecnica', dell'effettiva possibilità di applicazione del teorema di Bayes alle teorie scientifiche.

Ci serviremo, per questa discussione, soprattutto del volume di Howson e Urbach[18], i quali hanno discusso il problema della scelta tra *teorie* rivali (qualcosa, per intenderci, come la teoria della relatività) per mezzo del teorema di Bayes. Questi autori sostengono, e tendono a dimostrare, che il processo di scelta fra teorie rivali corrisponde alla logica bayesiana; ma le loro stesse conclusioni sembrano mettere in luce, viceversa, proprio l'inutilità del teorema di Bayes in questo tipo di questioni e come le teorie scientifiche non possano essere accettate sulla base di un metodo probabilistico.

caso, Hintikka si riferisce alle leggi empiriche e non alle «teorie» scientifiche. Cfr. J. Hintikka, *Towards a Theory of Inductive Generalization,* in Y. Bar-Hillel (a cura di), *Logic, Methodology and Philosophy of Science,* Amsterdam 1965; J. Hintikka, P. Suppes (a cura di), *Aspects of Inductive Logic,* Amsterdam 1966.

[18] C. Howson, P. Urbach, *Scientific Reasoning*, cit.

Dichiarando esplicitamente gli autori di non fare distinzione tra «ipotesi» e «teoria», ritengo di restare fedele al loro intendimento, definendo, sul piano formale, le teorie deterministiche come enunciati di forma universale, tali da non ammettere eccezioni, e le teorie statistiche come enunciati generali di forma statistica.

È facile constatare, ora, quanto segue, una volta concesso, per amore di discussione, che possano essere attribuiti valori quantitativi a *tutti* i fattori che compaiono nel teorema di Bayes, e cioè tanto alle probabilità a priori, quanto ai *likelihoods* corrispondenti a ciascuna teoria.

Deve essere inoltre precisato, anche se dovrebbe essere inutile, che il riferimento a «teorie rivali» non costituisce un caso particolare dell'argomento «probabilità di una teoria», sebbene l'espressione possa suggerire l'inverso. In realtà, affinché sia lecito parlare di probabilità di un qualsiasi tipo di ipotesi, occorre sempre determinare, come prima cosa, l'insieme delle ipotesi alternative; nessuna ipotesi (o teoria), perciò, può essere considerata e valutata da sola, per mezzo del calcolo delle probabilità, ma ogni ipotesi è sempre valutata rispetto ad un insieme di ipotesi concorrenti! L'accettazione o il rifiuto di una ipotesi isolata può, sotto il profilo logico, essere eventualmente prefigurata mediante il metodo ipotetico-deduttivo, ma non è possibile con il ricorso al calcolo delle probabilità. Perciò, se non si hanno 'rivali' ad una ipotesi data, non è neppure possibile sottoporla alla regola di Bayes. Ciò costituisce una prima, ma non piccola difficoltà, quando si vuol parlare di effettive *teorie* scientifiche (e non di ipotesi empiriche), perché è assai difficile, se non praticamente impossibile, determinare un insieme di «teorie» rivali che soddisfino le condizioni richieste dal calcolo delle probabilità.

Ciò premesso, si possono avere, sotto il profilo formale, le seguenti situazioni (sarà usato, qui di seguito, il termine generico «ipotesi», che meglio si adatta, a mio parere, alla forma enunciativa sopra indicata, mentre il termine «teoria» sarà riservato per indicare il suo significato specifico):

a) scelta tra ipotesi rivali di forma deterministica;
b) scelta tra una ipotesi deterministica ed una (o più) ipotesi statistica;
c) scelta tra ipotesi rivali statistiche.

Analizziamole caso per caso:

a) Date due (o più) ipotesi deterministiche, non contraddette dall'esperienza, nessuna informazione empirica sarà in grado di trasformare i valori di probabilità iniziali, qualunque essi siano. Infatti: essendo

$p(e/h) = p(e/\neg h) = 1$ e $p(h)+p(\neg h) = 1$, allora

$p(h/e) = p(h)/p(h)+p(\neg h) = p(h)$.

Anche Howson e Urbach rilevano tale fatto, ma, perseverando nella loro posizione, aggiungono: «il problema è allora quello di scoprire i criteri razionali con cui le teorie assumono particolari probabilità a priori»[19]. Il che equivale a dire, a mio avviso, che occorre trovare un modo indipendente da quello bayesiano per decidere tra teorie rivali deterministiche, perché è evidente che prevarrà sempre, con la regola di Bayes, quella teoria alla quale sarà stato attribuito un più alto valore di probabilità a priori!

b) Tra una ipotesi deterministica ed una ipotesi statistica, è facile verificare che la prima, essendo connessa con un valore di *likelihood* più alto, avrà sempre probabilità finale maggiore, sotto l'assunzione di una distribuzione iniziale uniforme. Ove si assuma, invece, una distribuzione iniziale non-uniforme a sfavore dell'ipotesi deterministica, quest'ultima vedrà aumentare il proprio valore di probabilità a posteriori ad ogni nuova applicazione del calcolo, fino a prevalere nettamente se il procedimento sarà ripetuto un numero sufficiente di volte (e ciò dipenderà dalla misura dello squilibrio iniziale), a patto che, ovviamente, l'evidenza empirica continui ad essere con essa compatibile.

c) Tra due ipotesi statistiche, prevarrà quella cui è connesso un più alto valore di *likelihood*: fin dalla prima applicazione della formula, se si parte da una distribuzione iniziale uniforme o favorevole all'ipotesi con più alto *likelihood*; con un sufficiente numero di ripetizioni del procedimento di conferma, altrimenti.

In conclusione, l'approccio bayesiano per dirigere la scelta tra 'teorie' rivali appare alquanto inconcludente ed i suoi risultati del tutto prevedibili a priori, almeno che non ci si imbatta in eventi che chiaramente contraddicano le ipotesi in gioco. Ma anche in tal caso l'uso del teorema di Bayes non sarebbe di nessuna utilità.

Naturalmente, quanto detto sopra non è che una conseguenza dei rapporti tra probabilità nel teorema di Bayes ed è valido in generale; ma, nelle applicazioni del teorema ad ipotesi empiriche (e non a teorie), la partizione delle ipotesi alternative è in genere più ampia e ben definita; inoltre, le osservazioni consistono in campioni estratti da una popolazione ignota. Le teorie scientifiche, invece, sono ideate in conformità ad un certo campo di fenomeni o ad un insieme di leggi empiriche che si vuole spie-

[19] Ivi, p. 117.

gare e con i quali intrattengono una relazione logica di natura differente. È difficile dire in che senso una teoria scientifica, specie se ad un livello elevato di astrazione, potrebbe essere considerata come una popolazione ignota, alla stessa stregua delle urne di Laplace!

Sotto il profilo dell'applicabilità del teorema di Bayes, dobbiamo dunque ripetere quanto già rilevato: il teorema di Bayes può essere applicato adeguatamente a situazioni che soddisfino le condizioni in precedenza esplicitate. Ciò rimanda, in primo luogo, come già detto, alla questione della determinazione completa e rigorosa dell'insieme delle ipotesi alternative, a quella partizione, sulla quale de Finetti giustamente insiste, che deve esprimere un insieme di ipotesi incompatibili ed esaustive. Ma, riferendosi alle teorie, come è possibile garantire, e in modo non semplicemente simbolico, che le teorie assunte come rivali indichino un insieme «ragionevole» di ipotesi alternative, fra di loro incompatibili ed esaustive? Se con h indichiamo la teoria della relatività, a cosa dovrebbe corrispondere il simbolo $\neg h$, con il quale gli autori in discussione indicano l'ipotesi rivale? Fino a che punto è ragionevole restringere o estendere a piacere l'insieme delle ipotesi alternative?

Assumendo come modello di partizione corretta le urne di Laplace, appare discutibile, ad esempio, che si possa assumere come universo del discorso un insieme di ipotesi rivali, tutte di forma universale: sarebbe come prendere in considerazione soltanto le ipotesi «tutte le palline sono bianche» e «nessuna pallina è bianca», due ipotesi, fra l'altro, che non possono essere entrambe compatibili neppure con l'estrazione di una sola pallina; perciò, avvenuta la prima estrazione, una delle due ipotesi cadrebbe inevitabilmente. Le teorie scientifiche rivali, seppure tra loro incompatibili, devono, al contrario, essere tutte compatibili con uno stesso insieme di fenomeni. Appare difficile, insomma, spiegare come sia possibile sottoporre al teorema di Bayes ipotesi rivali, tutte di forma deterministica, perché è impossibile stabilire come esse potrebbero formare quella partizione che è condizione necessaria per l'applicazione del calcolo delle probabilità! Considerazioni quasi analoghe possono essere fatte in relazioni a teorie rivali non, o non tutte, di forma deterministica, sebbene la contrapposizione tra ipotesi di forma deterministica e ipotesi statistiche appaia formalmente più vicina al senso di una partizione correttamente stabilita.

Per quanto concerne, poi, la questione dell'esaustività delle ipotesi, possiamo ricordare che la limitazione arbitraria del *numero* (ma solo di questo!) delle alternative è consentito soltanto all'interno dell'approccio soggettivistico (le ipotesi escluse sono ipotesi cui soggettivamente è attribuito un valore di probabilità a priori uguale a zero), che, se da un lato facilita la valutazione delle probabilità a priori, non tutti i fautori dell'induttivismo bayesiano sarebbero disposti a condividere, perché introduce una valutazione altamente soggettiva circa la credibilità di alcune ipotesi.

In ogni caso, non sembra ragionevole pensare che il soddisfacimento della condizione relativa alla partizione delle ipotesi possa consistere nel solo fatto, del tutto formale, che la somma delle probabilità iniziali delle ipotesi considerate sia uguale ad 1; viceversa, tale somma deve essere uguale all'unità *perché* la condizione di esaustività ed incompatibilità è stata soddisfatta.

Mi sembra utile ricordare, a questo punto del nostro discorso, una particolare modalità di applicazione del teorema di Bayes all'inferenza inversa, da parte di Laplace.

La prima modalità è quella riconducibile al modello dell'urna, di cui abbiamo già parlato: da un'urna, della quale sappiamo solo che contiene N palline, abbiamo estratto, con reimbussolamento, s palline bianche. Qual è la probabilità che l'urna contenga tutte palline bianche? Questo è il modello classico dell'inferenza induttiva, generalizzabile in ampi settori di applicazione, che non presenta difficoltà apprezzabili, né sotto il profilo logico, né sotto quello matematico: si tratta di un'inferenza circa la composizione possibile di una popolazione, ovvero di una generalizzazione empirica induttiva. Qui, naturalmente, l'insieme delle ipotesi alternative è formato da tutte le distribuzioni possibili di palline bianche nell'urna: N bianche, (N-1) bianche, (N-2) bianche, e così via fino a (N-N) bianche: si tratta, come si vede bene, di un insieme esaustivo di ipotesi incompatibili.

La seconda modalità, invece, è di tutt'altro tipo ed è assai interessante. Laplace l'ha usata frequentemente per trattare questioni di scienza naturale: fenomeni celesti, le maree, e altri fenomeni ancora. Questo genere di applicazione non appartiene alla classica tipologia induttiva e potremmo addirittura dire che, pur essendo un procedimento probabilistico, è soltanto indirettamente induttivo.

Per illustrarlo, basterà ricordare, fra tutte, la nota applicazione del teorema ai movimenti di rotazione e di rivoluzione dei pianeti e dei satelliti, con la quale Laplace mostrò essere estremamente improbabile (1 su circa 4.000 miliardi!) che tali movimenti si svolgessero tutti nello stesso senso della rotazione del Sole per effetto del puro caso.

Assunte come ipotesi alternative, con uguale probabilità a priori, il Caso e la Causa (con la quale egli intendeva la teoria di gravitazione universale), e assumendo inoltre il sistema planetario come composto dal Sole e da 11 pianeti e 18 satelliti, per un totale di 59 movimenti, dei quali ne erano all'epoca noti 43, Laplace procedeva a calcolare le probabilità (*likelihoods*) dei movimenti dei pianeti e dei satelliti sotto entrambe le ipotesi: dall'ipotesi deterministica discende un valore pari all'unità per il *likelihood* dell'insieme dei fenomeni in oggetto; più complesso il calcolo relativo alla probabilità dello stesso insieme di fenomeni, sotto l'ipotesi del Caso. Poiché qui si tratta di calcolare in quante, fra tutte le possibili combinazioni casuali dei movimenti, il fenomeno sia valido, basterà considerare solo le combinazioni che comportano variazioni tra i 16 movimenti non noti; troviamo pertanto il seguente valore per il *likelihood* dei fenomeni considerati, sotto l'ipotesi casuale:

$$\frac{1+2+2^2+2^3+2^4+2^5+2^6+2^7+2^8+2^9+2^{10}+2^{11}+2^{12}+2^{13}+2^{14}+2^{15}+2^{16}}{2^{59}} \approx \frac{1}{4398\,Mld}$$

con approssimazione per difetto.

Applicando quindi il teorema di Bayes, si vede che questo è anche il valore, approssimato, della probabilità dell'ipotesi che i movimenti in questione siano governati dal Caso.

È difficile dire quale ruolo effettivo, nell'ambito della scienza, Laplace attribuisse a questo procedimento, con il quale, come appare evidente, è solo possibile calcolare l'improbabilità del Caso, ma non è *mai* possibile giungere ragionevolmente ad una attribuzione di probabilità per una ipotesi deterministica specificata: qualsiasi teoria di forma deterministica ne risulterebbe infatti supportata allo stesso grado! È evidente, pertanto, che questo uso della formula di Bayes non può servire, e certamente non serviva neppure a Laplace, per giustificare l'accettazione di una teoria, nella fattispecie quella di gravitazione, ma piuttosto per escludere il Caso: con ciò Laplace forse riteneva di apportare una qualche giustificazione indiretta a favore di una concezione deterministica generale del mondo.

Questo uso, che possiamo definire improprio, del teorema di Bayes mi sembra tuttavia di estremo interesse: da un lato, anticipa la logica di quei test di significatività, che saranno sviluppati nei primi anni del XX secolo e che trovano ancor oggi ampia applicazione in molti campi della ricerca scientifica, e dei quali parleremo nel prossimo capitolo; dall'altro, contribuisce ad evidenziare la differenza già da me suggerita, e sulla quale torneremo in seguito, tra differenti contesti della ricerca, aiutando contemporaneamente a delimitare gli ambiti di applicazione delle metodologie induttive, e riducendone di conseguenza anche le pretese totalitaristiche.

Ma l'esempio di Laplace evidenzia anche un altro aspetto, sul quale mi sembra importante insistere: la difficoltà 'tecnica' di pervenire a calcolare la probabilità di sistemi complessi quali sono le teorie scientifiche.

3.6 Probabilità e teorie scientifiche

Veniamo ora, sia pur brevemente, alla questione lasciata in sospeso: ha senso parlare, con riferimento allo sviluppo della scienza contemporanea, di 'probabilità' di una teoria scientifica?

La risposta può variare, naturalmente, con il significato attribuito al termine «teoria», o meglio, per essere più precisi, a seconda della struttura formale e delle funzioni che associamo a questo termine.

Possiamo dire che la conoscenza ha avuto origine con i primi, talora ingenui, tentativi di generalizzare l'esperienza; un compito la cui utilità si rivelava soprattutto in relazione alle sue capacità predittive e, quindi, al contributo positivo per la stessa sussistenza umana. Forse ha ragione Hume, al riguardo, quando sostiene che la conoscenza sia solo frutto dell'abitudine basata sull'esperienza. La scienza, invece, sembra avere inizio con il desiderio di andare oltre le semplici generalizzazioni empiriche, per trovarne, per così dire, la 'causa'. Sembra nascere, cioè, da curiosità più profonde, da domande, non dall'osservazione. Questa ricerca di spiegazione è sfociata, agli inizi, nella formulazione di generalizzazioni di maggiore estensione, dalla quale le generalizzazioni ricavate direttamente dall'esperienza potevano esser fatte derivare. Così, ad esempio, la legge di Archimede sulla spinta dei fluidi «spiega» perché il ghiaccio ed il legno galleggino sull'acqua, ed anche perché possa galleggiare sul mare una nave di metallo, ma non una palla, non cava, di piombo. Pur unificando molteplici generalizzazioni induttive ricavate dall'esperienza (del tipo: per ogni *x*, se *x* è P, allora *x* galleggia), la legge di Archimede si presenta ancora come una legge empirica, estesa ad ogni corpo immerso in un liquido. Non nasce, però, come diretta generalizzazione empirica, ma piuttosto come ipotesi da sottoporre a prova mediante l'osservazione 'sperimentale' dei più disparati tipi di corpi. Una volta affermata, però, essa rende scientificamente inutili le singole generalizzazioni empiriche che essa spiega.

La scienza comincia così a costruire una gerarchia di leggi, a partire dalle generalizzazioni empiriche dirette. E comincia anche, lentamente, a costruire un metodo sperimentale e strumenti idonei.

In un processo di generalizzazione sempre più estesa ed astratta, troviamo che le teorie corpuscolare e ondulatoria della luce offrono una spiegazione delle leggi della riflessione, della rifrazione e della diffrazione e della propagazione rettilinea; la teoria gravitazionale di Newton riesce a spiegare le leggi del moto dei pianeti, della caduta libera dei gravi, dell'azione delle maree, delle forme delle masse rotanti, del galleggiamento dei liquidi e dei gas, dei fenomeni di capillarità, delle proprietà termiche dei gas, ed altre ancora. La teoria dei quanti è in grado di spiegare le leggi sperimentali dei fenomeni spettrali, delle proprietà termiche dei solidi e dei gas, della radioattività, delle interazioni chimiche e di molti fenomeni ancora.

La domanda, che affiora a questo punto, mi sembra la seguente: ferma restando la sempre maggiore generalità, che distingue le teorie dalle leggi empiriche, ma anche le leggi tra loro, vi sono altre caratteristiche determinanti e peculiari delle teorie?

Una prima osservazione consiste nel rilevare che, mentre le generalizzazioni e le leggi di infimo ordine (le quali ultime presentano uno status logico ambivalente, un po' descrizione di regolarità fenomeniche, un po' parti costituenti della complessiva intelaiatura teorica) muovono dall'osservazione e implicano direttamente specifiche osservazioni, ciò non vale per le leggi di ordine superiore, relativamente alle quali sembra aver senso non tanto il chiedersi se esse siano vere o probabili, quanto piuttosto in quali casi esse siano valide: ciò permette che si possa parlare di anomalie, cioè di deviazioni dalla legge, e di limitazioni nella loro applicazione, senza che ciò le confuti.

Si possono indicare ancora diverse altre caratteristiche peculiari delle teorie e non delle leggi e delle generalizzazioni empiriche.

Una differenza, essenziale per alcuni, consiste negli scopi che le teorie, ma non le leggi, sono chiamate ad assolvere: esse spiegano e riuniscono leggi sperimentali già stabilite, ma devono anche essere in grado di suggerire nuove leggi sperimentali e nuove linee di ricerca.

Tuttavia, la differenza più rilevante, che pone le teorie su un piano del tutto differente da quello delle leggi, riguarda, a mio parere, la loro struttura. Una legge sperimentale, così come ogni tipo di generalizzazione empirica, è sempre formulata in un enunciato singolo. Ciò non è possibile, nel caso delle teorie.

Per Hempel, ad esempio, una teoria è paragonabile ad una complessa rete sospesa nello spazio, i cui nodi corrispondono ai termini ed i fili alle definizioni ed alle ipotesi fondamentali derivate dalla teoria. L'intero sistema fluttua sul piano dell'osservazione, cui è ancorato mediante le regole interpretative[20].

Per Nagel[21], una teoria si presenta come una struttura logica e matematica astratta, collegata ad un insieme di differenti modelli e applicazioni. Essa evidenzia le connessioni sistematiche tra leggi sperimentali che riguardano argomenti qualitativamente disparati. Possiede perciò maggiore generalità e maggiore potere esplicativo.

Per Carnap, infine, una teoria si presenta come un linguaggio L, contenente, come suoi sottolinguaggi, il linguaggio osservativo L_O ed il linguaggio teorico L_T, oltre al sistema della logica, le regole della L-verità e le regole di corrispondenza. Trattandosi di un linguaggio empirico, il sistema della logica dovrebbe contenere, non solo tutta la logica deduttiva, ma anche tutta quella della probabilità logica, nonché i criteri per la verità fattuale[22].

[20] Cfr. C.G. Hempel, *Fundamentals of concept formation in empirical science*, Chicago 1952; trad. it. *La formazione dei concetti e delle teoria nella scienza empirica,* Feltrinelli, Milano 1976, pp. 46 sgg.

[21] Cfr. E. Nagel, *The Structure of Science*, Harcourt 1961; trad.it., *La struttura della scienza,* Feltrinelli, Milano 1968, p. 101.

[22] Cfr. R. Carnap, *The Methodological Character of Theoretical Concepts* e *Beobachtungsprache und theoretische Sprach»*; trad. it.: *Il carattere metodologico dei concetti teorici* e *Linguaggio osservativo e linguaggio teorico,* in R. Carnap,

Con ciò, Carnap viene implicitamente a negare che abbia senso parlare di probabilità di una teoria, esattamente come non ha senso parlare della verità, o della probabilità, di un linguaggio. La sua distinzione tra questioni *interne* e questioni *esterne* ha infatti relativizzato ogni conoscenza, di cui abbia senso parlare, ad un linguaggio, la cui adozione non può più essere ritenuta conoscitivamente giustificabile, ma solo 'praticamente' adeguata. Le teorie scientifiche, come linguaggi conoscitivi, rimanderebbero pertanto anch'esse a quelle decisioni che si collocano sul piano della prassi e che sono connesse ai fini ed ai contesti storici (nel senso più ampio) da cui sono animate. Di accettazione o di rifiuto di una teoria, quindi, sembra potersi parlare solo con riferimento ad un piano di 'adeguatezza' a fini 'pratici' prescelti e predefiniti, ma non in relazione ad un piano di pretesa sua oggettività o confermabilità empirica. Nessuna logica induttiva, o dell'incerto, sembra poter fornire criteri di scelta tra teorie rivali: questa logica, viceversa, appare necessariamente relegata alla soluzione di questioni induttive 'interne'.

3.7 Il caso clinico

Una situazione affatto particolare e tuttavia tipicamente bayesiana, anche se in pratica di difficile quantificazione (ma ciò non significa, a differenza dei casi discussi nei paragrafi precedenti, che si tratti di una situazione in linea di principio non determinabile o ambigua in un qualche senso), sembra configurarsi nel caso clinico, diagnostico e terapeutico. Nei casi tipici, la diagnosi è raggiunta per gradi, fase dopo fase, secondo un processo di ragionamento che si adatta bene al modello bayesiano. Il punto di partenza per il medico è dato da un insieme di informazioni consistenti nel quadro sintomatico offerto dal paziente, dall'anamnesi e da quanto altro derivato direttamente dalla visita oggettiva. Questo insieme di informazioni dovrebbe essere sufficiente ad orientare il medico ed a suggerirgli un insieme di ipotesi $\{H_i\}$ compatibili con questo quadro iniziale, ipotesi che egli già presumibilmente dispone in una sorta di graduatoria di valutazioni probabilistiche, guidato dalla propria competenza e dalla propria esperienza clinica, oltre che sulla base di ulteriori informazioni eventualmente disponibili (dati di diffusione delle patologie sospette, esistenza di forme epidemiche ecc.). In modo più o meno approssimato, comparativo e qualitativo, se non quantitativo, il clinico valuterà insomma le probabilità iniziali $p(H_i)$ relative ad ogni ipotesi diagnostica ragionevolmente possibile.

Analiticità, significanza, induzione, Il Mulino, Bologna 1971, pp. 263-315 e pp. 49-62; *Philosophical Foundations of Physics, New York 1966;* trad. it. *Fondamenti filosofici della fisica,* Il Saggiatore, Milano 1966; *Empiricism, semantics and ontology;* trad. it. *Empirismi, semantica, ontologia,* in *Neopositivismo,* a cura di A. Pasquinelli, UTET, Torino 1978, pp. 629-652.

La fase successiva consiste nella ricerca di nuovi dati (esami di laboratorio), in grado di trasformare le opinioni iniziali in un'alta probabilità finale a favore di una, tra le possibili cause in grado di spiegare i disturbi del paziente. Oltre alle analisi di routine, il medico richiederà quindi esami di laboratorio, radiologici ecc., mirati in tal senso, e proseguirà eventualmente per fasi successive, richiedendo man mano analisi più specifiche, finché la diagnosi non gli sembrerà sufficientemente sicura. Nel valutare i risultati degli esami di laboratorio, cioè nel valutare i *likelihoods* $p(r_j/H_i)$, il clinico avrà cura di tener conto non solo del grado di implicazione tra ciascuna delle ipotesi ed i risultati delle analisi, ma anche delle eventuali percentuali di falsi positivi e di falsi negativi, le cui statistiche sono generalmente note.

Una volta formulata la diagnosi, il processo bayesiano sarà nuovamente chiamato in causa, se occorre, in ausilio alla prescrizione terapeutica: può essere necessario, infatti, valutare la probabilità che il paziente appartenga a quella stessa popolazione di malati che ha mostrato di trarre giovamento da una determinata terapia.

Sotto il profilo logico, tanto la fase terapeutica quanto quella diagnostica si caratterizzano come un processo per adattamenti successivi, man mano che l'opinione iniziale, che altro non è che la formulazione di una strategia ipotetica, si trasforma sulla base dei risultati oggettivi riscontrati.

Una volta caratterizzato come tipicamente bayesiano il quadro logico entro cui si sviluppa il ragionamento clinico, e ferme restando le difficoltà oggettive per trasformare in valori di probabilità numerici tutti gli elementi che entrano nella valutazione, vorrei ancora soffermarmi su due questioni che mi sembrano di particolare rilievo.

1. *Insieme delle ipotesi diagnostiche.* Occorre a questo proposito ripetere per il caso specifico quanto già detto, poco sopra, per il caso generale. L'insieme delle ipotesi alternative deve essere un insieme esaustivo di ipotesi incompatibili *non fittizie*. In prima istanza, il clinico deve essere in grado di formulare un insieme di ipotesi, tali che ciascuna di esse potrebbe fornire una spiegazione accettabile della sindrome riscontrata nel paziente.

Non sembra pertanto una pratica accettabile quella di formulare un binomio alternativo di ipotesi h e $\neg h$, poiché, ammesso che h stia ad indicare una ben precisa patologia, $\neg h$ comprenderebbe invece, non solo tutte le patologie alternative possibili (cioè tali da essere anch'esse possibili cause dei disturbi in questione), ma anche quelle del tutto incompatibili con i sintomi rilevati e perfino lo stato di completa sanità; sarebbe equivalente, insomma, ad una disgiunzione inutilmente prolissa di tutti gli stati di salute possibili, il cui unico scopo sarebbe quello di permetter l'applicazione del calcolo delle probabilità. Oltre ad essere del tutto controintuitiva, $\neg h$ si rivelerebbe inoltre troppo comprensiva e finirebbe per riflettere la propria indeterminatezza sul *likelibood* $p(r_j/\neg h)$, rendendone la determinazione quantitativa impossibile o priva di senso.

Un razionale uso dell'approccio bayesiano, nel caso clinico, richiede invece che tutte le possibili cause dei disturbi del paziente siano chiaramente formulate: tra queste potrà eventualmente trovar posto anche l'ipotesi che i disturbi lamentati siano di origine funzionale o nervosa (che il paziente sia cioè organicamente sano), ove al medico tale possibilità sia suggerita da qualche dato discordante o da alcune caratteristiche del paziente. Il medico clinico, insomma, dovrebbe partire da una distribuzione iniziale di probabilità tra le cause ragionevolmente possibili del quadro sintomatico del paziente, per arrivare ad una distribuzione finale, conclusiva della fase diagnostica. Egli ha a che fare con un serio problema di valutazione e di decisione, nel quale impegna tutta la propria professionalità; non ha senso rappresentare questo problema come un futile giuoco di probabilità tra ipotesi fittizie. Piuttosto, laddove la situazione fosse tale da suggerire al medico una sola possibile causa del disturbo lamentato, dovremmo riconoscere che la logica qui impegnata cesserebbe di essere quella bayesiana, per adattarsi ad un tipo di ragionamento ipotetico-deduttivo, cioè:

$H \rightarrow S$, : il quadro sintomatico S fa pensare alla causa H,
H$\rightarrow(e_{1,\ldots,} e_n)$: deduzione da H dei risultati attesi dai test diagnostici di laboratorio
$e_1,\ldots e_n$: i risultati corrispondono alle attese

H è accettata come diagnosi ragionevole

Tale schema di ragionamento appare ricalcare abbastanza da vicino il modello di spiegazione probabilistica di Hempel, che traduciamo nel seguente modo:

$p(H/S)$ = prossima a 1
il quadro sintomatico s_i (del paziente) è un caso di S

(ciò rende altamente probabile che) s_i sia un caso di H

2. *Probabilità iniziali.* Molti bayesiani mostrano resistenze, peraltro ingiustificabili, all'uso di probabilità soggettive per le ipotesi diagnostiche e propongono di utilizzare per tali probabilità le frequenze di incidenza delle varie malattie nella popolazione.

Questa non sembra, in realtà, una strada percorribile, perché appare difficilmente compatibile con la fondamentale condizione che l'insieme delle ipotesi alternative considerate debba costituire una *partizione,* e che la somma delle $p(h_i)$ debba pertanto essere uguale all'unità.

Le frequenze di incidenza delle malattie nella popolazione, infatti, non sono, di norma, conformi a tale condizione; in generale le malattie non operano affatto una partizione reale della popolazione, come non è escluso che certe patologie ne facilitino altre: ad esempio, molti individui possono soffrire contemporaneamente, o aver sofferto successivamente, di svariate differenti malattie, anche consistenti con un medesimo quadro sintomatico. Di conseguenza, solo per un caso assolutamente fortuito e stravagante la somma delle frequenze di incidenza di un insieme di malattie, empiricamente rilevate nella popolazione e assunte come ipotesi esplicative del quadro sintomatico di un paziente, potrebbe essere uguale ad 1!

Si noti, per inciso, che ciò potrebbe costituire un argomento a favore dell'assunzione di una indeterminata ipotesi $\neg h$, quale unica alternativa ad una ipotesi diagnostica h determinata, perché, assumendo come probabilità a priori di h la sua frequenza di incidenza (f) nella popolazione, la probabilità a priori di $\neg h$ risulterebbe semplicemente determinata da $(1-f)$. Ciò solleverebbe, però, le questioni sopra indicate, e rischierebbe di portare a conclusioni inattendibili. Ove si rinunci invece al pregiudizio antisoggettivistico, il problema sembra porsi in modo più concreto e ragionevole.

Un medico riceve un paziente che presenta occhi gonfi, rossi, lacrimosi, naso gocciolante ecc. È raffreddore? È allergia? È ovvio che una persona soggetta ad allergia può ben prendersi anche un comune raffreddore, né le tavole di incidenza, se disponibili, consentono di attribuire valori *coerenti* di probabilità iniziale alle due ipotesi. Il clinico prenderà tuttavia in considerazione anche altri fattori: ad esempio, il fatto che si sia in pieno inverno oppure in primavera, e così via. In altre parole, il clinico terrà certamente conto *anche* degli eventuali dati relativi al grado di incidenza delle malattie, ma valuterà tali dati nel complesso di *tutte* le altre informazioni in suo possesso ed alla luce della propria personale esperienza, al fine di formulare valori soggettivi, ancorché non quantitativi, di probabilità.

In breve, il caso clinico non solo appare come tipicamente bayesiano, ma anche come tipicamente connesso con valutazioni soggettive di probabilità iniziale. In questa valutazione il medico clinico esprime la propria competenza, la propria esperienza, la propria professionalità. Pretendere di sottrarre tale processo alla valutazione soggettiva del medico equivale a disconoscere che «la clinica è, per definizione, valutazione individuale, giudizio sul caso singolo», nel quale «la diagnosi avviene fase dopo fase, a meno che non ci siano segni immediatamente discriminanti»[23].

[23] G. Prodi, *Teoria e metodo in biologia e medicina,* Bologna 1988, p. 54 s.

Capitolo 4
L’approccio oggettivista

4.1 Le critiche oggettivistiche alla regola di Bayes

L’approccio oggettivistico all’inferenza statistica trova origine nel clima della seconda metà dell’Ottocento, nel quale si è andata sviluppando una accesa polemica contro l’uso del teorema di Bayes e l’esigenza, successivamente raccolta da Fisher, di metodi alternativi non-bayesiani.

Ciò che la mentalità positivistica contestava, in nome del rigore e dell’oggettività scientifica, era il ricorso ad arbitrarie probabilità a priori nell’applicazione della formula di Bayes, il cui uso appariva giustificato soltanto se tali probabilità fossero ricavate dai dati oggettivi, empirici: i criteri di simmetria e di indifferenza, introdotti da Laplace, sembravano scandalosamente in contraddizione con l’oggettività perseguita dalla scienza.

In nome di tale oggettività, pertanto, l’inferenza inversa, mediante il teorema di Bayes, andava ripudiata: ogni inferenza induttiva legittima finiva per ridursi così quasi soltanto all’inferenza *diretta* (cioè da universo noto a campione da esso estratto), mentre la carenza di dati oggettivi, ben lungi dall’autorizzare assunzioni arbitrarie per la soluzione di un problema induttivo, non poteva che segnare il limite al di qua del quale la ricerca scientifica era costretta ad arrestarsi.

Così, per autori come Boole, Venn, Chrystal e altri non ci sono vie di mezzo: o i dati statistici oggettivi sono disponibili, ed allora siamo di fronte ad un problema di inferenza diretta, oppure non sono disponibili, ed allora non abbiamo il diritto di avanzare congetture.

G. Boole, nella sua opera citata[1], dichiarandosi contrario all’introduzione di costanti arbitrarie nella risoluzione di problemi induttivi,

[1] G. Boole, *An Investigation of the Laws of Thougth*, trad. it. cit. Le citazioni si riferiscono all’edizione italiana.

Maria Grazia Sandrini, *Filosofia dei metodi induttivi e logica della ricerca*, ISBN 978-88-6453-034-5 (print) ISBN 978-88-6453-037-6 (online) © 2009 Firenze University Press

riduce il campo di applicazione, non solo della formula di Bayes, ma dell'intera teoria delle probabilità. Ogni qualvolta le cause possibili di un dato fenomeno e le loro probabilità non siano note sulla base di dati oggettivi, Boole ritiene impossibile fornire una qualche soluzione definita ad un problema induttivo. In particolare, le probabilità devono poter essere desunte dai *dati,* i quali possono provenire soltanto da due fonti: possono esserci dati *causalmente,* nel senso che dalla conoscenza della struttura di un oggetto, o di una situazione, derivano deduttivamente i valori di probabilità che ci interessano. Così è ad esempio nel giuoco dei dadi, nel quale «dalla conoscenza della costituzione del pezzo si deduce la probabilità che gettando un dado venga una determinata faccia»: in questo caso, la *misura* della probabilità è il quoziente tra casi favorevoli e totalità dei casi, «purché tutti ugualmente possibili»[2]. Così è anche, naturalmente, in tutti i casi di inferenza diretta, nei quali dalla conoscenza delle caratteristiche dell'intera popolazione si vogliono trarre inferenze circa individui estratti a caso da essa.

Oppure, i dati possono essere desunti dall'esperienza, cioè «derivati dall'osservazione della ripetizione dei casi in cui gli eventi accadono o non accadono». In tal caso, alla probabilità è assegnato come valore «il limite a cui tende il rapporto tra i casi favorevoli e la totalità dei casi osservati (ammettendo l'uniformità delle leggi di natura) quando le osservazioni vengano proseguite indefinitamente»[3] .

L'inferenza induttiva è così di fatto sempre e soltanto inferenza diretta ed al calcolo delle probabilità è attribuito un compito strettamente deduttivo, mentre l'inferenza inversa è ridotta al calcolo della mutazione del valore noto di probabilità di una ipotesi dopo il verificarsi di nuovi eventi. Diversamente, il problema inverso ammette, per Boole, «solo una soluzione indefinita». Poniamo infatti, egli dice, che x rappresenti l'ipotesi che ci interessa, y un fenomeno che potrebbe accadere come possibile conseguenza di tale ipotesi e la cui probabilità subordinatamente ad x (cioè il *likelihood*), è p. Se ci chiediamo ora quale sia la probabilità dell'ipotesi x, nel caso che si osservi il fenomeno y, «già gli stessi dati del problema non possono essere espressi senza introdurre un elemento arbitrario. Possiamo solo scrivere: $Prob(x)=a; Prob(x.y)=ap$; dove a è perfettamente arbitrario»[4].

La principale conclusione che può dunque essere tratta dal teorema di Bayes consiste per Boole nelle relazioni matematiche che esso indica, e, in particolare, nella stretta proporzionalità tra il valore *a posteriori* della probabilità dell'ipotesi ed il valore del *likelihood*. Perciò, «quanto

[2] Ivi, p. 27.
[3] Ivi.
[4] Ivi, pp. 499 sgg.

maggiore (o minore) è la probabilità del fenomeno quando *l'ipotesi* è stata *assunta,* tanto maggiore (o minore) sarà la probabilità dell'ipotesi quando il *fenomeno* è stato osservato»[5].

La classe di problemi cui Laplace ed altri avevano dedicato particolare attenzione (Boole fa menzione del problema delle stelle doppie, affrontato da Mitchell, e di quello delle inclinazioni delle orbite dei pianeti, discusso da Laplace) rientra nei casi non suscettibili di una soluzione definita, a meno che non si ricorra ad assunzioni arbitrarie, per Boole non giustificabili. A questo proposito, anche il ricorso al principio di indifferenza, sostenuto da Laplace, è considerato da Boole un procedimento arbitrario e pertanto da rifiutare[6].

In conclusione, l'introduzione di costanti arbitrarie per la soluzione di problemi induttivi, in mancanza di dati oggettivi, sembra a Boole indizio sufficiente ad implicare l'impossibilità di una soluzione definita e «segnare il punto in cui la ricerca dovrebbe arrestarsi»[7].

Per certi versi ancor più rigida appare la presa di posizione di Venn[8], il quale non concepisce la probabilità se non come descrizione di un fatto empirico oggettivo: essa si identifica cioè con la frequenza osservata e rilevata nelle statistiche. È in quest'ottica empiristica e oggettivistica che egli polemizza contro coloro che definiscono la probabilità come grado di credenza, in particolare contro De Morgan[9]; sempre in quest'ottica si colloca la sua posizione nei confronti dell'inferenza inversa, che è rifiutata, in quanto suscettibile solo di soluzioni arbitrarie.

Più precisamente, Venn sostiene che debba essere abbandonata la distinzione, introdotta tra gli altri da De Morgan, tra «probabilità diretta» e «probabilità inversa». A suo parere, infatti, quando siano disponibili

[5] Ivi, p. 501.

[6] Cfr. ivi, p. 506. Al proposito, K. Pearson si chiede se, dopo tutto, anche al principio di indifferenza non possa essere riconosciuta una base nell'esperienza passata, che avrebbe mostrato come, nei casi di completa ignoranza, tutte le ipotesi si riveleranno, a lungo andare, equiprobabili. A sostegno di ciò, Pearson cita un esempio di F.Y. Edgeworth (*The Philosophy of Chance,* «Mind», vol. IX, 1884): supponiamo di dividere il numero 143678 per 7 e di terminare l'operazione dopo le prime quattro cifre. Abbiamo così ottenuto il numero 2052, ma non sappiamo quale numero seguirebbe ancora. Ebbene, in tale situazione di ignoranza, possiamo supporre che tutti i numeri da 0 a 9 abbiano la stessa probabilità di essere la prossima cifra, perché, se dividessimo una gran quantità di numeri di sei cifre per 7, troveremmo che tutti i numeri da 0 a 9 occorrerebbero come quinta cifra del quoziente in un numero di volte pressoché uguale. In altre parole, le statistiche giustificherebbero «l'uguale distribuzione della nostra ignoranza». Ciò, conclude Pearson, «può, forse, essere sufficiente a mostrarci che c'è un elemento di esperienza umana alla base dell'assunzione di Laplace» (*The Grammar of Science,* cit., pp. 174 sgg.).

[7] G. Boole, *An Investigation of the Laws of Thought,* cit., p. 503.

[8] J. Venn , *The Logic of Chance,* cit.

[9] A. De Morgan (1806-1871), *Formal Logic,* London 1847.

dati empirici adeguati, tale distinzione perde significato, dal momento che i due tipi di inferenza esibiscono una identità di metodo; mentre, laddove i dati empirici siano insufficienti, non abbiamo alcun diritto di avanzare conclusioni[10] .

A chiarimento, Venn analizza i due seguenti esempi:

1. una palla è estratta da un'urna contenente nove palle nere e una bianca; si chiede quale sia la probabilità che essa sia bianca;
2. una palla bianca è estratta da un'urna contenente dieci palle; si chiede quale sia la probabilità che nell'urna non vi sia che una sola palla bianca.

Il primo esempio (di inferenza diretta) è interpretabile, a suo parere, nel modo seguente, in accordo alla concezione frequentistica delle probabilità: se delle palle venissero continuamente estratte dall'urna, *con ripetizione* (cioè, reimbussolando ogni volta la palla estratta), la proporzione di quelle bianche, sull'intero numero di palle estratte, tenderebbe verso la frazione 1/10. Sebbene l'esempio faccia riferimento ad un evento singolo, questo può tuttavia essere considerato come appartenente ad una serie, nella quale gli eventi di un certo tipo tenderanno a verificarsi, a lungo andare, con la suddetta frequenza relativa. Dicendo allora che la probabilità di estrarre una palla bianca è uguale ad 1/10, intendiamo asserire *questo fatto empirico*.

Il secondo esempio (di inferenza inversa) sembra a Venn da interpretarsi invece nel seguente modo: palle bianche sono continuamente estratte da urne differenti, ognuna contenente dieci palle; quale sarà alla fine la proporzione dei casi nei quali le palle risulteranno estratte da urne contenenti una sola palla bianca?

Nei due esempi abbiamo a che fare, dunque, rispettivamente, con palle di colori differenti in una proporzione nota, e con urne con contenuti differenti, al fine di assegnare una misura alla probabilità di: a) estrarre una palla bianca; b) estrarre un'urna contenente una sola palla bianca.

Mentre nella soluzione del primo esempio non compare nulla di arbitrario, nel secondo caso la situazione appare a Venn assai differente. Generalmente, nell'affrontarlo, si procede nel modo precedentemente indicato: poiché l'estrazione avvenuta assicura la presenza di almeno una palla bianca, sono formulate dieci ipotesi compatibili con questa estrazione (cioè da una a dieci palle bianche) e ad ogni ipotesi viene assegnata la frazione 1/10 quale valore di probabilità a priori. Questa è appunto, per Venn, un'assunzione del tutto arbitraria, non giustificata da alcuna esperienza. Cosa ci autorizza ad assumere che ciascuna delle urne ipotizzate

[10] Cfr. J. Venn, *The Logic of Chance*, cit., cap. VII, sez. 9, pp. 179-189.

verrebbe estratta con la stessa frequenza delle altre? Eccettuata questa assunzione, tuttavia, l'intero procedimento restante sembra a Venn identico in ambedue gli esempi, dal momento che esso non richiede che un uso 'diretto' del calcolo delle probabilità[11]; l'assunzione di equiprobabilità a priori delle ipotesi è viceversa per Venn così ingiustificabile da rendere inammissibile l'uso della formula di Bayes.

Del resto, aggiunge Venn, anche concedendo di dare dei valori a priori alle possibili composizioni dell'urna, non è detto che l'equiprobabilità sia l'assunzione più ragionevole: Whitworth[12], ad esempio, assume come equiprobabili, non le possibili proporzioni tra palle bianche e nere nell'urna, ma tutte le combinazioni individuali di palline. Ne consegue che risultano più probabili quelle urne composte da un numero uguale, o quasi uguale, di palle bianche e nere (ad esempio, cinque e cinque)[13].

Poiché non è facile dire perché la distribuzione di probabilità generalmente adottata dovrebbe essere ritenuta più ragionevole di quella di Withworth, sembra a Venn che non vi siano alternative: o le statistiche tratte dai dati empirici sono disponibili, ed allora viene meno ogni differenza tra probabilità diretta e inversa (risultando ogni problema di inferenza sempre un problema di inferenza diretta), oppure esse non sono disponibili, ed allora non abbiamo in nessun caso il diritto di trarre una qualsiasi inferenza!

In conclusione, la teoria delle probabilità ha per oggetto, secondo Venn, soltanto le inferenze da proposizioni generali del tipo «alcuni A sono B» ad asserzioni particolari relative ad individui o gruppi di individui appartenenti ad A.

Se la probabilità non serve che a fare inferenze dirette, probabilistico-deduttive, è lecito chiedersi quale relazione con l'induzione Venn le attribuisca.

La ricerca scientifica, modellata sul metodo baconiano, parte, per Venn, dall'osservazione, per procedere quindi alla classificazione ed alla formulazione di asserti generali dai quali verranno, successivamente, tratte

[11] Assunte le ipotesi, il restante procedimento riguarda semplicemente il calcolo del valore dei *likelihoods*.

[12] W.A. Whitworth, *Choice and Chance,* 2° ed., p. 123; citato in J. Venn, *The Logic of Chance*, cit., p. 183.

[13] Infatti ci sono 252 modi per ottenere tale proporzione; mentre ve ne sono 210 per la proporzione 6 a 4; 120 per quella 7 a 3; 45 per quella 8 a 2 e soltanto 10 per la proporzione 9 a 1. Whitworth distribuisce cioè l'equiprobabilità tra quelle combinazioni individuali che Carnap chiama *descrizioni di stato*, anziché alle distribuzioni statistiche che corrispondono alle *descrizioni di struttura* di Carnap: queste seconde indicano soltanto la proporzione delle palline nell'urna, mentre le prime tengono conto delle singole palline, come se queste fossero riconoscibili individualmente (ad esempio perché indicate con una lettera). Anche la statistica di Boltzmann è basata sulle combinazioni individuali.

inferenze, deduttive o probabilistiche. Lo scienziato osserva la moltitudine dei fenomeni oggettivamente, senza pregiudizi aprioristici: il suo compito dovrebbe inizialmente «consistere in un graduale accumulo di istanze individuali, distinte una dall'altra secondo vari gradi di differenza e connesse secondo gradi di somiglianza». Successivamente, egli dovrebbe esprimere i risultati «in proposizioni generali dalle quali possano in seguito essere tratte inferenze. Queste inferenze non dovrebbero, naturalmente, contenere alcun fatto nuovo, ma dovrebbero solo essere ripetizioni di ciò che egli o altri hanno precedentemente osservato»[14]. L'esperienza stessa, rivelando come *dato di fatto* l'esistenza di leggi di natura, offre la possibilità di estendere il campo delle inferenze da pochi fatti ad una intera classe di oggetti, permettendo così di conseguire in breve tempo risultati che altrimenti avrebbero richiesto una lunga e faticosa ricerca.

Il processo conoscitivo si divide dunque, per Venn, in due parti distinte: nella prima sono stabilite, in un qualche modo imprecisabile, delle generalizzazioni; nella seconda sono determinate quali inferenze possano esserne tratte. Ebbene, per Venn l'induzione concerne esclusivamente la prima parte di questo processo. La seconda, invece, è affidata alla logica deduttiva o alla probabilità: laddove le generalizzazioni siano espresse in forma di proposizioni universali, sarà la logica deduttiva a determinare quali inferenze se ne possano trarre; laddove invece le generalizzazioni siano espresse nella forma di proposizioni statistiche, le inferenze saranno dominio della probabilità. Il calcolo delle probabilità deve limitarsi ad inferire da proposizioni generali statistiche, relative ad una determinata popolazione, caratteristiche riguardanti individui tratti da quella popolazione medesima: la probabilità, del resto, non è altro che una descrizione di fatti empirici, cioè la frequenza osservata e rilevata nelle statistiche, e non ha funzioni induttive. Per *induzione* determiniamo, ad esempio, che quattro uomini su dieci vivono oltre i 65 anni di età; ciò posto tocca alla probabilità trarne le inferenze possibili.

È così riconosciuto alla probabilità un ruolo del tutto analogo e complementare a quello della logica deduttiva, mentre il processo di induzione viene considerato precedente ed indipendente da essa. Senonché, le inferenze tratte da queste generalizzazioni, e le generalizzazioni stesse, nel momento in cui passano da mere descrizioni del 'già osservato' a leggi esprimenti regolarità naturali, travalicano la portata del piano empirico per proiettarsi, ingiustificatamente, verso il futuro e il non-noto. In ciò risiede il loro valore per la scienza e la loro utilità pratica; ma in ciò consiste, come Hume aveva mostrato, il punto delicato e discusso del 'problema dell'induzione'. Ma questa questione non sembra sfiorare Venn, per il quale anche l'assunto dell'uniformità della natura è una generalizzazione

[14] J. Venn, *The Logic of Chance,* cit., p. 205.

ottenuta per induzione, cioè in base al successo delle inferenze induttive passate[15]. Inutile ricordare ancora una volta, con Hume, che ciò involve un regresso all'infinito, perché l'uniformità della natura è il presupposto stesso dell'induzione, e che pertanto il tentativo di fondare l'induzione sull'esperienza, al fine di garantirne l'oggettività e la certezza, non sembra suscettibile, almeno per questa via, di un esito positivo!

Tra i detrattori dell'inferenza inversa e del teorema di Bayes, che si pongono sulla scia di Boole e di Venn, va annoverato, poco meno di un secolo più tardi, anche R.A. Fisher, il fondatore della statistica moderna, al quale dedicheremo un ben più lungo discorso.

4.2 K. Pearson e la teoria della correlazione

Tutte le generalizzazioni empiriche esprimono *correlazioni*, vale a dire connessioni parziali o totali tra due o più classi, o caratteristiche. L'enunciato «tutti i corvi sono neri» esprime una stretta correlazione tra la classe «corvo» e la classe «nero», mentre l'enunciato «tutti i giapponesi sono bassi ed hanno gli occhi a mandorla» esprime una correlazione tra tre caratteristiche, o classi: quella di un certo gruppo umano, quella dell'altezza e quella della forma degli occhi. I due enunciati precedenti esprimono connessioni universali, poiché la parola «tutti» non ammette eccezioni. Correlazioni più elastiche sono espresse da enunciati del tipo «il 90% dei malati di tumore al polmone è costituito da fumatori». Il termine *correlazione*, in breve, non è che l'equivalente statistico della connessione costante tra differenti caratteristiche o classi di oggetti che ogni enunciato generale empirico, se significativo, esprime.

Se volgiamo uno sguardo alla storia della scienza, ci accorgiamo come il reperimento di correlazioni abbia sempre avuto grande rilevanza nella ricerca. L'esperimento di Oersted, ad esempio, ha permesso di mettere in luce l'esistenza di una correlazione tra fenomeni elettrici e fenomeni magnetici: una correlazione non integrabile nella fisica newtoniana e che richiese, per essere spiegata, l'introduzione di un nuovo quadro teorico. Si tratta, in questo caso, di un esperimento mirato e guidato da precise ipotesi di lavoro; ma il rilevamento di correlazioni può aver luogo anche non intenzionalmente, come nel caso di un esperimento volto a finalità del tutto differenti, oppure accidentalmente, come nel caso già ricordato del dottor Semmelweis e della febbre da parto. Nel settore della ricerca epidemiologica, sono proprio le correlazioni l'oggetto stesso della ricerca: gli studi sull'associazione tra alimentazione e tumori, ad esempio, hanno messo in luce una correlazione significativa tra certi alimenti ricchi di vitamina C ed una ridotta incidenza di neoplasie.

[15] Cfr. J. Venn, *The Logic of Chance,* cit., p. 193.

Una correlazione può, naturalmente, sottintendere nessi di differente forza o valore: una pura generalizzazione empirica, un nesso causale, una connessione puramente accidentale. Vedo un lampo e sento il tuono: questa esperienza ripetuta fa sì che mi aspetti di sentir tuonare ogni volta che vedo un lampo. Ma non tutte le esperienze associate insieme, anche se ripetute nel tempo, possono essere prese come indicative di legami stabili ed utili per previsioni future. Suonano alla porta mentre sono sotto la doccia; squilla il telefono mentre sto per uscire precipitosamente, in ritardo per un appuntamento: per quanto ripetute nell'esperienza di ognuno, queste circostanze non possono esser considerate che spiacevoli coincidenze, a meno di non voler ammettere una qualche sorta di maligna forza telepatica che ci perseguita.

La correlazione, come connessione costante di eventi, costituisce la forma più elementare di legge empirica, tanto che possiamo idealmente immaginare una sorta di infanzia della conoscenza, costituita semplicemente da un *corpus* di asserti su uniformità universali esprimenti correlazioni osservate e generalizzate. Non è difficile rendersi conto tuttavia che una scienza basata su uniformità semplicemente 'date' finisce per ammettere tutto ed il contrario di tutto: come osserva J.S. Mill, le stelle fisse possono essere messe in correlazione con tutti i fenomeni occorrenti nel mondo, essendo ad ognuno coesistenti[16].

Eppure, nella concezione positivistica ed empiristica, la correlazione costituisce il fine della ricerca scientifica, perché alla scienza è affidato un mero compito descrittivo e, subordinatamente, previsivo. Hume, giova ricordarlo, aveva sottoposto a critica il concetto di «causa» riducendolo a nient'altro che ad una connessione costante di eventi, dei quali l'antecedente è detto «causa» ed il conseguente «effetto».

Circa un secolo più tardi, Mill riprende tale definizione; ma, dotato di fine sensibilità metodologica, mostra di trovarsi in difficoltà. I metodi di Mill, che possono essere correttamente caratterizzati come metodi qualitativi di *scoperta* di correlazioni di forma universale, si differenziano, per la loro capacità selettiva, nella misura in cui possano essere abbinati ad esperimenti volontari, e non alla semplice osservazione di eventi «dati»: qui Mill, oltrepassando i limiti imposti dall'empirismo, esprime l'esigenza di una distinzione tra uniformità del tipo 'notte-giorno', che pur essendo invariabile non esprime una effettiva intima relazione tra i due eventi (né la notte è causa del giorno, né questo di quella!), e correlazioni che possano a buon diritto essere chiamate causali[17]. Le difficoltà avvertite da

[16] J.S. Mill, *System of Logic,* cit., p. 519.

[17] Definita la causalità come successione invariabile di fatti e la causa come l'*antecedente invariabile*, Mill introduce poco dopo un nuovo concetto per definire la causalità, cioè l'*antecedente invariabile incondizionale*, per operare una distinzione tra i vari tipi di uniformità. Cfr. J.S. Mill, *System of Logic,* cit., pp. 322 e 334 sgg.

Mill rivestono particolare interesse perché anticipano quelle difficoltà che anche la teoria statistica della correlazione si troverà a dover affrontare, con l'ulteriore complessità derivante dalle leggi statistiche che si sono andate affermando, nel frattempo, nella scienza (si pensi, ad esempio, alle leggi di Mendel).

La teoria della correlazione, poggiando sulla più semplice e istintiva forma di induzione – quella per associazione ripetuta – intende venirci in aiuto per mettere ordine nel caos delle esperienze temporalmente o spazialmente connesse, isolando quelle troppo 'rare' per poter servire come fonte di apprendimento per previsioni future. In questo contesto, l'accento è posto, non tanto sull'esclusione degli eventi accidentali, quanto sull'importanza dell'individuazione delle correlazioni *significative*: sono queste il vero oggetto della ricerca scientifica, che ci forniscono l'indicazione di quali uniformità è lecito attendersi dalla natura e sulle quali poter contare.

Enunciata da Galton nel 1885, la «teoria della correlazione» fu matematicamente elaborata da Karl Pearson a partire dal 1893. Nel 1892 Pearson aveva però pubblicato la sua opera epistemologica, *Tbe Grammar of Science*[18], ampio studio che, pur trovando le sue radici nell'empirismo di Hume e nel metodo baconiano, sviluppa strette convergenze con il sensismo di Mach[19]; non mancano tuttavia neppure echi kantiani, in una prospettiva che, pur essendo certamente testimonianza dello spirito scientifico del suo tempo, non è priva di una propria originalità, sia pur con istanze talora difficilmente conciliabili fra loro. Poiché quest'opera permette di mettere a fuoco lo sfondo epistemologico, anche se spesso confuso e contraddittorio, sul quale vanno collocati importanti sviluppi della metodologia statistica, tra i quali, *in primis*, la teoria della correlazione, ci soffermeremo ad esporne gli aspetti più rilevanti, non senza, qua e là, qualche rilievo critico.

La funzione della scienza è caratterizzata da Pearson come *descrizione* di eventi osservati e *previsione* di eventi futuri. Fin dalle prime pagine della sua opera, egli dà voce alla propria concezione positivistica: «la classificazione dei fatti e la formazione di giudizi assoluti sulla base di questa classificazione – giudizi indipendenti dalle idiosincrasie della mente individuale – è peculiarmente lo scopo e il metodo della scienza moderna»[20].

La conoscenza trae origine, per Pearson come per Hume, dalle impressioni sensoriali immediate, immagazzinate poi nella memoria. Al di fuori di

[18] K. Pearson (1857-1936), *The Grammar of Science,* cit.

[19] Cfr. E. Mach, *DieAnalyse der Empfindungen und das Verhältnis des Physischen zum Psychischen,* 1896; trad. it. *L'analisi delle sensazioni e il rapporto tra fisico e psichico,* Feltrinelli, Milano 1975.

[20] K. Pearson, *The Grammar of Science,* cit., p. 7.

tali impressioni, ovviamente soggettive, non vi è che il mondo dei costrutti concettuali, elaborati dal pensiero sulla base delle impressioni stesse. Se sottoponiamo ad analisi ciò che comunemente chiamiamo «oggetto esterno», troviamo sempre e soltanto determinate impressioni sensoriali immediate, sulla cui base, mediante associazioni con altre impressioni sensoriali memorizzate, è possibile *inferire* ulteriori impressioni sensoriali normalmente associate ad un certo oggetto (per esempio, l'impressione sensoriale visiva di un fuoco lontano e l'inferenza del calore percepibile ad una distanza più ravvicinata). Il risultato di tutto ciò è un *costrutto*.

L'oggetto esterno non è dunque altro che una elaborazione concettuale che riassume e codifica un numero più o meno grande di impressioni sensoriali immediate, associate ad un gruppo di impressioni memorizzate. La realtà di un oggetto consiste semplicemente, perciò, nella sua possibilità di occorrere come gruppo di impressioni sensoriali immediate, e, poiché l'individualità di un oggetto è riducibile alla somiglianza tra impressioni sensoriali in due differenti istanti di tempo, non è necessario postulare la cosa in sé.

Il mondo esterno, nel quale i fatti avvengano indipendentemente da noi, come è comunemente supposto e come pretende il realismo, è semplicemente per Pearson una illusione metafisica. Anzi, essendo la conoscenza costruita a partire dalle impressioni sensoriali, il 'mondo esterno' è condizionato e limitato dalle nostre facoltà percettive e nervose; la 'realtà' è costruita sulla base delle impressioni sensoriali soggettive, e soltanto la somiglianza degli organi sensoriali e del sistema nervoso di tutti gli esseri umani 'normali' fa sì che si possa parlare di un mondo esterno praticamente uguale per tutti gli uomini: «due normali facoltà percettive costruiscono praticamente lo stesso universo». Su questo, e soltanto su questo, riposa la scienza: «la validità universale della scienza dipende dalla similarità delle facoltà percettive e di ragionamento di normali uomini civilizzati»[21].

Tale essendo per Pearson la sorgente ed il materiale della conoscenza, anche i fatti della scienza si caratterizzano come gruppi di impressioni sensoriali immediate connessi con altri gruppi di impressioni sensoriali memorizzate, il tutto organizzato secondo un costrutto concettuale che procede attraverso una lunga catena di inferenze: in pratica si procede da impressioni dirette ed immediate ad elaborazioni concettuali non più suscettibili di verifica diretta, per le quali cioè non sussiste un referente percettivo diretto.

Tali fatti, e le loro relazioni, possono poi essere riassunti e descritti mediante una *formula* che, in una stenografia mentale atta a realizzare una economia di pensiero, li riassuma e li descriva: una legge scientifica è

[21] Ivi, p. 57.

tale, per Pearson, quando alla capacità descrittiva di sequenze percettive unisce un'adeguata originalità concettuale.

Una legge scientifica, dunque, non esprime un ordine naturale indipendente nel senso del realismo, ma possiede significato soltanto in relazione alle facoltà percettive (se l'uomo fosse dotato di facoltà percettive differenti, anche il mondo sarebbe diverso) e razionali dell'uomo (una legge esiste solo nella misura in cui viene formulata dalla mente umana, la quale introduce l'elemento razionale nella natura)[22]. Generalmente, con «legge scientifica» si intende qualcosa di universalmente valido; in effetti, vi è, per Pearson, qualcosa di peculiare nella legge scientifica, che la differenzia, ad esempio, da un articolo di fede: una legge scientifica «sarà accettata da ogni mente razionale, una volta che questa ne abbia compreso i termini ed analizzato chiaramente i fatti che la legge riassume»[23].

Al proposito, Pearson solleva la seguente questione: ammesso che la natura sia condizionata dalle facoltà percettive e razionali dell'uomo, le sequenze sensoriali si susseguirebbero nello stesso modo, sia che l'uomo abbia già formulata una legge sia che ancora non l'abbia formulata? In altre parole, i fatti naturali si svolgevano in accordo alla legge di gravitazione universale ancor prima che Newton la scoprisse? La domanda è chiaramente tendenziosa, oltre che ambigua. Cosa significa, infatti, chiedersi se le sequenze sensoriali si susseguirebbero nello stesso modo, avendo o non avendo un determinato codice interpretativo? Le 'sequenze sensoriali', in quanto tali, *potrebbero* anche susseguirsi in un medesimo modo, ma ciò che sembra avere importanza, ai fini conoscitivi, non sono le sequenze in se stesse, bensì la coscienza soggettiva delle medesime, cioè il significato che attribuiamo loro, e ciò sembra strettamente dipendente dal contesto interpretativo adottato.

Ad ogni modo, la risposta di Pearson non è né un sì né un no, ma una via di mezzo, in cui sembra tuttavia privilegiata la funzione della ragione. La sequenza percettiva dei fenomeni, egli dice, era sicuramente la stessa anche prima che Newton formulasse la sua legge; tuttavia, «una sequenza di impressioni sensoriali non è per se stessa una legge». Affinché si possa parlare di legge,

> [...] la sequenza delle percezioni deve essere confrontata con altre sequenze, deve seguirne una classificazione e una generalizzazione; devono essere formati concetti e idee, puri prodotti della mente, prima che possa essere data, in relazione ad un campo di sequenze percetti-

[22] Qui appaiono chiaramente le due differenti istanze che Pearson cerca di conciliare: un empirismo che anticipa il sensismo di Mach, da un lato, ed una sorta di kantismo, dall'altro, che trova varie assonanze con la concezione di Whewell, come anche il termine «formula» suggerisce.

[23] K. Pearson, *The Grammar of Science,* cit., p. 94.

ve, una descrizione che, per la sua concisione e per la sua portata, sia meritevole del nome di legge scientifica[24].

Osserviamo, per inciso, che il rifiuto di ogni forma di realismo comporta, per Pearson, una prima e non lieve difficoltà in relazione ai criteri di validità che ogni concetto scientifico dovrebbe soddisfare. Pearson ne indica due: un criterio di autoconsistenza, che altro non è che il criterio empiristico di razionalità (un'idea è da ritenersi contraddittoria quando non si collega a possibili impressioni sensoriali: il centauro è, sotto questo profilo, un concetto non consistente), ed un criterio che potremmo definire di intersoggettività, il quale fa appello alle impressioni sensoriali di ogni individuo 'normale' (con ciò si vogliono escludere impressioni sensoriali troppo individuali come ad esempio quelle di un daltonico). Ambedue i criteri appaiono discutibili. Il primo appare eccessivamente limitativo, poiché una scienza che davvero vi si adeguasse non potrebbe che riproporre sempre nuove tautologie del già esperito, restandole addirittura precluso il diritto di 'pensare' campi di esperienza sinora inediti. Che il dato debba essere criterio di pensabilità razionale in sede scientifica è uno dei più grossi limiti che l'empirismo, ivi compreso il recente neopositivismo, ha voluto imporre alla scienza. Il secondo criterio, sembra troppo debole e vago: quale sarebbe, a sua volta, il criterio di 'normalità' e come effettuarne il controllo? Oppure, dovrebbe risolversi sul piano delle convenzioni linguistiche?

Riprendendo la nostra esposizione, una legge scientifica è dunque, per Pearson, una sintesi descrittiva di sequenze percettive, ma tale 'sintesi' non deve limitarsi ad esprimere, a suo parere, una semplice concatenazione di fenomeni: perché si possa parlare di 'legge' è indispensabile che vi sia produzione di concetti, una generalizzazione creativa: la 'ragione' che troviamo nei fenomeni, e che molti intendono in senso realistico, non è altro che la ragione umana.

Pearson cerca un difficile equilibrio tra le sensazioni di Hume e la funzione creativa dell'intelletto, di ispirazione kantiana. Perciò egli insiste, non so quanto congruamente, tanto sulla dimensione creativa di una legge scientifica, – pur restando il controllo da parte dei fatti «la sola prova che il nostro intelletto sia stato abbastanza acuto per giungere ad una formula coprente l'intero campo di fatti che essa professa di riassumere»[25] –, quanto sulla richiesta che la ricerca scientifica debba sempre rigorosamente iniziare con una mera raccolta di fatti, in accordo con i canoni metodologici indicati da Bacone nel *Novum Organum*. Esemplare, sotto questo profilo, gli appare il processo seguito da Darwin per giungere alla

[24] Ivi, p. 103.
[25] Ivi, p. 120.

formulazione della sua legge di evoluzione naturale, come è da lui stesso esposto: «raccogliendo tutti i fatti che si riferivano in qualche modo a mutazioni di animali e piante, [...] ho lavorato secondo i veri principi baconiani e, senza alcuna teoria, ho raccolto fatti»[26].

Jevons, nei suoi *Principles of Science,* aveva duramente attaccato il metodo induttivo baconiano, schierandosi a favore del metodo ipotetico-deduttivo di origine newtoniana. Il metodo scientifico proposto da Bacone appare a Jevons come una sorta di contabilità: i fatti dovrebbero, secondo tale metodo, essere indiscriminatamente raccolti da ogni parte ed elencati in un libro-mastro dal quale, col tempo, emergerebbe un bilancio di verità. «È difficile immaginare un metodo meno verosimile per giungere a grandi scoperte», egli dice; ed aggiunge: «centinaia di ricercatori possono essere costantemente impiegati nella ricerca sperimentale; essi possono compilare innumerevoli resoconti di fatti scientifici e infinite tavole di risultati numerici; ma, se la concezione dell'induzione qui sostenuta è vera, essi non potranno mai, attraverso questo unico lavoro, pervenire a nuove e grandi scoperte»[27]. Il metodo scientifico è assai meglio rappresentato, per Jevons, dai procedimenti di Keplero, di Faraday, di Newton, procedimenti nei quali i successi si alternano agli insuccessi, le speculazioni valide a quelle vane e senza fondamento; procedimenti, in breve, guidati dall'attività del pensiero che interviene a suggerire possibili nuovi significati dei fatti e ardite connessioni. In quest'ottica metodologica, i fatti non dicono nulla da soli, ma sono provocati, selezionati e vagliati in connessione con una ipotesi di lavoro preliminare, atta ad interpretarli. Tuttavia, pur lasciando ampio spazio ad un'ardita attività del pensiero, questo metodo non è privo, per Jevons, di rigore scientifico; nessuna ipotesi è infatti ammessa se non è definita esattamente nelle sue condizioni e interamente sviluppata nelle sue conseguenze deduttive; il confronto con i fatti sperimentali deciderà poi, in modo definitivo, del valore di ogni ipotesi.

La difesa del metodo baconiano è viceversa netta da parte di Pearson, ma non senza affermazioni contraddittorie, che talora lasciano perplessi. Un posto eminente, infatti, deve comunque essere riconosciuto anche all'immaginazione: «tutti gli scienziati sono stati in un certo senso grandi artisti; l'uomo sprovvisto di immaginazione può raccogliere fatti, ma non può fare grandi scoperte». Tuttavia, «l'immaginazione non deve sostituire la ragione nel ricavare relazioni e leggi dai dati classificati»[28]. Il primo passo della ricerca deve quindi consistere nella raccolta di fatti puri, seguita dalla loro classificazione, dall'analisi delle loro relazioni, dal confronto tra sequenze diverse, e così via. Una volta compiute queste operazioni,

[26] *The Life and Letters of Charles Darwin*, vol. I, p. 83, cit. ivi . p. 39.
[27] W.S. Jevons, *The Principles of Science*, cit., p. 577.
[28] K. Pearson, *The Grammar of Science,* cit., p. 37.

l'immaginazione «disciplinata», creativa, interviene nella scoperta di una *formula* atta a sintetizzare e descrivere l'intero gruppo di fatti.

La legge è così, per Pearson, una *descrizione*, per quanto «ragionata», di sequenze percettive, di correlazioni; essa non spiega *perché* certi fenomeni si presentino con un certo ordine né perché quell'ordine si ripeta. La legge, in breve, non introduce nessun elemento di necessità nella sequenza delle nostre impressioni sensoriali. Causa ed effetto non sono che nomi per indicare connessioni costanti tra fenomeni: «ovunque una sequenza di percezioni D, E, F, G sia invariabilmente preceduta dalla percezione C, oppure le percezioni C, D, E, F, G si diano sempre in quest'ordine, formino cioè una *routine* nell'esperienza, C è detta essere la causa di D, E, F, G, che sono descritti come suoi effetti»[29]. Al di là di questa *routine* percettiva non vi è altro senso della causalità: «che una certa sequenza sia occorsa e ricorsa nel passato è una questione di esperienza che esprimiamo nel concetto di *causalità*; che essa continuerà a ricorrere nel futuro è una questione di *credenza* che esprimiamo nel concetto di probabilità. La scienza in nessun caso può dimostrare una inerente necessità di una sequenza, né provare con assoluta certezza che essa debba ripetersi. La scienza per il passato è una *descrizione,* per il futuro una *credenza*; non è, e non è mai stata, una spiegazione »[30].

Il passo ora citato appare di importanza fondamentale per capire il nesso tra la correlazione statistica e la concezione epistemologica, in Pearson: se la scienza non è in grado di spiegare niente, allora la conoscenza non serve che a fondare aspettative per il futuro. A tal fine le è necessaria l'uniformità del ripetersi delle sequenze percettive, ovvero il crearsi di *routines* percettive: il potere del pensiero svanisce infatti di fronte alla casualità assoluta, al caos, alla mancanza di elementi permanenti; la necessità appartiene perciò al mondo dei concetti, non a quello delle percezioni. La conoscenza non è altro che ricerca e concettualizzazione delle *routines* dell'esperienza.

Per il Pearson statistico la raccolta indiscriminata di dati, con cui inizia ogni ricerca, si traduce con naturalezza nella compilazione di statistiche esaustive per ogni tipo di sequenza percettiva, fra le quali selezionare, successivamente, quelle sequenze abbastanza stabili, diventate *routines,* sulle quali stabilire le aspettative future[31].

[29] Ivi, p. 155.

[30] Ivi, p. 136.

[31] Risorge qui l'inquietante problema dell'induzione; cosa ci autorizza a credere che la *routine* non si interromperà? Cosa giustifica la nostra fiducia nella scienza? Pearson sembra rispondere a questo interrogativo in due modi sostanzialmente differenti, non so quanto fra loro congruenti. Per un verso, egli sostiene che la natura è un costrutto della mente umana ed è condizionata dalla nostra facoltà percettiva: in sostanza l'uniformità della natura è garantita dalla costanza di quest'ultima (cfr. *The*

Se l'idea appare semplice, meno semplice sembra precisarne il significato concreto. In ogni istante, infatti, ciascuno di noi è bombardato da un elevato numero di impressioni sensoriali concomitanti ed eterogenee. Come destreggiarsi in questo caos percettivo? La pretesa che i fatti debbano essere raccolti «senza alcuna idea preconcetta», se presa alla lettera, cozzerebbe contro difficoltà insormontabili; ma forse le parole di Pearson non vanno prese alla lettera: del resto, anche Darwin, pur dichiarando di aver raccolto i dati senza seguire nessuna teoria, aveva pur proceduto a raccogliere esclusivamente fatti relativi ad un campo omogeneo di fenomeni! Dobbiamo quindi ammettere che la raccolta dei dati non sia del tutto indiscriminata e che in pratica debba essere presente una qualche idea categoriale in grado di guidare almeno una prima selezione dei dati rilevanti da quelli che non sono tali. Ma anche così il numero e l'eterogeneità delle impressioni sensoriali suscettibili di essere raccolte rimane considerevole; tanto più che tutta questa operazione di raccolta, di classificazione e di generalizzazione dovrebbe basarsi sulla sola esperienza!

Per chiarire il senso delle *routines* dell'esperienza ricercate da Pearson, consideriamo le due seguenti sequenze numeriche: α) {5,9,78,24,91}; β) {1,2,4,8,16}. La sequenza β, rispecchiando una nota progressione, sembrerebbe di primo acchito la più attendibile per fondarvi una conoscenza. Per Pearson, viceversa, non sussisterebbe a priori alcuna differenza apprezzabile tra le due sequenze, poiché soltanto il ripetersi dell'esperienza può decidere circa la portata conoscitiva di una qualsiasi sequenza percettiva. Perciò, se la sequenza α, apparentemente una sequenza casuale di numeri, si ripetesse nell'esperienza fino a costituirvi una *routine,* essa acquisterebbe con ciò anche una portata conoscitiva e previsiva.

Diviene così evidente l'importanza di un criterio che metta in grado di dare un «peso» all'esperienza: quante volte deve essere ripetuta una sequenza percettiva affinché costituisca una *routine* e possa essere ritenuta stabile e non puramente casuale ?

È nell'ambito di questa problematica che vanno inseriti i contributi statistici di Pearson a cominciare dalla teoria della correlazione. Gli aspetti salienti della sua concezione epistemologica hanno caratterizzato il materiale della scienza come sequenze di fenomeni *correlati* e definito l'indagine scientifica come ricerca di *correlazioni* atte ad essere generalizzate mediante adeguate concettualizzazioni. Il problema che si pone ora è quello di come selezionare le correlazioni significative ai fini conoscitivi.

Il problema non è da poco per un empirismo abbastanza radicale, sia pure così contraddittorio come quello di Pearson! Dal punto di vista

Grammar of Science, cit., p. 176). Per altro verso, egli esprime l'opinione che la nostra fiducia nel costante ripetersi delle sequenze percettive nel futuro trovi sostegno nella teoria delle probabilità ed in particolare nella regola di successione di Laplace.

fenomenico, infatti, tutte le percezioni si collocano su uno stesso identico piano di esperienza soggettiva: non vi è nulla in esse che permetta di operare delle discriminazioni. Anche il sogno fa parte dell'esperienza soggettiva con la stessa forza di immediatezza di qualsiasi altra impressione sensoriale. Dal punto di vista della conoscenza scientifica, invece, non tutte le catene percettive possono essere poste sullo stesso piano. Ci troviamo così di fronte alla necessità di un metodo che ci guidi nella selezione delle correlazioni significative: occorre passare dal piano epistemologico a quello metodologico.

Pearson fu indotto a questo passo sicuramente anche dall'incontro con Galton e dal conseguente interesse per la biometria e l'eugenetica. La teoria della correlazione, del resto, era nata, come si è detto, proprio con Galton, ma, non avendola egli resa rigorosa mediante lo strumento matematico, spetta a Pearson il merito di aver messo a punto il primo metodo matematico per determinare il coefficiente di correlazione, in pratica una misura (compresa tra +1 e -1) dell'intensità di una correlazione.

Per misurare una correlazione sono sufficienti due variabili, cioè i rilevamenti statistici di almeno due gruppi di caratteristiche che si vogliono confrontare (ad esempio alimentazione e tumori intestinali; peso e statura ecc.). Se riportiamo ora le misurazioni di ciascuno dei due gruppi rispettivamente sull'asse delle ascisse e sull'asse delle ordinate di un sistema di coordinate cartesiane, otterremo due linee (il metodo di Pearson si limita alle regressioni lineari) che potranno avere andamenti divergenti o convergenti, ma che si incontreranno sempre nel punto corrispondente ai rispettivi valori medi.

Intuitivamente, allora, quanto più le due linee tenderanno ad unificarsi, identificandosi o formando tra loro un angolo acuto, tanto più alta sarà la correlazione; viceversa, tanto più le due linee tenderanno a divergere, quanto più bassa sarà la correlazione, fino ad essere nulla quando le due linee vengano a formare tra loro un angolo retto. Una correlazione può, ovviamente, essere diretta o inversa (le variabili cioè possono entrambe aumentare, o diminuire, di pari passo, oppure una può crescere al diminuire dell'altra): tale andamento è rilevato dal segno positivo o negativo del coefficiente. In breve, il coefficiente di correlazione permette di valutare andamento e intensità di una correlazione mediante la descrizione matematica del comportamento delle variabili considerate, descritto in termini di dispersione dai rispettivi valori medi.

In pratica, si può cercare una correlazione tra tutto ciò che si vuole; il guaio sta nel fatto che generalmente si riesce a trovarla! Basti ricordare la 'storica' correlazione inversa scoperta negli anni Cinquanta tra le quotazioni in borsa e la lunghezza delle gonne femminili! Negli anni Cinquanta fu riscontrata in Inghilterra un'alta correlazione diretta, relativamente al periodo 1930-56, tra il numero delle radio e delle televisioni diffuse tra la popolazione ed il numero dei decessi a seguito di disturbi coronarici. Una perfetta correlazione inversa fu rilevata, svariate decine di anni addietro, tra il calo della temperatura in Inghilterra e l'aumento delle vendite di cappotti in Germania. E si potrebbe continuare!

Su correlazioni statisticamente significative si basano, del resto, anche certe previsioni nel giuoco del lotto, giustificate con strampalate teorie circa la 'simpatia' tra i numeri, i quali manifesterebbero una spiccata tendenza ad uscire... con gli amici!

Il fatto é che la significatività statistica di una correlazione non è che l'espressione tautologica, seppure sintetica e quantitativa, dell'esperienza; essa richiama l'attenzione su particolari aspetti di questa, ma non dice nulla circa il valore da attribuirvi.

Una correlazione statisticamente significativa, non solo non è garanzia di uniformità effettivamente stabili, ma, in particolare, non afferma niente circa l'esistenza di nessi causali tra i fenomeni correlati. Se A e B sono fenomeni statisticamente correlati, potrebbe essere A la causa di B, o viceversa; ma potrebbero entrambi dipendere da una ignota causa comune ed essere tra loro del tutto indipendenti. Infine potrebbe non essere nulla di tutto questo, perché anche l'improbabile ha un posto nel mondo. Questo fatto va sempre tenuto presente, se non si vuol rischiare di interpretare, come spesso succede, una correlazione in termini di causa-effetto.

In breve, il coefficiente di correlazione permette di restringere il campo dell'esperienza, isolandone i fenomeni significativamente correlati dal punto di vista statistico da quelli che non lo sono. Ma niente di più. Possiamo osservare, sulla base di quanto ora detto, che, se l'analisi delle correlazioni può valere come metodo per la descrizione statistica di eventi passati, più problematico sembra il tentare di fondarvi le previsioni future. La significatività matematica, così valutata, esprime infatti ancora un rapporto troppo debole fra i fenomeni per garantire, entro limiti ragionevoli, la stabilità di quelle stesse correlazioni rilevate. Se in passato certe sequenze percettive sono state altamente correlate tra loro, cosa esclude che tale correlazione si sia verificata per caso e che essa sia destinata a non più ripetersi ?

Per le finalità conoscitive la teoria della correlazione rivela la propria debolezza e insufficienza, mostrandosi inadeguata come criterio empiristico idoneo a rilevare quegli aspetti uniformi dell'esperienza che siano anche conoscitivamente rilevanti. La possibilità che risultino statisticamente significative delle correlazioni del tutto inconsistenti dal punto di vista scientifico e l'impossibilità di discriminare, con questo solo criterio, le correlazioni scientificamente rilevanti, da quelle che non lo sono, sollecitano un ulteriore sforzo metodologico in direzione dell'elaborazione di un metodo che permetta anche di escludere, con un grado di probabilità da determinare, che certe correlazioni rilevate siano puramente accidentali. Tale passo ulteriore troverà realizzazione nel test

χ^2 (*chi-quadro*), elaborato dallo stesso Pearson nel 1900, e, in generale, con i *test di significatività*.

4.3 Fisher e i test di significatività

I test di significatività sembrano idonei a risolvere la questione ora sollevata, ossia quella di fornire un metodo che permetta di capire quale valore possa essere attribuito ad una correlazione e di eliminare le correlazioni accidentali. Perciò i test di significatività rappresentano lo sviluppo logico della teoria della correlazione. Essi sono costituiti da un insieme di numerosi test statistici, ognuno dei quali è idoneo a trattare nel modo più adeguato determinate questioni in determinati tipi di situazione. Ad esempio, un test adatto per l'analisi dei dati quantitativi, come il *t-Student*[32], non è adatto anche per l'analisi dei dati qualitativi, per la quale si presta assai meglio il test χ^2. Test atti a confrontare il comportamento di due fucili non necessariamente sono adatti anche per confrontare quello di tre; un test idoneo per la valutazione del tasso di incidenti in differenti turni di lavoro di una fabbrica non può in genere essere utilizzato per confrontare l'influenza di differenti fertilizzanti sui raccolti agricoli. Vi sono inoltre test ragionevolmente accurati per piccolissimi campioni (fino a 50 individui), come alcuni proposti da Fisher, ed altri che risultano accurati solo per campioni assai grandi. I test di significatività, insomma, variano per estensione della loro applicabilità, nonché per la loro accuratezza ed efficienza, oltre che per la maggiore o minore complessità dell'apparato matematico necessario. Dal punto di vista operativo, quindi, si presenta al ricercatore non solo la questione di come eseguire il test, ma anche quella di quale test sia più opportuno utilizzare nella situazione specifica.

Tutti i test di significatività, però, sottostanno ad una medesima ed unica logica, che Fisher[33] ha contribuito a mettere in chiaro.

Per quanto abusato, il noto esempio fisheriano della «signora del tè» è forse il più semplice modo per introdurre alla logica dei test di significatività: una signora afferma di saper distinguere, al sapore, se una tazza di tè con latte sia stata preparata mettendovi per primo il tè oppure

[32] Questo test fu scoperto nel 1908 da «Student», pseudonimo sotto il quale William Sealy Gosset, un chimico che lavorava per Guinness, la famosa fabbrica inglese di birra, pubblicò il suo lavoro. Il test *t-Student*, nella forma riveduta successivamente da Fisher, costituisce, assieme al χ^2, uno dei più usuali strumenti di analisi e di inferenza statistica.

[33] R.A. Fisher (1890-1962), statistico e genetista. Ha contribuito allo sviluppo delle tecniche statistiche ed alla loro applicazione specialmente in biologia e genetica. Insieme a K. Pearson ha dominato la scena statistica della prima metà del novecento. Oltre alle opere di statistica, ha scritto *Genetical Theory of Natural Selection* (1930), che costituisce un punto di riferimento obbligato nella storia della teoria evoluzionistica.

il latte. Si tratta quindi di progettare un esperimento di prova. L'ipotesi in riferimento alla quale l'esperimento sarà condotto, che Fisher chiama *ipotesi-nulla* e che è caratteristica di tutti i test di significatività, è l'ipotesi che le risposte della donna siano date completamente a caso; l'esperimento ha l'unico compito di confutare tale ipotesi.

Fisher propone quindi che l'esperimento consista nel presentare in ordine casuale alla donna, per l'assaggio, otto tazze di tè, quattro preparate in un modo e quattro nell'altro. La signora dovrà dividere le otto tazze in due gruppi omogenei di quattro.

Per comprendere il disegno dell'esperimento gioverà tener presente il ragionamento che segue: ci sono settanta modi di scegliere due gruppi di quattro da otto oggetti (cioè: 8!/4!.4!), e tra questi uno solo è quello giusto. C'è quindi una probabilità pari a 1/70 (=0,014) che la signora indovini a caso. Tale probabilità è ritenuta sufficientemente bassa per permettere di respingere l'ipotesi-nulla nel caso che le otto tazze siano correttamente individuate.

L'esempio ora esposto ci permette di mettere subito a fuoco la forma caratteristica dei test di significatività; essi si presentano essenzialmente come metodi volti a sondare l'incidenza del caso nei risultati di un esperimento. Questa loro caratteristica, del resto, è appunto quella che li lega, su una stessa linea di sviluppo metodologico, alla teoria della correlazione. Abbiamo appena finito di notare, infatti, come quest'ultima non fosse sufficiente a soddisfare le esigenze della scienza a causa della sua scarsa capacità selettiva: individuate delle alte correlazioni statistiche, non si è ancora in grado di asserire se esse possano risultare anche significative in senso scientifico, finché non sia stata esclusa l'accidentalità del loro verificarsi.

Il ragionamento contro il caso, a dire il vero, non è nuovo. Connesso con il calcolo delle probabilità lo troviamo ripetutamente riproposto a partire dal sec. XVIII. Già nel 1710 J. Arbuthnot aveva usato il calcolo delle probabilità per dimostrare che l'equilibrio tra nascite maschili e nascite femminili non poteva essere ascritto al caso, bensì alla divina provvidenza; De Moivre notava esplicitamente, nella prefazione alla prima edizione del suo trattato, che la dottrina delle probabilità poteva essere usata per distinguere gli eventi prodotti dal caso da quelli prodotti secondo un piano ben definito. Argomentazioni simili furono avanzate successivamente anche da Daniel Bernoulli (in relazione alle inclinazioni dei piani delle orbite dei pianeti), da L. de Maupertuis (sulla polidattilia), da Mitchell (sulle Pleiadi). Estesa è infine, l'applicazione, da parte di Laplace, del teorema di Bayes ai fenomeni naturali, dove esso è utilizzato, come abbiamo visto, *esattamente come un test di significatività*, volto cioè a dimostrare l'improbabilità del caso.

La forza del ragionamento, che presiede ai test di significatività, riposa logicamente, come osserva Fisher, sulla forma della disgiunzione «o un evento eccessivamente raro è occorso oppure la teoria della distribuzione casuale è falsa»[34]. Si tratta, in sostanza, del ben noto argomento *modus tollens,* tradotto in termini di probabilità, e quindi applicato ad un'unica, determinata, ipotesi, l'ipotesi del Caso o *ipotesi nulla.* Tale struttura logica è costante in tutti i test di significatività.

In generale, si può dire che un test di significatività opera la distribuzione dei risultati possibili di un esperimento in due classi: da un lato i risultati che mostrano una discrepanza significativa con l'ipotesi-nulla, dall'altro quelli che non la mostrano. La linea di demarcazione tra le due classi, detta *livello di significatività,* corrisponde a quel grado di probabilità che si è disposti ad accettare come sufficientemente basso da permettere di considerare un risultato così 'raro', sotto l'ipotesi che esso sia dovuto al caso, da basarvi il rifiuto dell'ipotesi-nulla.

Nella pratica, si usano livelli standard di significatività: 0,05 generalmente, oppure, specie in campo medico, 0,01; talora addirittura 0,001. Tuttavia, «nessun operatore scientifico ha un livello di significatività fissato al quale anno dopo anno, e in tutte le circostanze, rifiuta le ipotesi; egli valuta piuttosto ogni caso particolare alla luce della sua evidenza e delle sue idee»[35]. Una volta scelto tale livello, l'ipotesi-nulla è respinta se i risultati sperimentali cadranno al di sotto di esso.

L'esecuzione di un test di significatività consiste sempre in un esperimento atto a valutare quanto i suoi risultati differiscano da ciò che ci si dovrebbe attendere in base al caso, ad un livello di significatività prescelto. Si tratta sempre, in sostanza, di valutare una differenza.

Alcuni esempi serviranno a comprendere varie situazioni sperimentali.

In una fabbrica si sono verificati in un mese 15 incidenti, differentemente distribuiti nei tre diversi turni lavorativi: 7 nel turno A, 7 nel turno B, 1 nel turno C. L'ipotesi da sottoporre al test (ipotesi-nulla) è la seguente: non sussiste alcuna differenza significativa nei tre turni (le differenze sono accidentali). Al caso in esame si adatta il test χ^2.

Calcoliamo anzitutto il valore teorico E, vale a dire la frequenza di incidenti, in ogni turno di lavoro, che ci dovremmo attendere sulla base dell'ipotesi-nulla: poiché i tre turni hanno una stessa durata e poiché gli incidenti sono riferiti ad uno stesso periodo di tempo, non vi è motivo per non attendersi una uguale ripartizione degli incidenti nei tre turni. Perciò $E = 15/3 = 5$.

[34] R.A. Fisher, *Statistical Methods and Scientific Inference,* London 1956, p. 39.

[35] Ivi, p. 42. Come si vede, la pretesa oggettività di Fisher trova un limite in un inevitabile elemento di valutazione soggettiva. Ciò è quanto non manca di rilevare de Finetti. Lo stesso Fisher ammette, del resto, che «il livello di significatività soddisfa le condizioni di una misura dei motivi razionali di non credenza».

Procediamo ora al relativo calcolo per χ^2. Poiché la formula per calcolarlo è la seguente

$$\chi^2 = \sum \frac{(O-E)^2}{E} \text{ (O = frequenze osservate; E = frequenze teoriche),}$$

avremo:

$$\chi^2 = \frac{(7-5)^2 + (7-5)^2 + (1-5)^2}{5} = \frac{4+4+16}{5} = \frac{24}{5} = 4{,}8$$

Dovremo ora consultare l'apposita tabella per il test χ^2, al fine di scoprire a quale livello di significatività si trova il valore ora ottenuto. Ma per poter consultare la tabella occorre prima calcolare anche i *gradi di libertà* (che sostengono un ruolo importante in tutti i test di significatività), perché il valore di χ^2 varia non solo in funzione dell'ampiezza degli scostamenti dei valori osservati da quelli teorici, ma anche del loro numero. Essi servono a dare una misura del margine di giuoco entro cui si trovano i risultati sperimentali. I gradi di libertà (GDL) per il test χ^2 possono essere trovati facilmente applicando la regola seguente:

a) quando un singolo insieme di C elementi deve essere confrontato con una proporzione nota, o teorica, GDL = *C*-1;
b) quando due o più classi R, divise in C sottogruppi, sono confrontate tra loro, GDL = (R -1)(C -1).

Il nostro esempio rientra nel primo caso e pertanto i gradi di libertà sono (3 - 1) = 2. Nella tabella per il test χ^2 troveremo ora che il valore 4,8, a *2* gradi di libertà, indica che le differenze osservate di incidenti nei diversi turni hanno una probabilità di verificarsi per puro caso superiore allo 0,05. Perciò, a questo livello, l'ipotesi-nulla non può essere respinta.

Nell'esempio ora considerato, il valore teorico di riferimento era facilmente calcolabile, essendo costituito dalla frequenza media «ideale» calcolata sotto l'ipotesi di una distribuzione casuale. Ma tale valore può essere offerto anche da una frequenza empirica nota (ad es., i morti annui per infarto in una data popolazione), oppure può essere costituito sperimentalmente mediante un parallelo campione di controllo. Per sottoporre a test la significatività dei risultati ottenuti dalla somministrazione di un farmaco si procede, generalmente, prendendo in esame due campioni casuali di pazienti affetti dalla medesima patologia e trattando uno di essi con il farmaco, l'altro con un diverso trattamento (un placebo o altro); l'ipotesi-nulla affermerà che i due trattamenti si equivalgono, cioè che qualsiasi differenza riscontrata tra i due gruppi di pazienti sarà puramente accidentale.

A Calcutta, durante l'epidemia di colera del 1894, furono tenute sotto osservazione 818 persone esposte al rischio di contrarre la malattia. Di queste, 279 erano state vaccinate e tra di esse vi furono 3 casi di colera. Le altre 539 persone non erano state vaccinate e tra esse vi furono 66 casi di colera. L'ipotesi-nulla afferma che la differenza riscontrata tra il gruppo dei vaccinati e quello dei non-vaccinati è puramente accidentale.

Al fine di calcolare i valori teorici si ragiona, in questo caso, nel modo seguente: dal momento che i vaccinati sono i 279/818, cioè circa il 34% del totale, dovremmo attenderci tra di essi il 34% dei 69 casi infetti, vale a dire un valore di 23,5.

Di conseguenza, (279 - 23,5) = 255,5 dovrebbero essere i casi di persone non infettate tra i vaccinati. Tra i non-vaccinati dovremmo attenderci invece (69 - 23,5) = 45,5 casi di colera (cioè il 66% di 69), mentre (539 - 45,5) = 493,5 dovrebbero essere i casi di persone non infettate.

A questo punto, si può procedere al calcolo del χ^2 applicando la formula precedente:

$$\chi^2 = \frac{(3-23{,}5)^2}{23{,}5} + \frac{(276-255{,}5)^2}{255{,}5} + \frac{(66-44{,}5)^2}{44{,}5} + \frac{(473-493)^2}{493{,}5} = 29{,}6$$

Tale valore di χ^2 ad 1 grado di libertà corrisponde ad un livello di significatività minore dello 0,2%. La differenza tra i due gruppi è perciò altamente significativa e l'ipotesi-nulla può essere respinta.

Differenti tipi di problemi devono essere affrontati con test di significatività diversi.

Darwin cercò di determinare se vi fosse una differenza nella crescita delle piante ibridate rispetto a quelle autofecondate. In questo caso è preferibile un esperimento che operi un confronto 'a coppie', perché in tal modo è possibile ottenere la massima uniformità delle condizioni ambientali: ogni pianta di un tipo viene cioè associata ad una dell'altro. Darwin scelse 15 coppie di piante e misurò le rispettive altezze ad una data prefissata. In rapporto a questo esperimento, l'ipotesi-nulla asserisce che qualsiasi differenza in altezza riscontrata in ogni coppia sia da ritenersi puramente accidentale e che perciò i limiti a cui tendono le medie delle altezze dei due tipi di piante sono uguali. Ciò significa che le differenze in altezza si distribuiscono normalmente attorno ad un valore medio uguale a zero, o, ciò che è lo stesso, che tutte le misurazioni sono campioni di una medesima popolazione 'normale'[36]. Si tratta quindi di pervenire a giudicare se le differenze rilevate tra le altezze delle coppie di piante siano compatibili con l'ipotesi nulla, o se siano invece così grandi da consentirne il rifiuto.

Il test indicato in questo caso è il *t-Student,* che dipende solamente dal numero dei confronti indipendenti (i *gradi di libertà,* che, in questo caso,

[36] *Normale*, in senso statistico, è una popolazione che si distribuisce secondo una curva 'a campana'.

sono 15 - 1). In pratica, il calcolo è abbastanza semplice. Una volta rilevate tutte le altezze, avremo, per ogni coppia, una misura della differenza tra le loro altezze (tale differenza può avere segno positivo o negativo, a seconda dalla pianta più alta). Siano, ad esempio, 49, -67, 8, 16, 6, 23, 28, 41, 14, 29, 56, 24, 75, 60, -48 le differenze rilevate nelle 15 coppie. La somma di queste differenze è perciò 314, che, divisa per 15, il numero delle coppie, dà il valore 20,933, che è la *differenza media,* la quale, moltiplicata per 314, darà 6573. Calcoliamo ora il quadrato di ogni differenza e sommiamo i valori ottenuti: otteniamo il numero 26518 dal quale verrà sottratto il numero 6573 per ottenere il valore della discrepanza tra le differenze osservate nelle 15 coppie: 19945. Calcolato ora, con i valori sopra ricavati, l'*errore standard,* il valore di *t* potrà essere ottenuto dal quoziente tra la differenza media e l'errore standard; nel nostro caso, *t* risulterà uguale a 2,148. Non resterà poi che ricorrere alle tavole per la distribuzione del *t-Student* e cercarvi, in corrispondenza del valore 2,148 e 14 gradi di libertà, il rispettivo livello di significatività.

È opportuno ricordare, sia pur brevemente e per inciso, che un test statistico può essere, come si suol dire, a 'una coda' o a 'due code'; può cioè prendere in considerazione l'allontanamento da una media, o da un valore atteso, in una sola direzione oppure in ambedue le direzioni possibili (maggiore e minore). La scelta del test dovrebbe logicamente dipendere dall'ipotesi che si vuol controllare: una 'differenza' da una media rimanda ad un test a due code; 'essere maggiore' di un dato valore richiede un test a una coda. Capita purtroppo che gli sperimentatori, invece di far dipendere la scelta del test dall'ipotesi sotto controllo, decidano quale tipo di test usare solo dopo aver visto i risultati sperimentali ed optino, ad esempio, per un test ad una coda sulla base del fatto che la deviazione riscontrata si estende in una sola direzione. Le tabelle di consultazione per i livelli di significatività sono generalmente predisposte tanto per i test a una coda che per quelli a due code, ma è necessario fare attenzione, perché le risposte saranno assai differenti a seconda del test che si usa (tanto per farsi un'idea: con riferimento ad una distribuzione normale, il livello di significatività di un test a due code sarà il doppio di quello di un test a una coda).

Ho ritenuto doveroso menzionare tale fatto; tuttavia, poiché i test di significatività devono essere interpretati come prove contro il caso, mi sembra ragionevole assumere che si debba sempre fare riferimento ad un test a due code, perché le deviazioni casuali da una media, o da un valore teorico, non hanno una direzione predeterminata.

4.4 L'interpretazione dei test di significatività

Siamo ora in condizione di poter analizzare e discutere più dettagliatamente le caratteristiche peculiari dei test di significatività.

La prima caratteristica, sulla quale non si insisterà mai abbastanza, è rappresentata, come si è detto, dall'ipotesi chiamata a confrontarsi con i dati sperimentali, l'ipotesi-nulla. Tale ipotesi è sempre, nei test di significatività, esclusivamente l'ipotesi del caso, l'ipotesi cioè che non afferma nulla di significativo nei fenomeni osservati, ma che attribuisce ai capricci del caso qualsiasi apparente relazione, o differenza, riscontrata fra i dati sperimentali. L'ipotesi-nulla è per Fisher «caratteristica di tutte le sperimentazioni» ed è l'unica ipotesi in relazione alla quale un esperimento può mostrarsi 'significativo' perché è la sola che, a causa della sua determinatezza, permetta di programmare un esperimento in grado di contraddirla: «in relazione ad ogni esperimento possiamo parlare di questa ipotesi come della "ipotesi nulla", e si potrebbe notare che l'ipotesi nulla non è mai provata o stabilita, ma è possibile confutarla nel corso della sperimentazione. *Ogni esperimento può esser detto esistere solo al fine di dare ai fatti una possibilità di confutare l'ipotesi nulla*»[37].

Poiché l'ipotesi-nulla non può mai essere 'provata', un esperimento che non riesca a contraddirla non comporta, per converso, la sua accettazione; nei test di significatività non ci si muove all'interno della dicotomia vero/falso, o accettato/rifiutato: sarà bene non dimenticare mai che essi non sono test per l'accettazione di ipotesi, sebbene tale confusione sia spesso assai diffusa anche tra gli stessi ricercatori.

Si può mettere meglio a fuoco il senso e la funzione dell'ipotesi-nulla se si tiene presente il quadro epistemologico sopra delineato. Era stato posto in evidenza, poco sopra, come il pensiero empiristico e positivistico assegnasse alla scienza l'unica funzione di trovare correlazioni stabili e di generalizzarle, ed avevamo osservato che la teoria della correlazione non permette di operare una selezione scientificamente significativa tra le correlazioni osservate. I test di significatività sono stati quindi introdotti come strumenti più efficaci per gli obiettivi scientifici, in particolare per sgombrare il campo dalle correlazioni accidentali, espresse dall'ipotesi-nulla. È chiaro quindi che l'ipotesi-nulla non possa essere mai accettata: l'interesse che muove la ricerca scientifica è volto verso le correlazioni significative, non verso quelle che non sussistono. Una correlazione significativa, in breve, è un fatto empirico, conoscitivamente rilevante; una correlazione non significativa esprime solo un accidente empirico e, come tale, non ha significato scientifico. Perciò l'ipotesi-nulla non si pone come un'ipotesi da accettare: essa è solo un indice di rilevanza o irrilevanza scientifica dei fenomeni, o, al massimo, della compatibilità che può esservi tra i dati empirici ed un'ipotesi. L'ipotesi-nulla, dunque, può essere solo respinta, confutata: *è l'ipotesi che non afferma nulla di significativo*!

37 R.A. Fisher, The *Design of Experiments*, London 1949, p. 16.

Che cosa succede, però, quando l'ipotesi-nulla viene confutata dai dati sperimentali? Siamo autorizzati, sulla base di quei medesimi risultati, ad accettare un'ipotesi alternativa? Se la 'signora del tè' individua correttamente le otto tazze con la bevanda, siamo autorizzati a riconoscere che essa possiede la straordinaria capacità di percepire differenze di sapore che forse nessun altro percepisce? Se i risultati sperimentali rendono insostenibile l'ipotesi che le guarigioni riscontrate in un gruppo di malati di diabete curati con un certo farmaco siano puramente accidentali, siamo autorizzati ad accettare come *provata* l'ipotesi che il farmaco usato nella sperimentazione sia efficace nella cura del diabete?

La risposta è: *no*, non siamo autorizzati a farlo; o, quanto meno, non siamo autorizzati a farlo sulla sola base dei test di significatività. Nel modo più categorico, il rifiuto dell'ipotesi-nulla non comporta, come sua conseguenza, l'accettazione di nessuna ipotesi alternativa. Ciò vale in generale per i test di significatività: essi sono in grado di esprimersi esclusivamente in relazione al rifiuto dell'ipotesi-nulla.

Fisher richiama l'attenzione anche sui motivi 'tecnici' che impediscono che l'ipotesi sottoposta a test di significatività sia diversa da quella casuale. Dice al proposito Fisher, a conclusione della discussione sull'esempio della 'signora del tè':

> Si potrebbe arguire che se un esperimento può confutare l'ipotesi che una persona non possiede alcuna capacità sensoriale discriminatoria tra due tipi di oggetti, dovrebbe di conseguenza essere in grado di provare l'ipotesi opposta, che essa può compiere una tale discriminazione. Ma quest'ultima ipotesi, per quanto ragionevole o vera possa essere, non può essere eletta come una ipotesi nulla da sottoporre a prova sperimentale, perché non è puntualmente determinata. Se essa asserisse che la persona non sbaglia mai nei suoi giudizi, di nuovo avremmo una ipotesi puntuale, ed è facile vedere che questa ipotesi potrebbe essere confutata da un solo errore, ma non potrebbe mai essere provata da nessun finito ammontare di esperimenti. È evidente che l'ipotesi nulla deve essere puntualmente determinata, libera da vaghezza e ambiguità, perché essa deve fornire la base del «problema di distribuzione» del quale il test di significatività fornisce la soluzione»[38].

Queste condizioni sono soddisfatte soltanto dall'ipotesi-nulla, cioè dall'ipotesi del Caso. Sebbene, infatti, anche le ipotesi universali siano ben determinate e falsificabili da un solo insuccesso, non si vede come potrebbe essere progettato un esperimento volto alla loro confutazione. Se in un'urna di infinite palline vi fosse, ad esempio, una sola pallina

[38] *Ibid.*

nera, come potremmo programmare un esperimento atto a confutare l'ipotesi che tutte le palline sono bianche? Sebbene tale ipotesi possa essere empiricamente falsificata, non sembra tuttavia esservi alcun test in grado di programmare un esperimento volto alla sua confutazione. Resta pertanto solo l'ipotesi casuale quale ipotesi che permetta di programmare un esperimento in grado di contraddirla.

Siamo qui di fronte ad uno dei punti forse più delicati per la comprensione del significato dei test di significatività; il punto in cui si sono innestati tutti i fraintendimenti, denunciati dallo stesso Fisher, di cui tali test sono stati e sono tuttora oggetto.

A che servono dunque i test di significatività, se essi sono in grado di esprimersi soltanto a proposito del rifiuto dell'ipotesi-nulla? Qual può essere il loro contributo alla ricerca scientifica?

La natura dei test di significatività apparirà più comprensibile se, come Fisher stesso suggerisce, terremo per fermo il fatto che essi devono essere pensati come destinati ad essere applicati ad una sola, isolata, ipotesi (quella *nulla* appunto), e non ad una coppia o ad un insieme di ipotesi alternative. Se ne deve concludere, quindi, che il rifiuto dell'ipotesi-nulla non *può* implicare l'accettazione di una ipotesi alternativa: in primo luogo, per il fatto che *non ci sono ipotesi alternative nella logica dei test di significatività*; in secondo luogo, perché l'esperimento programmato per confutare l'ipotesi-nulla non è idoneo a provare contemporaneamente alcunché: «un test di significatività non contiene alcun criterio per "accettare" ipotesi»[39] .

Dobbiamo pertanto cercare una interpretazione dei test di significatività a partire da questo assunto fondamentale, che essi siano chiamati ad esprimersi esclusivamente contro l'ipotesi del caso. Si tratta allora di vedere se questo compito abbia una sua funzione scientifica e quale essa sia.

Per affrontare questa questione, immaginiamo che sia stato eseguito un test per valutare la significatività dei risultati relativi alla somministrazione di un certo farmaco per la cura del diabete e che esso abbia portato ad un netto rifiuto dell'ipotesi-nulla. Può sembrare a prima vista strano che non si possa anche dedurne che il farmaco sperimentato sia effettivamente efficace nella cura del diabete, dal momento che il farmaco è stato appunto usato nella sperimentazione. Tuttavia, prima di comprometterci con prese di posizione che potrebbero risultare poco sostenibili, immaginiamo che venga effettuato un esperimento in tutto simile al precedente, ma con l'unica differenza che la sostanza somministrata consista in una mistura, di sospetta composizione, messa insieme da una specie di stregone; siano, anche in questo caso, i risultati sperimentali nettamente significativi

[39] R.A. Fisher, *Statistical Methods and Scientific Inference*, cit., p. 42.

contro l'ipotesi-nulla. Sia dato inoltre per scontato che i test siano stati eseguiti nel migliore dei modi e che nulla vi sia da eccepire sui risultati ottenuti. Troveremmo ora altrettanto paradossale la resistenza ad accettare l'ipotesi che la mistura sia davvero efficace nella cura della patologia considerata? Oppure, non preferiremmo fermarci a questo risultato ed effettuare altre ricerche, sollecitate proprio da esso? Probabilmente il nostro auspicabile spirito scientifico ci porterebbe a non accettare quella mistura come cura valida, se non dopo che un'accurata analisi chimico-farmacologica avesse chiarito che tra i componenti di essa non ve ne sono di nocivi, ma che ve ne sono alcuni, anzi, le cui proprietà terapeutiche non possono essere disconosciute.

È chiaro che i due differenti atteggiamenti, di immediata propensione per l'accettazione delle proprietà terapeutiche del farmaco, nel primo caso, di diffidenza e di cautela nel secondo, non possono essere messi a carico dei test eseguiti. Anzi, essendo nella nostra supposizione i due test assolutamente identici, non sarebbe neppure possibile giustificare due comportamenti differenti. La logica sottesa all'interpretazione dei test di significatività non può essere cambiata a piacere: o il rifiuto dell'ipotesi-nulla implica sempre l'accettazione di una ipotesi ad essa alternativa, oppure non la implica mai.

Riflettiamo sul motivo per il quale ci sembrerebbe ragionevole accettare, nel primo caso, l'ipotesi che il farmaco sia efficace, e, nel secondo, no. Il fatto è che né la propensione ad accettare la bontà del farmaco, in un caso, né quella a rifiutare l'ipotesi di efficacia della mistura, nell'altro, derivano logicamente dai test di significatività. Tali atteggiamenti sono piuttosto determinati dalle informazioni aggiuntive in nostro possesso (ad esempio, la composizione del farmaco, una competenza clinica, ecc.), e da un generale atteggiamento culturale di fondo (ad esempio, una mentalità scientifica in contrapposizione ad una mentalità superstiziosa), nonché dalla conoscenza delle modalità di esecuzione dell'esperimento medesimo.

Su quest'ultimo aspetto, quello della programmazione e dell'esecuzione dell'esperimento voglio richiamare l'attenzione, perché è su questa base che può, eventualmente, essere giustificato oltrepassare la portata logica dei test di significatività per attingere conclusioni (ad es. «il farmaco è efficace») assolutamente non implicate, anzi neppure contemplate, dalla logica di tali test. È possibile, ad esempio, almeno in via teorica, immaginare di strutturare in laboratorio, sia logicamente che sperimentalmente, una questione sotto indagine in modo tale che il rifiuto dell'ipotesi casuale implichi logicamente l'accettazione di una specificata ipotesi causale; che sia cioè possibile, anche se difficile, costruire in laboratorio una situazione in cui effettivamente valga $h_o \vee h_1$, così che $\neg h_o \rightarrow h_1$. In concreto, ciò significherebbe costruire in laboratorio una situazione nella quale l'unica differenza tra due gruppi di oggetti sotto osservazione sia quella da noi stessi immessa in uno dei due gruppi, una sostanza chimica, ad esempio. In effetti, spesso una programmazione sperimentale mirata tende verso questa situazione ideale; ma essa non è mai rigorosamente ottenibile, neppure nell'ambiente asettico e controllato di un laboratorio!

Sotto il profilo puramente logico, non sempre può essere sbagliato, dunque, accettare una ipotesi sulla base dei risultati sperimentali dei test di significatività; ma sarebbe senz'altro sbagliato credere che ciò sia logicamente autorizzato dai test medesimi, perché, in tal caso, prima o poi sbaglieremmo sicuramente. L'ipotesi casuale, infatti, non può essere pensata come una tra le ipotesi di un insieme esaustivo, come se si trattasse di una delle distribuzioni possibili di palline bianche e nere in un'urna; e neppure come uno degli elementi di una coppia dicotomica di ipotesi A ∨ B, a meno che non la si voglia ridurre ad una pura banalità, del tipo caso/non-caso.

L'ipotesi casuale, a rigore, non ha ipotesi alternative, ovvero ne ha potenzialmente un numero quantitativamente e qualitativamente indeterminato. Dalla logica dei test di significatività non consegue, a favore di una qualche ipotesi alternativa all'ipotesi-nulla, conferma maggiore di quanta non ne ricavasse Laplace, dall'alta improbabilità dell'ipotesi che i moti dei pianeti fossero puramente casuali, a favore dell'ipotesi di gravitazione universale: qualsiasi altra ipotesi deterministica ne avrebbe tratto lo stesso grado di conferma. Se l'esclusione dell'ipotesi del Caso comportasse *ipso facto* l'accettazione di una qualche ipotesi alternativa, causale, trarrebbero facilmente sostegno dai test di significatività anche ipotesi come quelle del flogisto e dei vortici di Cartesio e, in generale, qualsiasi ipotesi causale, per quanto cervellotica, venisse in mente di contrapporre ad una data ipotesi-nulla!

È mia opinione, pertanto, che la natura dei test di significatività possa essere meglio compresa non solo se, come evidenzia Fisher, essi siano concepiti come rivolti a misurarsi esclusivamente con l'ipotesi-nulla, ma anche e soprattutto ove si tenga presente che questa non è nient'altro che l'ipotesi del puro verificarsi accidentale dei fenomeni, e si rifletta su ciò che questo significhi. L'esclusione del caso, infatti, è da sempre il primo passo della ricerca empirica impegnata a scoprire nessi significativi o a cercare risposte esplicative per fenomeni nuovi o anomali.

Cade così l'apparente paradossalità dei test di significatività: essi esauriscono il loro compito con il rifiuto dell'ipotesi-nulla. Con ciò essi danno al ricercatore un'indicazione preziosa, basata sull'analisi statistica dei dati, contribuendo, nel contesto della scoperta, al progresso della ricerca. L'eventuale restante compito, di spiegazione o di elaborazione dei risultati ottenuti, ricade sulla branca scientifica competente; solo essa potrà decidere come utilizzare tali indicazioni, inserendole in un coerente sistema teorico o promuovendo nuove questioni e nuovi esperimenti.

Si fondono così, nei test di significatività di Fisher, due linee di riflessione convergenti: la prima è quella peculiare del contesto della scoperta, nel quale ha importanza primaria per la scienza empirica eliminare le correlazioni accidentali per fare emergere quelle significative sotto il profilo scientifico. Questa linea, come si è visto, ha radici nell'empirismo classico, matura nel clima positivistico e passa attraverso l'elaborazione della teoria delle correlazioni di Pearson. La seconda è quella statistico-matematica, che individua l'ipotesi del caso come l'unica ipotesi determinata per la quale sia sempre possibile programmare un esperimento atto a confutarla, e logica, che individua nel *modus tollens* il metodo deduttivo valido per il controllo di ipotesi generali.

4.5 Test di significatività e programmazione degli esperimenti

Un secondo aspetto caratterizzante i test di significatività riguarda la situazione sperimentale. Ogni test è infatti una procedura sperimentale che connette strettamente lo strumento statistico di analisi con la programmazione e l'esecuzione dell'esperimento stesso. È merito di Fisher l'aver messo in evidenza questa stretta connessione mostrando come solo l'intervento delle procedure statistiche, basate sul calcolo delle probabilità fin nella fase di programmazione dell'esperimento, possa garantire l'affidabilità dei risultati sperimentali.

La situazione sperimentale merita di essere considerata sotto due aspetti strettamente connessi tra loro, quello del campionamento, cioè degli accorgimenti da osservare per la scelta di un campione in grado di dare ai risultati la massima affidabilità possibile, e quello della programmazione vera e propria dell'esperimento.

Per quanto riguarda la scelta del campione, affinché l'esperimento non risulti già compromesso in partenza, devono essere osservate certe condizioni atte a garantire al massimo la rappresentatività del campione.

Per essere affidabile, entro limiti calcolabili, un campione deve anzitutto essere *casuale,* cioè tale che ogni singolo individuo della popolazione abbia la stessa probabilità di entrare a far parte del campione, e, nel contempo, essere effettivamente rappresentativo dell'intera popolazione da cui è tratto. Generalmente, il procedimento di selezione casuale garantisce anche la rappresentatività, ma non sempre è così, specie se la popolazione è molto varia: in tal caso il procedimento casuale deve accompagnarsi ad altri accorgimenti che preservino la rappresentatività del campione. In pratica, ottenere un campione casuale non è sempre facile: non basta ad esempio porsi all'uscita di un sottopassaggio pedonale e scegliere tra i passanti una persona ogni cinque: l'esclusione di chi si muove abitualmente in macchina potrebbe rendere il campione non rappresentativo e scarsamente affidabile.

Un procedimento sicuramente casuale è quello di servirsi della tavola dei numeri casuali, assegnando ad ogni individuo della popolazione uno di tali numeri ed estraendo quindi a sorte, oppure selezionando sistematicamente su tale tavola (ad esempio scegliendo un numero ogni dieci, od ogni cinque, ecc.). Anche una serie consecutiva di pazienti che si presentano spontaneamente ad uno specialista, per la terapia di uno stesso stato patologico, può essere considerata equivalente ad un campione casuale. Invece, quando la forte variabilità esistente tra gli individui di una popolazione richiederebbe un campione assai grande, e quindi costi eccessivamente alti, per preservare la rappresentatività del campione e garantirne l'affidabilità, è preferibile ricorrere ad un campionamento *stratificato,* che si ottiene suddividendo la popolazione in vari strati omogenei e selezionando quindi, casualmente, campioni da ogni strato (ad esempio una popolazione può essere suddivisa per sesso, per età, per professione ecc.). In tal modo è possibile ottenere un buon grado di precisione con campioni assai più piccoli di quanto occorrerebbe con il campionamento casuale ordinario. La formazione del campione, per concludere, richiede un'accuratissima selezione, che in parte dipende sia dalla conoscenza che possiamo avere delle caratteristiche della popolazione, sia dalla rigorosa formulazione del problema che ci interessa; ad ogni modo, sembra opportuno tener presente che un campione scelto male comporta sempre conclusioni sicuramente fuorvianti.

Un ulteriore importante aspetto della fase di campionamento è quello della scelta dell'*ampiezza* del campione. Il calcolo delle probabilità ci insegna che un campione casuale riflette le caratteristiche della popolazione in modo tanto più accurato quanto più il campione è ampio. Questa accuratezza, o attendibilità che dir si voglia, cresce in misura della radice quadrata della numerosità del campione: in altre parole, occorre aumentare di cento individui il campione per accrescere di dieci volte la sua accuratezza. In pratica, quindi, il principale fattore per determinare la numerosità del campione è il grado di accuratezza che si desidera raggiungere. Come misura del grado di precisione, o accuratezza, che è possibile raggiungere con un dato campione, è usata la *deviazione standard campionaria*, o errore standard, la cui formula, con riferimento al parametro p (una proporzione) e ad un campione di ampiezza *n, è*

$$S = \sqrt{\frac{p.(1-p)}{n}}$$

Questa formula può essere usata, nei casi più semplici, per calcolare l'ampiezza del campione ad una precisione voluta; infatti:

$$n = \frac{p.(1-p)}{S^2}$$

Sebbene la distribuzione campionaria tenda alla distribuzione normale, tuttavia l'adeguatezza di tale approssimazione dipende dal valore di p. Se p=1/2, allora l'approssimazione è buona anche per piccoli campioni; man mano che p tende verso 0, o verso 1, occorrerà invece aumentare l'ampiezza del campione per preservarne una buona adeguatezza. La formula

$$N_{\min} = \frac{2}{3(p.(1-p))^2}$$

fornisce l'ampiezza minima del campione affinché l'approssimazione sia valida. Tuttavia, come già si è detto, anche le caratteristiche della popolazione possono influenzare l'ampiezza del campione: si può ottenere, con un piccolo campione tratto da una popolazione uniforme, lo stesso grado di accuratezza che si ottiene con un grande campione tratto da una popolazione assai varia.

È ancora da notare che l'attendibilità del campione è una funzione della sua ampiezza, indipendentemente dalla proporzione che il campione medesimo ricopre nella popolazione, se quest'ultima è grande (a meno che il campione non costituisca più del 20% della popolazione medesima): un campione di 1.000 individui tratto da una popolazione di 500.000 non risulterà in genere più accurato in modo apprezzabile di un campione di 1.000 individui tratto da una popolazione di 3 milioni! D'altra parte, se è generalmente vero che un campione più ampio risulta più attendibile di uno più piccolo, è altresì vero che l'ammontare di errore sistematico si accresce, viceversa, con l'aumentare dell'ampiezza del campione. Inoltre, il modo stesso in cui è programmato l'esperimento può notevolmente influenzarne la sensibilità senza dover ricorrere ad un campione più numeroso.

Motivi di riflessione a questo proposito ci sono offerti dall'esempio della 'signora del tè'. Abbiamo visto che l'esperimento suggerito da Fisher (4 tazze preparate in un modo, e 4 nell'altro) offre un soddisfacente grado di improbabilità all'ipotesi che la donna possa indovinare per puro caso. Tuttavia, la sensibilità dell'esperimento potrebbe essere ulteriormente accresciuta, non solo, ovviamente, aumentando il numero delle tazze (raddoppiandone il numero, la probabilità di indovinare per caso passerebbe da 1/70 a 1/12.870), ma, ad esempio, ripetendo l'esperimento stesso (in tal caso, mantenendo inalterato il numero delle tazze, la probabilità passerebbe a 1/4.900). Viceversa, suddividendo il campione in due gruppi di differente ampiezza (ad esempio tre tazze preparate in un modo e cinque tazze preparate nell'altro) si otterrebbe una diminuzione di sensibilità e non un aumento (la probabilità che la donna indovini per caso scenderebbe a 1/56).

In generale, un esperimento raggiunge la sua massima sensibilità quando il campione è suddiviso in due gruppi di uguale ampiezza.

Nella programmazione di un esperimento, inoltre, è raro che si possa completamente prescindere dal considerarne anche i relativi costi. Ciò comporterà in genere la ricerca di una strategia di mediazione tra un grado minimo di accuratezza, sotto cui non si è disposti a scendere, ed un costo massimo, che non si riterrà possibile superare: si potrà, per esempio, tener fermo il grado di accuratezza e scegliere quindi il programma sperimentale con il minor costo, oppure, viceversa, tener fermo il costo e scegliere il programma più accurato fra quelli che comportano una uguale spesa.

La programmazione dell'esperimento deve inoltre curare che le condizioni dell'esperimento siano quanto più costanti e uniformi possibile per ogni singolo individuo del campione. Tale richiesta mira ad eliminare dai risultati dell'esperimento influenze diverse da quelle sottoposte alla prova. Nella ricerca medica, dove occorre disporre di un gruppo di 'trattati' e di un gruppo di controllo, è molto importante, non soltanto che il campione nel suo insieme, e ciascun gruppo individualmente preso, siano stati formati con procedura casuale, ma che venga inoltre eliminato ogni elemento esterno di disturbo, come ad esempio l'effetto psicologico connesso al sapere di essere o di non essere tra i pazienti del gruppo sotto trattamento farmacologico. Per questo motivo, al gruppo di controllo è generalmente somministrato un placebo o un altro differente trattamento; si ritiene inoltre consigliabile eliminare anche le eventuali involontarie suggestioni che possono nascere dal comportamento degli stessi medici che seguono la terapia: a tal fine l'esperimento sarà condotto, come si usa dire, in 'doppio cieco', cioè in una situazione nella quale né i pazienti né i medici conoscono la composizione dei due gruppi.

In altri tipi di esperimento è invece preferibile una procedura 'a coppie', opponendo ad ogni individuo di un gruppo un individuo dell'altro, come nell'esperimento di Darwin, già ricordato. È preferibile questo sistema, ad esempio, là dove siano richieste misurazioni di differenze quantitative (differenze in altezza, in peso ecc.). Tale procedimento sembra conciliare bene due requisiti talvolta in conflitto tra loro, quello della massima uniformità del materiale, per cui è accresciuta la sensibilità dell'esperimento, e quello di moltiplicare le osservazioni al fine di rendere i risultati più affidabili e consistenti.

Al fine di confrontare due fertilizzanti agricoli si possono disegnare differenti esperimenti, tutti ragionevoli, ma non ugualmente sensibili ed affidabili. Un primo esperimento potrebbe consistere nel provare un fertilizzante sull'intero terreno per un anno e l'altro fertilizzante nell'anno successivo, e poi confrontare i risultati. Tale procedura ha però lo svantaggio di essere troppo esposta all'influenza di elementi esterni: basti pensare alle possibili differenze climatiche, che possono verificarsi nei due anni, per concludere che tale esperimento potrebbe rivelarsi di scarsa utilità. Più sicuro appare il disegno sperimentale di usare ambedue i fertilizzanti nello stesso anno, trattando metà del terreno con uno di essi e l'altra metà con l'altro. È tuttavia ancora possibile che differenze nella composizione del suolo, nell'esposizione al sole, nell'umidità, ecc., rendano una parte del terreno naturalmente più fertile dell'altra, togliendo così sensibilità all'esperimento. Il miglior procedimento sembra pertanto quello di suddividere il terreno in molte porzioni non grandi, trattando quindi la metà di ogni porzione con un fertilizzante e l'altra metà con l'altro. In tal modo anche le eventuali differenze naturali del terreno vengono ripartite in modo uniforme, accrescendo la sensibilità del test.

Un ultimo esempio può servire a mettere in maggiore evidenza come la programmazione dell'esperimento influisca in maniera decisiva sull'affidabilità dei risultati. Immaginiamo di voler provare la differenza di consumo prodotta dall'aggiunta di un additivo alla normale benzina per auto. A tal fine, potremmo scegliere casualmente un campione di otto vetture, tra loro differenti sia per marca che per cilindrata, così che varie tipologie di auto siano rappresentate, e dividerle quindi, con procedura casuale, in due gruppi di quattro, uno alimentato con benzina più additivo, l'altro con sola benzina.

Tuttavia, essendo le auto tutte differenti per tipo e per cilindrata, la differenza delle medie dei consumi dei due gruppi sarà soggetta a variazioni anche molto sensibili, in conformità con la composizione di essi: se, ad esempio, le auto di grossa cilindrata finissero casualmente nello stesso gruppo, la differenza delle medie dei consumi risulterebbe molto più grande di quanto sarebbe se le auto più potenti compensassero, in ogni gruppo, quelle meno potenti. In ogni caso, un esperimento così programmato non potrebbe evitare che all'eventuale differenza di consumo dovuta all'aggiunta dell'additivo si vada ad aggiungere la differenza dovuta alle caratteristiche intrinseche delle vetture. Un procedimento 'a coppie', che associ tra loro vetture con caratteristiche simili, risulterebbe senz'altro più vantaggioso. L'esperimento guadagnerebbe però ulteriormente in sensibilità se venissero usate le medesime auto per provare ambedue i consumi, sia quello della benzina addizionata con l'additivo sia quello della sola benzina: in tal modo verrebbe eliminata ogni differenza che non sia quella prodotta dall'aggiunta dell'additivo.

Solo un breve cenno, infine, ad esperimenti più complessi.

Nel 1935 Fisher mostrò come fosse possibile progettare anche esperimenti che permettessero di sottoporre a prova simultaneamente differenti fattori. Nacquero così gli esperimenti *fattoriali,* nei quali ogni fattore è confrontato su differenti livelli (differenti livelli di temperatura, o differenti dosi di un farmaco, o per differenti periodi di tempo ecc.). Volendo sperimentare due diversi fattori A e B sui due diversi livelli di assenza e di presenza, si potrebbero ad esempio rilevare i risultati di quattro blocchi sperimentali, nel primo dei quali siano registrati gli effetti del solo fattore A, nel secondo quelli del solo B, nel terzo quelli di A e B insieme e nel quarto gli effetti in assenza di ambedue i fattori. L'insieme di queste procedure è noto come *analisi della varianza,* ed il suo compito è quello di distinguere se i differenti risultati ottenibili variando i fattori in ogni gruppo sperimentale possano essere attribuiti, ed in quale misura, al caso oppure alle variazioni stesse[40].

4.6 Fraintendimenti sui test di significatività. Test di ipotesi

È probabile che a generare i fraintendimenti, di cui sono stati e sono oggetto i test di significatività, abbia contribuito lo stesso Fisher, con la sua critica al teorema di Bayes ed all'inferenza inversa, che potrebbe aver preparato un terreno favorevole alla distorsione del significato di questi test.

Trascinato dalla polemica antibayesiana ed oggettivistica, infatti, Fisher sembra non essersi reso conto della contraddizione esistente tra la funzione peculiare dei test di significatività da lui stesso sostenuta, quella cioè di test volti alla confutazione dell'ipotesi-nulla, e la pretesa che tali test si pongano come metodi alternativi all'approccio bayesiano.

Afferma Fisher: «seguiamo autori come Boole, Venn e Chrystal nel rifiutare l'argomento inverso come privo di fondamento ed inoltre non suscettibile di applicazione consistente», aggiungendo che la dottrina bayesiana dell'inferenza inversa, sebbene sia apparsa a molti «fondamentalmente falsa e priva di fondamento», deve probabilmente la sua lunga sopravvivenza al fatto che, fino a quel momento, non vi era stata a disposizione nessun'altra teoria razionale dell'apprendimento dall'esperienza, atta a sostituirla.

La tecnica di Bayes, assicura Fisher, è ingegnosa:

> I suoi predecessori avevano fornito metodi adeguati per stabilire, data una ben definita popolazione, la probabilità di un particolare tipo di campione. Il suo problema era: dato un particolare tipo di campione, stabilire con quale probabilità un particolare tipo di popolazione potrebbe averlo originato. Egli immagina, di conseguenza, che i possibili tipi di popolazione siano stati essi stessi estratti, come campioni, da una super-popolazione, ed il suo assioma definisce questa super-popolazione con esattezza. Il suo problema diventa così un problema puramente deduttivo al quale possono essere applicati i metodi usuali[41].

[40] L'analisi della varianza può essere considerata il corrispettivo statistico del metodo delle variazioni concomitanti di Mill.

[41] R.A. Fisher, *Uncertain Inference*, «Proc. Of the Am. Ac. Of Arts and Sc.», 71, 1936; ristampato in R.A. Fisher, *Contributions to Mathematical Statistics*, New York 1950, p. 531.

Fisher riconosce a Bayes il merito di aver compiuto il primo serio tentativo di razionalizzazione del ragionamento induttivo mediante una tecnica assai ingegnosa, motivo questo, a suo parere, già sufficiente, da solo, ad assicurargli un posto nella storia della scienza. Tuttavia, a parere di Fisher, Bayes merita un ulteriore ricordo onorevole per il fatto che, «avendo percepito il problema e divisato un assioma che, se ne fosse garantita la verità, collocherebbe l'inferenza inversa nel campo della probabilità matematica, fu sufficientemente critico circa la sua validità da rifiutare di pubblicare il suo saggio finché i suoi dubbi non fossero stati risolti»[42].

Tale assioma, che Fisher contesta, consiste appunto nella distribuzione delle probabilità a priori, che a suo parere introduce un ingiustificabile elemento arbitrario, tale da inficiare l'intero metodo. Sembra a Fisher ancor meno giustificabile questa arbitrarietà per il fatto che nella ricerca scientifica, specialmente nell'interpretazione degli esperimenti, non sembra esservi ragione per fare ricorso all'introduzione di probabilità a priori.

A chiarimento, egli introduce un semplice esempio. Secondo la teoria mendeliana, esistono topi neri di due tipi genetici: omozigote, che, accoppiato con un topo marrone, genera sempre ed esclusivamente topini neri, ed eterozigote, che, accoppiato con un topo marrone, produce una progenie per metà nera e per metà marrone. Dall'accoppiamento di due topi eterozigoti possiamo quindi aspettarci una distribuzione di nascite del seguente tipo: 1 omozigote nero, 2 eterozigoti neri, 1 marrone. Un topo nero nato da due eterozigoti ha perciò 1/3 di probabilità di essere omozigote e 2/3 di essere eterozigote. Se ora, dall'accoppiamento con un topo marrone, un topo nero avrà dato alla luce sette topini tutti neri, avremo una situazione perfettamente analizzabile anche con il metodo bayesiano, essendo note le probabilità a priori di essere monozigote o eterozigote.

La probabilità che un topo nero omozigote dia alla luce sette topi neri è ovviamente pari all'unità; perciò la probabilità dell'evento composto è uguale ad 1/3.

La probabilità che i sette topini siano stati generati da un eterozigote è $(1/2)^7 = 1/128$; perciò la probabilità composta è uguale a $2/3. (1/2)^7 = 1/192$.

È possibile perciò calcolare, con il teorema di Bayes, la probabilità che il topo sia omozigote (BB) o eterozigote (Bb), dopo aver dato alla luce i sette topini neri (E):

$$p(BB/E) = \frac{\frac{1}{3}}{\frac{1}{3} + \frac{1}{192}} = \frac{64}{65} = 0{,}98$$

[42] Fisher non ha invece dubbi sul fatto che il concetto di probabilità di Bayes sottintenda la possibilità di «verifica ad un certo grado di approssimazione attraverso prove ripetute» e che sia perciò equivalente al valore limite della frequenza relativa.

$$p(Bb/E)=\frac{\frac{1}{192}}{\frac{1}{3}+\frac{1}{192}}=\frac{1}{65}=0{,}015$$

Perciò, se lo sperimentatore fosse in grado di sapere che il topo in questione è nato da genitori eterozigoti, «sarebbe disponibile una cogente conoscenza a priori ed il metodo di Bayes potrebbe essere correttamente applicato», nel modo sopra esposto. Ma se questa informazione non fosse disponibile, «nessun sperimentatore si sentirebbe giustificato a ragionare come se conoscesse ciò che di fatto ignora, e, per mancanza di dati adeguati, il metodo di Bayes sarebbe inapplicabile a questo problema»[43].

In assenza di rilevante evidenza genealogica, infatti, ci sono solo due possibilità da considerare: o il nostro topo è omozigote, oppure è eterozigote. In tal caso, il ragionamento bayesiano potrebbe essere applicato soltanto assumendo uguali probabilità a priori per le due alternative, come usa fare quella linea di pensiero che va da Laplace a Jeffreys; ma, per Fisher, ciò significherebbe «trattare il problema, nel quale non abbiamo nessuna evidenza genealogica, esattamente *come se* sapessimo che il topo da sottoporre a test è nato da un accoppiamento che produce metà omozigoti e metà eterozigoti»[44]. Ciò rappresenta per lui un modo di procedere scandalosamente arbitrario ed erroneo; erroneo proprio perché arbitrario: lo scienziato non deve sopperire con assunzioni aprioristiche soggettive alla mancanza di dati oggettivi. Naturalmente, «possono essere trovati, o costruiti, casi nei quali valide probabilità a priori esistono e possono essere dedotte dai dati. Tuttavia, più frequentemente, e specialmente quando siano in questione le probabilità di teorie scientifiche contrapposte, un esame imparziale dei dati a disposizione dello scienziato mostra che nulla del genere può essere asserito»[45].

In conclusione, ci sono per Fisher due possibilità: o si è in possesso di dati che forniscano le probabilità a priori, ed allora è possibile usare il teorema di Bayes, al quale egli riconosce validità e correttezza matematica; oppure non ci sono dati disponibili appropriati, ed allora dobbiamo ricorrere ai test di significatività per confutare l'ipotesi nulla.

In questa prospettiva, i test di significatività vengono evidentemente, ma anche con eccessivo semplicismo, sostenuti da Fisher come metodi *alternativi* per tutti quei casi, che nella scienza sperimentale sono a suo parere la maggioranza, ai quali il teorema di Bayes non può essere applicato per carenza di dati oggettivi. Con ciò Fisher ha implicitamente

[43] R.A. Fisher, *Statistical Methods and Scientific Inference,* cit., p. 20.
[44] *Ibid.*
[45] Ivi, p. 17.

suggerito che la decisione di ricorrere ai test di significatività o al teorema di Bayes dipenderebbe dalla presenza od assenza di dati oggettivi, e non dal contesto e dalla tipologia delle questioni induttive; senonché, proprio le caratteristiche e le condizioni da lui stesso imposte all'ipotesi-nulla, ed i limiti di applicabilità dei test di significatività, avrebbero dovuto portarlo invece ad approfondire il discorso nella seconda direzione.

Non c'è quindi da meravigliarsi se l'ambiguità da Fisher stesso proiettata in tal modo sui suoi test abbia portato altri statistici a svilupparne dei nuovi, sulla base di alcuni fondamentali fraintendimenti dei test di significatività fisheriani.

La natura di questi fraintendimenti potrà essere meglio compresa dopo che saranno stati introdotti i *test di ipotesi*. Tali test furono elaborati da J. Neyman ed E.S. Pearson (il figlio di Karl) e presentati come «perfezionamenti» dei test di significatività e, in genere, dei metodi statistici fisheriani.

L'atteggiamento epistemologico, che sta alla base di questo nuovo approccio induttivo, è facilmente rintracciabile negli scritti di Neyman, ed in particolare in un saggio nel quale egli contrappone al 'ragionamento induttivo' di Fisher il proprio 'comportamento induttivo'[46].

In tale scritto, Neyman, ricollegandosi con ogni evidenza al metodo ipotetico-deduttivo classico, distingue tre momenti nel procedimento induttivo:

1. il momento in cui si visualizza un insieme di ipotesi rilevanti per i fenomeni studiati;
2. il momento della deduzione delle conseguenze osservative dalle ipotesi formulate;
3. il momento dell'accettazione o del rifiuto dell'ipotesi sotto prova, sulla base dei risultati sperimentali.

Ora, proprio in relazione al terzo momento del procedimento induttivo ora enunciato, sembra a Neyman che i metodi fisheriani siano scorretti, o comunque carenti: infatti, essi gli sembrano comportare un passaggio quasi automatico dai risultati sperimentali alla conclusione circa l'ipotesi, come se tra i primi e la seconda sussistesse una sorta di implicazione logica. Per Neyman, invece, il terzo momento del procedimento induttivo rimanda sempre a un 'atto di volontà' che si concretizza in una decisione riguardante un comportamento da adottare: un 'comportamento' induttivo, quindi, che fra l'altro richiede anche che si tenga conto delle sue stesse conseguenze, e non un 'ragionamento' induttivo, come gli sembra essere il procedimento di Fisher.

[46] Cfr. J. Neyman, *"Inductive Behavior" as a Basic Concept of Phylosophy of Science,* «Revue de l'Institut Intern. de Statistique», vol. 25, 1957.

Al proposito, appare abbastanza singolare il fatto che, nel saggio in questione, Neyman parli dei test di significatività come test volti all'accettazione o al rifiuto di una ipotesi e non evidenzi invece quel 'difetto', che il metodo da lui elaborato tenderà ad eliminare, e cioè che i test di significatività di Fisher non contengono criteri per l'accettazione di ipotesi. Inoltre, nessun accenno è fatto alle caratteristiche peculiari che Fisher attribuisce all'ipotesi-nulla[47].

Per introdurre il metodo dei test di ipotesi, possiamo immaginare che un produttore debba periodicamente controllare, mediante campioni casuali, se la sua produzione è sempre sotto controllo, vale a dire se dalla sua fabbrica esce un numero di pezzi difettosi superiore a quello che il mercato sarebbe disposto ad accettare. Il produttore ha davanti a sé due linee di azione possibili:

a) ritenere la produzione fuori controllo e ritirare la merce;
b) ritenere la produzione sotto controllo ed immettere la merce sul mercato.

La scelta tra le azioni *a* e *b* dipenderà dai risultati dell'esperimento; tuttavia il produttore deve anche tener presente la possibilità di commettere due tipi di errore: respingere erroneamente l'ipotesi che la produzione sia tuttora sotto controllo, a causa di un campione accidentalmente più difettoso del dovuto (*errore di prima specie*); accettare erroneamente, viceversa, tale ipotesi ed immettere la merce sul mercato, a causa di un campione casuale assai migliore dell'intera partita (*errore di seconda specie*). È evidente che le conseguenze di tali errori potrebbero risultare piuttosto gravi: una rilevante ed inutile perdita finanziaria nel primo caso, una progressiva perdita di credibilità e quindi di mercato, nel secondo; ma, fra i due, questo secondo errore sarebbe alla lunga il più grave.

Passando ora ad una più dettagliata esposizione, il metodo di Neyman e Pearson può essere sintetizzato come segue:

1. Il risultato dell'esperimento è rappresentato come un punto *s* (punto-campione) in uno spazio-campione *W,* consistente dell'insieme di tutti i campioni possibili.
2. L'ipotesi H_o, che è sottoposta al test, attribuisce una ben determinata distribuzione alla popolazione dalla quale *s* è estratto, ed è tale da determinare la probabilità di ogni campione possibile.

[47] Si potrebbe ipotizzare, a parziale spiegazione di ciò, che Neyman e gli altri abbiano confuso fra loro anche i vari differenti usi dei test di significatività, che Fisher utilizzerà, ma in maniera affatto particolare, anche con riferimento ad alcuni problemi di stima.

3. Si assume che ci sia sempre una ipotesi alternativa determinata.
4. Come risultato dell'esperimento si deciderà:
   a) di rifiutare *Ho;*
   b) di accettare *Ho;*
   c) di sospendere il giudizio (questa eventualità sarà tralasciata).
5. Il test è perciò equivalente alla seguente regola:
   a. rifiuta H_o se *s* cadrà nella regione critica $w_{1.}$
   b. accetta H_o se *s* cadrà nella regione W- w_1.

In relazione alle decisioni *a* e *b si* dovrà tener conto della possibilità di cadere nei due possibili errori seguenti: rifiutare H_o quando è vera (errore di 1° specie); accettare H_o quando è falsa (errore di 2° specie).

In generale, Neyman e Pearson ritengono che gli errori di 1° specie siano tutti equivalenti e che non comportino per lo più gravi conseguenze (rifiutare una ipotesi comporta solo la sospensione del suo uso), mentre gli errori di 2° specie possono avere assai più gravi conseguenze, e tale gravità è proporzionale alla differenza esistente tra l'ipotesi accettata e l'ipotesi vera[48].

In concreto, l'errore di 1° specie, detto anche *livello di significatività*, corrisponderà a $p(w_1/H_0)$, cioè alla probabilità che *s* cada nella zona critica di rifiuto w_1 sotto l'assunzione che H_o sia vera, mentre l'errore di 2° specie corrisponderà a $p(W\text{-}w_1/H_1)$, cioè alla probabilità che *s* cada nella zona di accettazione sotto l'assunto che sia invece vera l'ipotesi alternativa.

La determinazione del test ottimale si identifica perciò con la scelta della zona critica w_1. Neyman e Pearson fanno dipendere tale scelta dalla cosiddetta *potenza del test,* vale a dire da $p(w_1/H_1)$, la probabilità della zona critica sotto l'assunzione della verità dell'ipotesi alternativa H_1.

I due statistici mostrano quindi che si ottiene un test ottimale quando sono resi ottimali ambedue i rischi di errore, e ciò si ottiene scegliendo, quale zona critica w_1, quella zona dello spazio-campione costituita da tutti i punti-campione (ovvero campioni possibili) aventi le più alte probabilità sotto l'assunzione che sia vera H_1 (ciò equivale a minimizzare il rischio di errore di 2° specie).

Facciamo un esempio. Si voglia sottoporre a prova l'ipotesi H_o che afferma che la proporzione di palline nere in un'urna è pari ad 1/4. Sia presa come alternativa H_1 l'ipotesi che attribuisce all'urna i 3/4 di palline nere. L'esperimento consisterà nell'estrazione con ripetizione di un campione casuale di 5 palline. È possibile definire subito lo spazio-campione *W* come l'insieme di tutti i risultati possibili, e cioè:

[48] Cfr. J. Neyman, E.S.Pearson, *The Testing of Statistical Hypotheses in Relation toProbabilities a Priori*, in «Procedings of the Cambridge Philos. Society», XXIX, 1932-33, pp. 492-510.

s_1	5N
s_2	$4N \wedge 1\neg N$
s_3	$3N \wedge 2\neg N$
s_4	$2N \wedge 3\neg N$
$s_{5:}$	$1N \wedge 4\neg N$
s_6	$5\neg N$

Possiamo quindi calcolare la probabilità di ciascuno di questi possibili risultati, subordinatamente a ciascuna delle due ipotesi:

$$p(s_1/H_1) = p(s_6/H_0) = 1\left(\frac{3}{4}\right)^5 = 0{,}23$$

$$p(s_2/H_1) = p(s_5/H_0) = \frac{5!}{4!1!}\left(\frac{3}{4}\right)^4 \frac{1}{4} = 0{,}39$$

$$p(s_3/H_1) = p(s_4/H_0) = \frac{5!}{3!2!}\left(\frac{3}{4}\right)^3\left(\frac{1}{4}\right)^2 = 0{,}26$$

$$p(s_4/H_1) = p(s_3/H_0) = \frac{5!}{2!3!}\left(\frac{3}{4}\right)^2\left(\frac{1}{4}\right)^3 = 0{,}087$$

$$p(s_5/H_1) = p(s_2/H_0) = \frac{5!}{4!1!}\frac{3}{4}\left(\frac{1}{4}\right)^4 = 0{,}014$$

$$p(s_6/H_1) = p(s_1/H_0) = 1\left(\frac{1}{4}\right)^5 = 0{,}0009$$

I risultati con la più alta probabilità, sotto l'ipotesi alternativa H_1, sono i punti-campione s_1, s_2 ed s_3. Se scegliamo l'insieme di questi punti come zona critica w_1 per il rifiuto dell'ipotesi H_0, otterremo una potenza del test pari a $p(s_1 \vee s_2 \vee s_3/H_1) = 0{,}88$ ed avremo reso ottimali i due rischi di errore.

Avremo infatti che $p(w_1/H_0) = p(W\text{-}w_1/H_1) = 0{,}10$: l'errore di 1° specie e l'errore di 2° specie sono stati resi uguali. L'esempio ha ovviamente valore esclusivamente esplicativo perché rischi di errore del 10% sarebbero generalmente ritenuti troppo alti e non accettati; d'altra parte, per ottenere dei rischi di errore più piccoli, occorrerebbe aumentare l'ampiezza del campione.

Sarebbe stato naturalmente possibile anche scegliere una zona critica differente, ad esempio quella compresa dai soli punti campione s_2 ed s_3; in tal caso i due rischi di errore sarebbero risultati diversi fra loro. È da tener presente, infatti, che non è detto che il test ottimale consista sempre nel rendere uguali i due rischi di errore; quando le conseguenze di un tipo di errore siano di gran lunga più disastrose delle altre può eventualmente essere opportuno valutare se non sia il caso di cautelarsi maggiormente nei confronti dell'errore più pericoloso. Ma, in generale, specie quando si tratta di rischi che ricadono su differenti soggetti, come ad esempio nel caso di una partita di merce in cui un rischio ricade sull'acquirente e l'altro sul commerciante, la scelta suggerita sopra appare ottimale.

Infine, si richiama ancora l'attenzione sul fatto, talvolta trascurato, che anche con i test di ipotesi l'ipotesi sotto controllo, quella cioè oggetto di accettazione o di rifiuto, è sempre soltanto una, indicata con h_o, *l'ipotesi-zero*. La ragione di ciò è semplice: i test di ipotesi non richiedono che sia preso in considerazione l'insieme esaustivo delle effettive ipotesi possibili, ma solo ipotesi matematicamente esaustive; perciò l'ipotesi alternativa non è in genere che un'ipotesi idonea a minimizzare l'errore di 2° specie. Pertanto, il rifiuto dell'ipotesi-zero lascia ancora aperte diverse possibilità che potranno eventualmente essere controllate separatamente. Per poter passare dal rifiuto dell'ipotesi-zero all'accettazione dell'ipotesi alternativa occorrerebbe, anche in questo caso, che le due ipotesi considerate costituissero effettivamente, e non in modo fittizio, l'insieme esaustivo delle ipotesi possibili.

4.7 Test di ipotesi e test di significatività a confronto

Se, come dicevamo, Neyman ed altri guardavano ai test di ipotesi come ad una elaborazione migliore, più corretta non solo sul piano teorico, ma anche sul piano tecnico, dei test di significatività elaborati da Fisher, quest'ultimo, dal canto suo, andava gridando allo scandalo e all'equivoco, non peritandosi di gratificare gli avversari di incompetenza, e non soltanto, forse, nel campo delle procedure sperimentali delle scienze empiriche. Ma cediamo la parola allo stesso Fisher:

> È notevole [egli dice parlando dei test di significatività] che gli uomini che sentirono la necessità di questi test e che per primi li concepirono, o che più tardi li resero matematicamente precisi, erano tutti attivamente interessati a ricerche nel campo delle scienze naturali. Ma più recentemente un considerevole corpo dottrinale ha cercato di spiegare, o piuttosto di reinterpretare, questi test sulla base di un modello affatto differente, cioè come mezzi per prendere decisioni in una procedura di accettazione. Le differenze tra queste due situazioni sembrano a chi scrive molte e grandi e non credo che sarebbe stato possibile trascurarle se gli autori di questa reinterpretazione avessero avuto una qualche effettiva familiarità con il lavoro delle scienze naturali, o la consapevolezza di quelle caratteristiche di un rapporto di osservazione che permettono una migliore comprensione scientifica, così come esse sono viste in particolare nell'ottica della programmazione degli esperimenti. I fraintendimenti invero appaiono andare più in profondità di quanto ci sarebbe stato da aspettarsi da un puro trasferimento di tecniche da un campo di studio ad un altro.

Ed aggiunge, forse un po' pesantemente:

> Si può anche sospettare, naturalmente, che quegli autori, come Neyman e Wald, che hanno trattato questi test con poco riguardo per la loro funzione nelle scienze naturali, possano non aver avuto maggior successo nell'applicazione delle loro idee alle necessità delle procedure di accettazione. È tuttavia un evidente vantaggio di entrambi i tipi di applicazione che le teorie sviluppate ed insegnate ai matematici non confondano le loro differenti esigenze[49].

Fisher ha senz'altro ragione nel denunciare un profondo equivoco alla base della reinterpretazione dei test di significatività da parte di Neyman e della scuola americana in genere; ed ha ragione, in particolare, quando collega questo equivoco con le differenti esigenze di differenti campi di applicazione, sebbene egli manchi poi di cogliere l'effettiva sorgente del fraintendimento.

Se guardiamo alla struttura metodologica cui Neyman intende riferirsi, quella ipotetico-deduttiva, appare chiaro che per lui il procedimento induttivo si configura come una decisione da prendere sulla base di certi risultati empirici, ed inizia perciò soltanto *dopo* che certe ben determinate ipotesi siano state proposte; il procedimento induttivo da lui contemplato consiste, insomma, in un procedimento di prova che si colloca in un contesto *giustificativo* per il comportamento, di accettazione o di rifiuto, che verrà successivamente assunto nei confronti delle ipotesi considerate. È altresì evidente, dalle critiche che egli muove a Fisher, che egli ritiene che anche i test di significatività rientrino nello stesso contesto, in riferimento al quale essi si mostrerebbero, appunto, carenti. Pertanto, non solo egli è naturalmente indotto ad interpretare i test di significatività come metodi per l'accettazione di ipotesi, ma non pone neppure sufficiente attenzione alle peculiari caratteristiche dell'ipotesi-nulla di Fisher, che con l'ipotesi h_o dei test di ipotesi non ha nulla da spartire[50].

[49] R.A. Fisher, *Statistical Methods*, cit., pp. 76 sgg.

[50] I fraintendimenti relativi ai test di significatività della scuola statistica americana hanno finito per imporsi, con effetti spesso confusi. Esemplare, sotto il profilo di una macroscopica ignoranza dell'origine del termine «ipotesi-nulla», e presumibilmente dell'intero lavoro statistico di Fisher, mi sembra W. Hays (*Statistics*, 5°ed., London, 1994, pp. 269 sgg.), il quale, cercando di spiegare il significato del termine «ipotesi-nulla» (da lui impropriamente riferito anche all'ipotesi zero dei test di ipotesi), afferma che tale termine sta solo ad indicare «l'ipotesi che è assunta come vera nel generare la distribuzione campionaria usata nel test»; aggiunge quindi, dissentendo da chi pensa che tale termine derivi dal fatto «che nel lavoro sperimentale il valore del parametro specificato in H_0 è spesso 0» (posizione che si avvicina al senso originario, perché l'ipotesi casuale postula *sempre* che le differenze rilevate siano zero), che tale ipotesi «può specificare ogni valore possibile per uno o più parametri e questo uso della parola *nulla* è solo fortuito». Anche Howson e Urbach mostrano

Sfugge del tutto a Neyman che la funzione peculiare dei test di significatività nella ricerca scientifica (all'interno, in particolare, di quella concezione positivistica della scienza che Fisher ha ereditato tanto da Darwin quanto da K. Pearson) sta proprio nel loro configurarsi come metodi di «scoperta» nel processo di eliminazione dei fenomeni accidentali dal campo di studio.

I test di significatività, dice Fisher, sono costantemente usati nella ricerca sperimentale per «distinguere effetti di reale importanza per un programma di ricerca» da effetti apparentemente tali, ma che potrebbero risultare da errori di campionamento casuale, o da variabilità incontrollate, di qualsiasi tipo, nel materiale fisico o biologico sotto esame; le conclusioni tratte da tali test «costituiscono i passi con cui l'operatore scientifico guadagna una migliore comprensione del suo materiale sperimentale e dei problemi che esso presenta»[51]. I test di significatività mirano all'analisi dei dati sperimentali, non all'accettazione di una ipotesi; perciò, mentre le decisioni raggiunte con le procedure di accettazione sono definitive, lo stato di opinione che deriva dai test di significatività è provvisorio e sempre suscettibile di ulteriori revisioni o approfondimenti.

Fisher ha ragione, ripeto, nel sottolineare le differenze applicative e logiche tra test di significatività e test di ipotesi. In particolare, gli sembrano sufficienti a caratterizzare tali differenze due aspetti dei test di significatività, che sono appunto quelli su cui ci siamo già diffusamente soffermati: il loro essere rivolti esclusivamente al rifiuto dell'ipotesi-nulla e le caratteristiche affatto peculiari di questa ipotesi. Pur riconoscendo alle procedure di accettazione una utile funzione nel mondo industriale, Fisher difende le peculiarità di applicazione dei propri test di significatività. Per esempio, gli appare importante che, contrariamente a quanto richiesto dai test di ipotesi, i test di significatività non contemplino l'introduzione di alcuna funzione di costo, in relazione a decisioni errate. Fare così, egli dice, implicherebbe che gli scopi, per i quali si ricercano nuove conoscenze, erano già noti ed in grado di essere valutati. Gli appare invece auspicabile che la ricerca scientifica sia sempre aperta a nuove scoperte e che i risultati man mano raggiunti servano ad illuminare altre menti di ricercatori e che possano, prima o poi, servire per un numero di scopi dei quali non si possa, al momento, conoscere nulla. Il contributo al progresso della conoscenza naturale, che la ricerca scientifica può fornire, è disseminato dalla speranza e dalla fiducia che, con il crescere della conoscenza stessa,

altrettanta ignoranza in proposito: la teoria dei test di significatività di Fisher, essi scrivono, «prescrive come le ipotesi statistiche dovrebbero essere testate». Per loro, l'ipotesi-nulla di Fisher indica semplicemente una qualunque ipotesi statistica sotto prova (*Scientific Reasoning*, cit., p. 8).

[51] R.A. Fisher, *Statistical Methods,* cit., p. 76.

«verrà facilitata una gran varietà di scopi, per mezzo di una gran varietà di uomini, e di gruppi di uomini»[52]

Tornando alle differenze tra le ipotesi sotto prova, nei due metodi, preme richiamare l'attenzione sul fatto che l'ipotesi h_o dei test di ipotesi (anch'essa purtroppo spesso chiamata ipotesi-nulla, ma che, per evitare confusioni, indicheremo qui sempre con *ipotesi-zero*) non è mai l'ipotesi casuale di Fisher, ma è sempre, allo stesso modo della sua alternativa, una ipotesi parametrica relativa alla distribuzione di qualche caratteristica in una popolazione determinata.

Semplificando, la situazione considerata da Neyman sembra essere del tipo seguente: si vuol sapere se l'ipotesi che un certo parametro di una popolazione sia r (tale parametro può esprimere una frequenza, una media, una varianza ecc.), sia da accettare o da rifiutare. Per procedere alla prova ci si serve di una ipotesi alternativa la cui unica funzione è quella di intervenire nella determinazione dell'errore di 2° specie, errore che non ha senso determinare nei test di significatività, nei quali non si pone alcun problema di accettazione.

Si deve anche porre attenzione al fatto che, sebbene l'errore di 1° specie sia chiamato anche «livello di significatività» e sia spesso assunto nella pratica come livello standard, esso non coincide affatto con il livello di significatività dei test fisheriani di significatività, con il quale condivide soltanto il senso generico di valore critico di probabilità al di sotto del quale l'ipotesi sotto prova viene rifiutata: in un caso infatti esso riguarda verosimiglianze (*likelihoods*) di campioni possibili, nell'altro si riferisce a differenze.

Le inferenze relative alla distribuzione ipotetica di una popolazione sono per lo più trattate da Fisher, invece, come problemi di *stima,* e quindi affrontate con altre strategie *(maximum likelihood,* argomento fiduciale, ecc.), alle quali accenneremo successivamente.

In breve, i test di ipotesi si differenziano dai test di significatività per essere metodi di accettazione o di rifiuto di una ben determinata ipotesi riguardante un parametro di una popolazione, laddove i test di significatività sono esclusivamente volti al rifiuto dell'ipotesi che asserisce che le differenze riscontrate tra i risultati sperimentali sono dovute al caso.

La distribuzione casuale, naturalmente, può anch'essa comparire come ipotesi parametrica relativa ad una popolazione (ad esempio, la popolazione dei lanci di un dado, o dei lanci di una moneta) e come tale può essere soggetta ai test di ipotesi; ma in tal caso essa non vi compare con lo stesso significato che essa ha nei test di significatività. In generale, ovunque si tratti di valutare dati empirici sotto il profilo della loro rilevanza scientifica, oppure laddove una ipotesi non sia esprimibile altro che in termini

[52] Ivi, p. 104.

qualitativi, indeterminati o indeterminabili quantitativamente (ad esempio, «il farmaco A è più efficace del farmaco B»), i test di ipotesi di Neyman e Pearson non sembrano applicabili a meno di evidenti ed opinabili forzature, e la questione non può essere impostata che come problema di eliminazione dell'ipotesi casuale. Rientra in questo tipo, ad esempio, la maggior parte delle ricerche in campo clinico, farmacologico ed epidemiologico.

Di fatto, la situazione appare oggi alquanto confusa: test di significatività e test di ipotesi sono, sotto l'influenza della scuola statistica americana, generalmente fusi insieme (chiamati spesso indifferentemente test di significatività o test di ipotesi) e interpretati come il procedimento per confrontare due ipotesi alla luce dell'evidenza, al fine di accettare o rifiutare una qualche ipotesi indicata con il simbolo h_o.

L'applicazione dei test di ipotesi a situazioni di tipo qualitativo (del tipo, «il farmaco A è efficace nella patologia F») è resa possibile mediante il trucco di trasformare l'ipotesi in oggetto in una coppia di ipotesi alternative parametriche. Laddove i test di significatività si confrontano con l'ipotesi di irrilevanza scientifica dei risultati sperimentali (l'ipotesi, cioè, che le differenze rilevate nei campioni siano dovute alle normali oscillazioni casuali che si hanno nelle operazioni di campionamento da una *stessa* popolazione), al fine di confutarla, i test di ipotesi mirano viceversa ad accettare o rifiutare l'ipotesi che i valori dei parametri P_1 e P_2, nelle *due* diverse popolazioni, dalle quali sono stati estratti i campioni, siano uguali, contro l'ipotesi che siano differenti. In pratica è come se si assumessero due urne, l'una contenente i malati della patologia F trattati con il farmaco A, l'altra contenente i malati della stessa patologia non trattati o trattati diversamente, e si ipotizzasse che in ciascuna delle due urne vi sia una certa percentuale di individui con la caratteristica di essere «guariti o sensibilmente migliorati» (naturalmente questa espressione richiederebbe di essere resa più precisa). L'ipotesi da accettare o rifiutare è ora quella che le due urne contengano una identica percentuale di miglioramenti ($P_1=P_2$), mentre l'ipotesi alternativa affermerà che tale percentuale è diversa nelle due urne.

Appare evidente quanto questo artificio snaturi il problema originario ed il significato dei test di significatività.

4.8 Analisi di una scheda sperimentale

Proviamo ora ad esaminare criticamente una scheda di rapporto sperimentale che costituisce un esempio, fra i più dettagliati ed accurati fra quanti mi sia capitato di vedere, dei resoconti sperimentali forniti dalla letteratura medica corrente a tutti disponibile[53]. Dopo aver ripor-

[53] La scheda è tratta da R. Marchioli, G. Tognoli, *Cause-effetti in medicina*, Roma 1994, p. 21.

tato fedelmente la scheda, ne daremo, in primo luogo, i risultati secondo l'interpretazione dei test di significatività in senso proprio, per procedere poi ad alcune osservazioni.

Scheda

Premessa I : *Nel post-infarto molti pazienti muoiono per aritmia ventricolare.*

Premessa II: I *farmaci antiaritmici riducono le aritmie ventricolari.*

Ipotesi: I *farmaci antiaritmici migliorano la prognosi del post-infarto.*

Nei trials CAST 1 e 2 (Cardiac Arrhythmia Suppression Trial) circa 1.500 pazienti con infarto miocardico nei 6-24 mesi precedenti, con aritmie ventricolari (asintomatiche o solo minimamente sintomatiche) sensibili alla terapia antiaritmica (in base all'Holter), sono stati randomizzati ad un trattamento antiaritmico con farmaci della classe 1C (flecainide o encainide o moricizina) o a nessun trattamento antiaritmico. Entrambi gli studi sono stati chiusi prima della conclusione a causa di un eccesso di mortalità nei pazienti assegnati al trattamento antiaritmico.

Risultati del CAST 1:

	I		II		III	
	Encaimide	Placebo	Flecainide	Placebo	Entrambi	Placebo
Pazienti randomizzati	432	425	323	318	755	743
Mortalità totale Arrestocardiaco	44	19	19	7	63	26

Risultati del CAST 2:

	I (risultati a 2 settimane)		II risultati a18 mesi di *follow-up*	
	Moricizina	Placebo	Moricizina	Placebo
Pazienti randomizzati	665	660	691	683
Mortalità totale o Arresto cardiaco	17	3	104	86

Non solo i risultati degli studi CAST hanno falsificato l'ipotesi iniziale, ma si è ottenuto un effetto opposto a quanto ci si attendeva in base alla 'razionalità' delle ipotesi iniziali.

Termine della scheda

Risultati con i test di significatività

Ipotesi-nulla: nessuna differenza significativa è rilevabile tra i due gruppi.

Test utilizzato: χ^2 a due code.

(Il ragionamento, per calcolare i valori teorici, è analogo a quello esemplificato a proposito dei casi di colera riscontrati a Calcutta).

CAST 1, I : per approssimazione, χ^2 =9,82 ed 1 grado di libertà. L'ipotesi-nulla è respinta ad un livello di significatività inferiore allo 0,002.

CAST 1, II: approssimativamente χ^2= 5,75 ed 1 grado di libertà. L'ipotesi-nulla è respinta ad un livello di significatività di poco più alto dello 0,01.

CAST 1, III: χ^2= 15,41 per approssimazione, ed 1 grado di libertà. L'ipotesi-nulla è rifiutata ad un livello di significatività inferiore allo 0,001. Notiamo che CAST 1 III non è che la meta-analisi dei primi due, cioè è soltanto un esperimento teorico costruito sommando insieme i dati di CAST 1 e CAST 2. Pertanto non poteva che riconfermare i risultati precedenti.

CAST 2, I: χ^2= 9,95 per approssimazione, ed 1 grado di libertà. L'ipotesi-nulla è rifiutata al livello di significatività decisamente inferiore allo 0,002.

CAST 2, II : χ^2= 1,6 per approssimazione, ed 1 grado di libertà. In questo caso l'ipotesi-nulla non può essere rifiutata: tra i due gruppi le differenze non sono significative.

Il risultato di CAST 2, II non può mancare di sorprendere, se confrontato con CAST 2, I: a più lungo termine, i risultati sono passati da fortemente significativi contro l'ipotesi-nulla a non significativi.

Osservazioni

I risultati sperimentali di CAST 1, I e II, e CAST 2, I sono altamente significativi contro l'ipotesi-nulla, vale a dire che le differenze riscontrate nei due gruppi non possono essere ritenute casuali; inoltre mostrano che tale significatività va nella direzione opposta a quella attesa. Ciò non implica tuttavia la conclusione, che ne è stata tratta, che «i risultati [...] hanno falsificato l'ipotesi iniziale»[54], – ipotesi che, nella scheda, afferma che «i farmaci antiaritmici migliorano la prognosi nel post-infarto» –, perché questa non può essere l'ipotesi sotto prova nei test di significati-

[54] Ivi.

vità e nessuna implicazione logica diretta esiste tra essa e l'ipotesi-nulla. Il test pertanto può dare solo indicazioni assai indirette circa l'effetto dei farmaci e non può neppure escludere che le differenze riscontrate siano dovute ad una differente causa ignorata.

La scheda, in realtà, non offre maggiori indicazioni di quelle sopra riportate circa il tipo di test utilizzato, ed è probabile che si tratti di un test di significatività interpolato, come è divenuto frequente soprattutto nell'ambito delle sperimentazioni cliniche e farmacologiche[55]. Resta comunque il fatto che 'l'ipotesi sotto prova' difficilmente può coincidere con quella espressa nella scheda, la quale oltretutto è troppo indeterminata e vaga per dar luogo ad un test statistico ben programmato. Il fraintendimento circa l'ipotesi sulla quale il test è chiamato ad esprimersi può, ovviamente, compromettere la portata dell'esperimento e la corretta interpretazione dei risultati.

I dati della scheda, nel loro complesso, suggerirebbero l'opportunità di riconsiderare più criticamente sia l'esperimento che l'ipotesi. Relativamente all'esperimento, è sempre possibile che non vengano osservate tutte le cautele connesse con la randomizzazione. Il clinico, peraltro, si trova spesso in conflitto tra le esigenze della sperimentazione e la propria etica professionale, e può accadere che faccia in modo da includere i malati più gravi nel gruppo da trattare, quando nutra fondate speranze che il trattamento abbia un effetto positivo. Ciò, naturalmente, può alterare considerevolmente i risultati.

Relativamente all'ipotesi, possiamo in primo luogo cercare di rendere rigorosa la struttura logica del ragionamento, che sembrerebbe riducibile al seguente schema (dove *Arv* = aritmia ventricolare; *Mp* = morte nel post-infarto; *An* = antiaritmici; $\prec$ sta per «implica spesso ma non sempre»; > sta a significare «diminuzione numerica»; $\rightarrow$ sta per «implica»):

$$
\begin{aligned}
&Arv \prec Mp\\
&An \rightarrow\ > Arv\\
&An \rightarrow\ >Mp
\end{aligned}
$$

Così formalizzata, l'argomentazione appare corretta; tuttavia le espressioni usate nella scheda appaiono tutte di una vaghezza eccessiva, che suona al logico alquanto ambigua. Consideriamo la prima premessa, «nel post-infarto *molti* pazienti muoiono *per* aritmia ventricolare»: non è chiaro se *molti* sia un sottoinsieme della classe dei pazienti deceduti o coincida con tale classe, né è chiaro se «*per*» voglia indicare un nesso

[55] Al proposito, uno studio pubblicato nel 2004 su *BMC Medical Research Methodology* da Emili Garcia-Berthou e Carles Alcaraz, dell'università di Girona (Spagna), rilevava che buona parte degli articoli scientifici sulle maggiori riviste mediche contiene errori statistici, anche di grado tale da falsare i risultati.

causale o una mera correlazione. La seconda premessa, inoltre, non precisa se ad essa sia da attribuire un valore universale (se, cioè, l'effetto positivo degli antiaritmici sia da attendersi *sempre*), oppure soltanto in una (quanto larga?) percentuale dei casi. Infine, dato il grado solo parziale di implicazione esistente tra *Arv* e *Mp*, è evidente che non vi sarebbe comunque da attendersi, come effetto di *An,* una diminuzione di uguale intensità tra *Arv* e *Mp*.

Analoghi rilievi di incompletezza e di ambiguità possono essere rivolti anche alle espressioni riportanti i risultati degli esperimenti: ad esempio, la mortalità e/o l'arresto cardiaco rilevati possono essere connessi *causalmente* con l'aritmia ventricolare, o soltanto correlati? Mi permetto di sottolineare ancora una volta, limitatamente agli aspetti logici, e non a quelli clinici, della questione, che risultati sperimentali significativi contro l'ipotesi-nulla possono non essere affatto rilevanti a proposito di ipotesi riguardanti i farmaci usati nel trattamento: vale a dire che, pur dovendosi rifiutare l'ipotesi-nulla, i farmaci antiaritmici potrebbero non essere responsabili degli esiti infausti.

Un clamoroso esempio di erronea interpretazione dei test di significatività fu costituito dal caso della Tolbutamide, che provocò un vero pandemonio vari decenni addietro. Questo farmaco, usato da tempo nella terapia del diabete, risultò significativamente correlato, nel corso di un test clinico sui vari regimi antidiabetici, con un alto tasso di mortalità, per cause cardiovascolari, riscontrato nel gruppo trattato, rispetto al gruppo di controllo, cui era stato somministrato un placebo. Anziché limitarsi a rifiutare l'ipotesi che tale correlazione potesse essere casuale, e cercarne una spiegazione nella sede appropriata, si passò immediatamente alla conclusione che la Tolbutamide causasse disturbi, con esito infausto, di natura cardiovascolare; conclusione che si dovette in seguito correggere.

4.9 Stima

Tra le questioni di inferenza inversa, che è opportuno tener separate dalle precedenti, si pongono i problemi di *stima*. Tali questioni, invero, non sempre sono tenute distinte dalle questioni di accettazione, ed in effetti i test di ipotesi di Neyman-Pearson possono essere considerati, in senso lato, anche metodi di stima, perché l'ipotesi che essi sono chiamati a controllare, per accettarla o respingerla, esprime appunto una stima di un valore ignoto di un parametro di una popolazione considerata. Anche il teorema di Bayes può, naturalmente, essere utilizzato per stimare un parametro: il valore del parametro sarà quello espresso dall'ipotesi che risulterà più probabile alla luce dei dati sperimentali.

Tuttavia, la formulazione di un problema di stima suona sempre in modo diverso dalla formulazione di una questione di accettazione, e la logica che guida i due metodi non può che essere differente. In un procedimento di accettazione, la questione è se l'ipotesi, che «il valore della grandezza non nota *x* è *r*», sia conforme all'esperienza e quindi accettabile: perciò, la formulazione di un insieme di ipotesi alternative circa il parametro ignoto della popolazione precede e determina la raccolta delle informazioni o l'esecuzione dell'esperimento. Un problema di stima si esprime, invece, come ricerca diretta del valore della grandezza ignota che ci interessa; pertanto in esso non si perviene a valutazioni di probabilità.

Anche i test di significatività svolgono una loro funzione nei problemi di stima. Essi svolgono, per così dire, un ruolo propedeutico al procedimento di stima. Per loro mezzo viene messa alla prova la compatibilità di certi valori teorici (ad esempio, valori attesi o, appunto, stime di un qualche parametro) con i dati sperimentali. Si controlla, cioè, in primo luogo, se certe caratteristiche trovate possono essere accidentali. Se l'ipotesi-nulla non può essere rifiutata, allora nessuna ipotesi di stima valida può essere avanzata. Viceversa, se l'ipotesi nulla viene rifiutata, allora si passerà alla stima del parametro ignoto, con una delle procedure di stima: argomento fiduciale, massima verosimiglianza, intervalli di confidenza. L'argomento fiduciale, dovuto a Fisher, può essere usato soltanto in mancanza di informazioni circa eventuali probabilità a priori.

Le stime si differenziano tra loro a seconda che riguardino una grandezza continua (ad esempio una temperatura) o una grandezza discreta (ad esempio il numero di abitanti di una certa città): si parla rispettivamente di stime intervallari e di stime puntuali. Nel primo caso la stima riguarda un intervallo entro il quale il valore effettivo della grandezza che si vuol stimare cade con alta probabilità (si assumono in genere probabilità standard del 95% o del 99%). Nel caso di stime intervallari si può ricorrere al metodo degli *intervalli fiduciali* di Fisher oppure al corrispondente metodo degli *intervalli di confidenza* di Neyman e Pearson.

Per quanto concerne le stime puntuali, è sempre di Fisher il metodo della *massima verosimiglianza (maximum likelibood),* che consiste, in breve, nel preferire come stima della grandezza ignota quel valore, sotto l'assunzione del quale i risultati osservati hanno la massima probabilità di verificarsi. Siano state estratte, ad esempio, con reimbussolamento, due palline da un'urna contenente esattamente 4 palline, e siano, le due palline estratte, una bianca ed una nera; si voglia quindi stimare quante palline nere siano contenute nell'urna. Per far questo, dovremo semplicemente calcolare tutti i valori $p(e/r)$, cioè le probabilità dell'evento «estrazione di 1 pallina bianca e di 1 nera», subordinatamente a tutti i possibili valori *r* di palline nere nell'urna, che, nella fattispecie, sono solo tre: 1, 2, 3. Eseguiti i calcoli, vedremo che la stima di 2 palline nere nell'urna risulta l'assunzione che dà alle due estrazioni eseguite il massimo valore di *likelihood*: p= 0,25 di contro a p= 0,18 degli altri due valori possibili di *r*.

Sebbene non si faccia ricorso, con tale metodo, ad alcuna probabilità a priori, è evidente la stretta connessione esistente tra metodo della massima verosimiglianza ed il teorema di Bayes: infatti quest'ultimo porterebbe alla stessa stima, sotto l'assunzione di equiprobabilità a priori per le tre ipotesi alternative. Ciò è vero in generale.

Vi sono varie proprietà che gli stimatori (cioè le funzioni di stima) debbono soddisfare. Essi devono essere:

- *consistenti:* uno stimatore T è detto consistente se al crescere di *n* (numero delle osservazioni) tende al vero valore θ, mentre, di conseguenza, la sua varianza tende ad annullarsi; cioè: $E(T) \rightarrow \theta$ e $Var(T) \rightarrow 0$, per $n \rightarrow \infty$. Classico esempio di stimatore consistente è il teorema di Bernoulli;
- *corretti (unbiased):* uno stimatore T è detto corretto quando, per qualunque *n,* tende al vero valore θ;
- *efficienti:* uno stimatore T è detto efficiente se, tra tutti gli stimatori corretti, tende ad avere la minima varianza al crescere di *n;*
- *sufficienti:* è detto sufficiente uno stimatore T se riassume tutta l'informazione rilevante contenuta nel campione ai fini della stima di θ. Un esempio di stimatore sufficiente è dato dal numero di successi in *n* prove bernoulliane (per esempio il numero di teste in *n* lanci di moneta).

Il metodo della massima verosimiglianza offre degli stimatori che soddisfano tutte le proprietà suddette, ed in particolare quella della sufficienza, sulla quale proprio Fisher ebbe a richiamare l'attenzione.

Il metodo degli *intervalli di confidenza,* usato per stimare un parametro ignoto di una popolazione, consiste nell'assumere un intervallo, nel quale il valore del parametro da stimare sia compreso con un grado di confidenza prestabilito. Sia, ad esempio, il parametro da stimare una proporzione di una caratteristica di una popolazione, e sia f la proporzione riscontrata in un campione di ampiezza n. Se il campione è sufficientemente ampio, la sua distribuzione si approssima alla normale, cosicché l'intervallo di confidenza può essere facilmente ottenuto dalla formula $\boldsymbol{f} \pm z.\sigma$, dove z è un valore connesso con il grado di confidenza prescelto e σ è la deviazione standard dalla proporzione nella popolazione. Poiché, in genere, il valore di σ non è conosciuto, è possibile usare come sua stima la corrispondente deviazione standard campionaria s, che può essere calcolata dai dati del campione:

$$s = \sqrt{\frac{f.(1-f)}{n-1}}$$

Quanto a z, occorre tener presente che il metodo degli intervalli di confidenza è strettamente connesso con i test di ipotesi e con il rischio di errore di 1° specie (che molti chiamano anche livello di significatività e che è spesso un valore standard). Un intervallo di confidenza al 95% è perciò associato ad un rischio di errore di 1° specie di 0,05, a cui corrisponde $z = 1{,}96$. Un intervallo al 99% è associato ad un rischio di errore di 0,01, con $z = 2{,}58$. In generale, un intervallo sarà detto essere al $100\cdot(1-\alpha)\%$ di confidenza (dove α indica il prescelto rischio di errore). Naturalmente, ogni campione darà luogo ad un differente intervallo. Il grado di confidenza indica in quale proporzione ci si aspetta che il vero valore del parametro sia compreso nell'intervallo: se l'intervallo è al 95%, ad esempio, ci si aspetta che in 95 volte, su 100 campionamenti, il vero valore del parametro cada nell'intervallo dato.

Gli intervalli di confidenza, nati per stimare grandezze continue, sono sempre più frequentemente usati anche per stime puntuali, quando si ritenga sufficiente un'approssimazione al valore da stimare. Inoltre, sembra vada estendendosi la tendenza (opposta a quella di Neyman-Pearson, i quali tendevano a risolvere le questioni di stima, specie se puntuale, in questioni di accettazione) a tradurre in termini di intervalli di confidenza molte questioni di accettazione di ipotesi. Quest'ultima tendenza abbraccia anche problemi di tipo 'qualitativo', già da tempo discutibilmente sottratti ai test di significatività, i quali vengono affrontati con una procedura finora inconsueta.

Cercherò di spiegare tale procedura con un esempio: sia il nostro test connesso con un campione suddiviso in due gruppi, il primo, n_1, trattato con il farmaco A, il secondo, n_2, con un placebo, e si abbiano, come risultati, s_1/n_1 casi di successo nel gruppo trattato con il farmaco e s_2/n_2 casi di successo nel gruppo di controllo. Se poniamo $p_1=s_1/n_1$ e $p_2=s_2/n_2$, possiamo operare una serie di confronti tra vari parametri dei due gruppi, anche in considerazione del fatto che, qualunque sia la distribuzione della popolazione, le distribuzioni campionarie tendono sempre alla curva normale, se il campione è abbastanza grande.

A questo punto, si possono 'stimare' gli intervalli di confidenza entro cui siano compresi (con probabilità standard del 95%, o del 99%) i valori seguenti: $D=(p_1-p_2)$; $R=p_1/p_2$ (detto *rischio relativo);* $\Omega=(p_1/1-p_1)/(p_2/1-p_2)$ *(odds ratio).* I sostenitori di questo procedimento pretendono che tali intervalli di confidenza forniscano maggiore informazione dei test di significatività, pur essendo consistenti con questi. Ma quali sono queste informazioni? I rapporti di test sperimentali, che si trovano talora disponibili in varie trattazioni, consentono un cauto scetticismo circa la perspicuità, la correttezza e la fertilità dei risultati ottenuti. Del resto, parametri quali R e Ω non sono del tutto intuitivi: cosa effettivamente stiamo cercando di valutare quando procediamo alla loro stima? D, che è il più perspicuo, indica evidentemente di quanto p_1 e p_2 differiscano fra loro, o meglio di quanto la loro differenza si allontani da zero; R indica di quante volte p_1 è maggiore, o minore, di p_2; Ω, infine, senz'altro il meno intuitivo, rappresenta il quoziente dei rapporti successi/insuccessi nei due gruppi, ed indica di quante volte uno di essi è maggiore dell'altro.

D'altra parte, è anche evidente che l'interpretazione di una espressione come $r_i \leq \Omega \leq r_j$ è legata da un lato al significato di Ω, ma dipende, dall'altro, dalla conoscenza delle caratteristiche matematiche e di comportamento di tale parametro, dei limiti entro cui è compresa la sua variabilità, e così via. L'intervallo di variabilità di D è, ad esempio, differente da quello di R e di Ω: mentre D varia tra +1 e -1, R e Ω variano tra 0 e ∞, di modo che l'uguaglianza $p_1=p_2$ è espressa da D=0, ma da R=1 e Ω=1. Ciò significa che un intervallo di confidenza contenente lo 0, per D, e 1, per R e per Ω, indica che l'ipotesi-nulla non può essere respinta. Un esempio concreto permetterà di comprendere meglio la procedura e le osservazioni successive.

Operiamo un confronto tra trattamenti sulla base dei seguenti dati campionari: 78 pazienti sono stati trattati con il farmaco A e 68 di essi ne hanno riportato una perdita di memoria; un altro campione di 65 pazienti è stato trattato con il farmaco B e 41 di essi hanno riportato la stessa conseguenza. Applicando un test di significatività, si ottiene χ^2=11,21 e 1 grado di libertà: l'ipotesi-nulla è respinta al livello di significatività 0,002.

Poniamo ora p_1= 68/78 =0,87 e p_2= 41/65 =0,63. Procediamo a valutare gli intervalli di confidenza per D e per R; tralasceremo invece il calcolo per l'intervallo di Ω, a causa della sua eccessiva lunghezza: D= 0,87–0,63 =0,24. Calcoliamo ora la deviazione standard campionaria *s* relativa a D:

$$s_D = \sqrt{\frac{0,87 \cdot (81 - 0,87)}{78} + \frac{0,63 \cdot (1 - 0,63)}{65}} = 0,071$$

L'intervallo di confidenza al 95%, per D, sarà quindi dato da 0,24 ± 1,96·0,071, che dà un intervallo compreso fra 0,38 e 0,10. Possiamo cioè aspettarci dal 10% al 38% in più di perdite di memoria fra i trattati con il farmaco A rispetto a quelli trattati con B. Poiché l'intervallo non include lo zero, il risultato è consistente con il test di significatività.

Veniamo ora a calcolare l'intervallo per R. Poiché R= p_1/p_2, otteniamo R=1,38. Dal momento che abbiamo già ricavato il valore della deviazione standard campionaria, possiamo passare subito al calcolo dell'intervallo di confidenza, che, al 95%, sarà dato da 1,38 ±1,96·0,071, che porterà ad un intervallo, sempre al 95%, compreso fra 1,24 e 1,5. Ci sarebbe, cioè, un rischio relativo, di perdita di memoria, maggiore, da 1,2 a 1,5 volte, per i trattati con A, rispetto ai trattati con B. Anche in questo caso, poiché l'intervallo non comprende 1, il risultato è consistente con il test di significatività.

Non si può tuttavia sfuggire ad una sensazione di vacuità, simile a quella destata da un affascinante, ma sterile, virtuosismo non finalizzato che a se stesso, davanti a simili procedure. La tendenza a trasformare questioni di significatività contro l'ipotesi-nulla in questioni di stima di grandezze molto artificiose non comporta probabilmente un guadagno, ma piuttosto un rischio ulteriore di confusione e forse di errore. Anche perché un esperimento, che dovrebbe essere significativo soltanto con riferimento all'ipotesi-nulla, viene letto immediatamente, come risulta evidente nell'esempio ora descritto, in chiave di causa-effetto.

Il fatto è che si può stimare tutto ciò che si vuole, frequenze e medie, differenze di medie e differenze di probabilità, quozienti di probabilità e quozienti di quozienti di probabilità, e via dicendo, e la statistica matematica non manca di idonei strumenti, talora assai sofisticati. Occorre tuttavia chiedersi se non si rischia di perdere il senso dei problemi, ed il significato dei risultati, nell'applicazione di un apparato forse inutilmente sovrabbondante. Il clinico, lo scienziato, l'operatore di laboratorio non sempre conoscono a fondo la portata ed il significato delle formule che usano; eppure, la loro analisi dei dati non può prescindere da questa conoscenza; altrimenti, le loro interpretazioni dei risultati sperimentali dovranno essere accolte con molta cautela. Lo statistico, d'altra parte, non conosce la materia specifica cui i metodi statistici sono stati applicati e la sua interpretazione andrebbe parimenti accolta con molta cautela. Lo statistico, inoltre, spesso, come abbiamo rilevato, non conosce l'origine dei propri strumenti ed il loro sviluppo storico, che è essenziale per comprenderne il significato e l'effettiva portata: di qui si originano fraintendimenti che possono ripercuotersi sulla programmazione e sull'interpretazione dei test sperimentali. Ciò vale in generale per tutta la metodologia statistica.

Tornando ora al calcolo degli intervalli di confidenza dei parametri R e Ω, si deve inoltre considerare che esso comporta in genere una lunga serie di approssimazioni di numeri molto piccoli, spesso assai prossimi allo zero, cosicché il risultato potrà risentire, talora anche pesantemente, di tale accumulo di valori approssimati.

Di quali informazioni, che non siano già racchiuse nei test di significatività ed in una intelligente analisi qualitativa dei risultati sperimentali, si può dunque ancora legittimamente parlare?

Vediamo un altro esempio: la Tab. 1, qui di seguito riportata, è relativa ad un test sul trattamento con farmaci Ca^{++} antagonisti nell'infarto miocardico[56] .

Tabella I

		Morti	Sopravvissuti
Trattati	3.195	210	2.985
Controlli	3.225	191	3.034
Totali	6.420	401	6.019

[56] Anche questa tabella è ripresa da R. Marchioli, G. Tognoni, *Cause-effetti in medicina,* cit., p. 22.

Il test χ^2 mostra che i risultati sperimentali non sono significativi e che l'ipotesi-nulla (le differenze tra i due gruppi sono accidentali) non può essere respinta. Ora, il procedimento per il calcolo dell'intervallo di confidenza di Ω *(odds ratio)* ha dato i seguenti valori: Ω=1,11; intervallo di confidenza di Ω: 0,90-1,36. Tale intervallo, contenendo il valore 1, mostra la propria consistenza con il test χ^2, e cioè la non-significatività dei risultati. In breve, il gruppo trattato con il farmaco non si differenzia significativamente da quello trattato con il placebo, e, di conseguenza, i due campioni dovrebbero essere considerati come appartenenti ad una medesima popolazione; la leggera differenza tra di essi rientra nella piena normalità delle fluttuazioni campionarie. Vi è qualche altra rilevante informazione, oltre queste derivanti dal test χ^2, che il calcolo dell'intervallo di confidenza per Ω è in grado di dare? A rigore nessun'altra, se si richiede coerenza logica. In effetti, lo stesso calcolo dell'intervallo di confidenza di Ω non sembra avere molto senso, perché, se i due campioni appartengono ad una stessa popolazione, possiamo aspettarci che le fluttuazioni campionarie, in ripetuti esperimenti, tendano a coprire l'intero intervallo entro il quale l'ipotesi-nulla appare ancora consistente con i dati sperimentali. Viceversa, la stima dell'intervallo di confidenza relativo ad un parametro, che esprime a sua volta un confronto tra parametri attribuibili a *due* popolazioni differenti, si pone in evidente contraddizione con la conclusione del test di significatività.

Che tale osservazione non sia basata su una impressione puramente soggettiva, sembra trovare conferma nell'interpretazione offerta da McPherson ad un altro esempio: in due località, indicate con i numeri 1 e 2, vengono eseguiti controlli alle automobili in circolazione al fine di identificare le auto difettose. I risultati dei controlli sono riassunti nella seguente Tab. 2.

Tabella 2

	Località 1	Località 2	Totali
Auto non-difettose	38	51	89
Auto difettose	7	19	26
Totali	45	70	115

Il test χ^2 non permette di respingere l'ipotesi-nulla (non vi sono differenze significative tra le due località; i campioni appartengono ad una medesima popolazione), ed anche l'intervallo di confidenza di Ω risulta compreso tra 0,71 e 5,9. McPherson offre di tali risultati la seguente interpretazione: poiché l'intervallo di confidenza include 1, «i dati sono consistenti con l'assunzione che le probabilità *(odds)* di trovare un'auto non difettosa sono le stesse nelle due località. Tuttavia possono essere almeno sei volte più grandi le vere probabilità *(odds)* di trovare un'auto non difettosa nella località 1 anziché nella località 2»[57]: asserzione che evidentemente, non solo contraddice l'ipotesi-nulla, ma viene addirittura ad implicare che le auto controllate appartengono contemporaneamente ad una stessa popolazione (*ipotesi-nulla*) ed a due differenti (*le vere probabilità*)!

Come può spiegarsi una conclusione così paradossale? Per tentare una risposta, occorre esplicitare anzitutto alcuni presupposti impliciti nella procedura in questione.

Il più rilevante tra essi è che l'ipotesi-nulla non è l'ipotesi casuale (come è in Fisher e come, seguendo Fisher, è stata qui costantemente considerata, in relazione ai test di significatività), bensì, in accordo con l'interpretazione di Neyman-Pearson, una determinata ipotesi parametrica. La differenza può sembrare soltanto interpretativa, ma è tuttavia sostanziale ed implica importanti differenze a livello metodologico: mentre, come ipotesi casuale, l'ipotesi-nulla ha soprattutto il significato negativo di contraddire ogni determinato e significativo rapporto di correlazione, come ipotesi parametrica essa ha il senso positivo che ha ogni distribuzione determinata, come ad esempio quella dei lanci di un dado. In breve, mentre l'ipotesi casuale asserirebbe che i campioni considerati sono campioni casuali estratti da un'unica popolazione (e quindi che le loro differenze rientrerebbero nelle normali fluttuazioni di campionamento), l'ipotesi parametrica asserisce che i campioni appartengono a popolazioni differenti, ma che queste popolazioni hanno una uguale distribuzione. In tal modo, una questione, relativa alla compatibilità dei dati con una ipotesi, è trasformata in una questione di stima, ed i test di significatività in una procedura di accettazione: si apre allora anche la possibilità di ipotizzare che le due popolazioni non siano in realtà del tutto uguali, ma che piccole differenze tra di esse potrebbero ancora essere compatibili con i dati, qualora essi non rivelassero differenze significative.

Del resto, è questo uno dei punti su cui insiste maggiormente Fisher: nel cercare di identificare un test di significatività, quale esso è usato nelle scienze naturali, con un test di accettazione, una delle più profonde differenze cade sulla popolazione, e la confusione su questo aspetto ha portato in molte occasioni a valori numerici erronei. In una procedura di accettazione, infatti, la popolazione da cui i campioni sono estratti è inequivocabilmente definita: essa ha una realtà empirica. Viceversa, la sola popolazione cui i test di significatività possono riferirsi non ha realtà oggettiva, essendo esclusivamente il prodotto della mente dello statistico[58].

[57] G. McPherson, *Statistics in Scientific Investigation. Its Basis, Application and Interpretation,* New York 1990, p. 260.

[58] Cfr. R.A. Fisher, *Statistical Methods and Scientific Inference*, cit., p. 78.

Una seconda assunzione della procedura che stiamo discutendo è che le frequenze osservate in ogni campione siano stime attendibili delle corrispondenti frequenze nelle popolazioni di appartenenza. In tal modo è possibile assegnare valori quantitativi a p_1 e a p_2, ed affrontare quindi, in termini di stima, qualsiasi tipo di relazione tra essi. Se p_1 e p_2 sono stime di parametri ignoti, ma reali e determinati, è evidente che ci sarà anche un valore determinato, anche se altrettanto ignoto, per D, R e Ω.

Il procedimento riflette il peso di tali assunzioni. Mentre i test di significatività sono basati sulla valutazione degli scostamenti da un valore medio determinato, in accordo con l'ipotesi-nulla, la procedura in questione sottintende il riferimento ad un test di accettazione di ipotesi (o meglio ad un test di significatività interpretato secondo la logica dell'accettazione), in sintonia con la scuola statistica americana. Il calcolo degli intervalli di confidenza implica, in particolare, che siano considerate tutte le ipotesi $p_1 \neq p_2$ compatibili con i risultati sperimentali. Le conclusioni paradossali, poc'anzi rilevate, possono essere così spiegate: sebbene le osservazioni siano compatibili con l'ipotesi $p_1 = p_2$, appare ragionevole ipotizzare (ma in base a quale criterio di ragionevolezza?) che i 'veri' valori di p_1 e di p_2 (o di uno soltanto fra essi) siano tali da implicare una percentuale maggiore di individui di un certo tipo in una popolazione che nell'altra.

Resta il fatto che il procedimento ora discusso, applicato a questioni propriamente non quantitative, come è il confronto tra differenti trattamenti farmacologici, risulta troppo artificioso e, d'altra parte, incidendo profondamente sul senso della questione che è oggetto di investigazione, sembra snaturarne il significato originario, non si sa bene in quale direzione.

In generale, possiamo dire che i fraintendimenti sui test di significatività hanno dato origine ad una situazione metodologica abbastanza strana, per quanto riguarda le ipotesi qualitative, della quale, spesso, non vi è neppure consapevolezza. Accade infatti che una ipotesi qualitativa, come quella sull'efficacia di un farmaco, che originariamente non poteva essere direttamente né rifiutata né accettata, ma che poteva contare soltanto sul sostegno indiretto offertole dalla significatività delle correlazioni rilevate nell'esperimento volto a confutare l'ipotesi della loro accidentalità, è sottoposta a test ibridi derivanti, per una metà, da test in grado soltanto di rifiutare l'ipotesi della casualità delle correlazioni rilevate, e, per l'altra metà, da test volti ad accettare o rifiutare ipotesi parametriche determinate: come conclusione di tale procedura, una ipotesi qualitativa causale, per la quale nessuno dei due metodi era adeguato, è accettata o rifiutata. Abbiamo perso, introdotto o alterato qualcosa del senso del problema nel corso del procedimento? Siamo sicuri che questo test sia in grado di dirci davvero qualcosa sull'ipotesi che ci interessa? Che cosa esattamente può dirci, visto che tutti i significati sono stati sconvolti, senza che sia stata apportata nessuna nuova chiarezza?

Abbiamo visto che il metodo degli intervalli di confidenza e quello dei test di ipotesi si basano su una logica del tutto differente da quella che guida i test di significatività, cosicché quelli non possono essere considerati né variazioni migliorative, né metodi alternativi a questi, ma semplicemente metodi idonei a trattare problemi differenti da quelli affrontati con i test di significatività.

Personalmente, trovo che le procedure più vecchie siano fondate su una logica più convincente e meno ambigua. Per questo motivo, mi sembra anche preferibile, alla luce della metodologia statistica disponibile, distinguere, nell'inferenza inversa, questioni attinenti ad ipotesi quantitative e questioni circa correlazioni qualitative, e trattare i due tipi di questioni con metodi differenti.

Le ipotesi quantitative sono connesse, come si è detto, con un valore ignoto di un parametro della popolazione e possono essere affrontate con i metodi di stima, oppure con i test di ipotesi di Neyman-Pearson, oltre che, naturalmente, con il teorema di Bayes.

Le ipotesi qualitative concernono caratteristiche non quantificabili, come ad esempio nell'asserto «il farmaco A è efficace nella patologia F». Tali questioni sono state affrontate da Fisher con la metodologia tipica della logica della scoperta perché anch'esse sono caratterizzate, in definitiva, dal tentativo di confutare l'ipotesi di una pura casualità dei risultati sperimentali.

Vi è tuttavia anche chi pensa che si possano quantificare concetti, quali *l'efficacia* di un farmaco, e che sia legittimo farlo in termini dei successi ottenuti, dimenticando, però, che «efficace» è un termine disposizionale, come «fragile» e «solubile»; un farmaco può dare risultati positivi sotto certe condizioni, che dipendono solo in parte da caratteristiche intrinseche del farmaco, ma che, per altro verso, dipendono dalla «stabilità» di certe caratteristiche della popolazione dei pazienti considerati. Perciò, mentre la potenziale 'efficacia' del farmaco dovrebbe essere connessa con certe leggi della chimica, della fisiologia, ecc., il numero dei successi si connette invece con l'uniformità o disuniformità della popolazione e con le particolarità proprie della risposta individuale alla terapia. Ciò fa sì che la demarcazione tra successo ed insuccesso di un farmaco non sia così netta come l'estrazione di una pallina bianca o nera: vi sono presumibilmente delle differenze di grado, sia in relazione alla 'forza' del successo (ad esempio guarigione più o meno completa), sia al tempo occorso per conseguirlo.

Aggiungiamo che l'enunciato che esprime la frequenza relativa dei *successi* di un farmaco è un'asserzione descrittiva di fatti avvenuti, ma esso non sembra assumibile come generalizzazione relativa ad una qualche caratteristica del farmaco, denominata «efficacia», proprio come il numero di bicchieri che si sono rotti non può essere assunto come 'stima' di un fantomatico ignoto valore della 'fragilità' del vetro. Perciò, per quanto sia indiscutibile che l'efficacia di un farmaco abbia a che fare con il numero dei successi, perché solo questi ne rendono visibile testimonianza, non sembra convincente identificare i due concetti ed assumere la frequenza relativa dei successi come misura dell'efficacia del farmaco, esattamente come se quest'ultima fosse una reale grandezza da stimare; fosse, anzi, addirittura la popolazione da cui i successi ottenuti sono estratti!

Un qualche tipo di ipostatizzazione di parametri inesistenti pare costituire, tuttavia, un'assunzione quasi costante, implicita nelle procedure che hanno finito per sostituire i test di significatività.

Conclusione

Possiamo ora tentare un bilancio complessivo del nostro lavoro.

L'analisi della logica sottesa ai vari metodi di inferenza statistica, coadiuvata dalla puntuale attenzione alla storia dell'origine e dello sviluppo di ciascun metodo, così come è documentata dalle opere più significative sull'argomento, ha portato ad individuare, come chiave interpretativa costante, le peculiarità dei contesti della scoperta e della giustificazione, il primo caratterizzato dalla esigenza della eliminazione del caso, il secondo dalla ricerca di criteri di accettazione per ipotesi determinate.

L'uso di tale chiave interpretativa, oltre a contribuire, come spero, al chiarimento di fraintendimenti cronicizzati, come quelli relativi ai test di significatività, ha permesso di esplicitare in modo consistente il significato e la portata di ciascuno dei metodi considerati; ha permesso, cioè, non solo di delimitarne il campo peculiare di applicazione, ma anche di precisare i limiti di senso delle risposte che ciascuno di essi è in grado di fornire.

Naturalmente, la collocazione di ciascun metodo nel proprio specifico contesto non preclude in assoluto la possibilità di estenderne l'uso anche oltre l'ambito di peculiarità. L'uso di un test di scoperta al fine di offrire supporto favorevole all'accettazione di una ipotesi, come generalmente avviene, richiede tuttavia sempre cautele e accorgimenti (sia nella formulazione del problema che nell'interpretazione del test) che soltanto una consapevolezza critica delle specificità e dei limiti dei metodi usati può efficacemente mettere in atto.

L'uso corretto e proficuo di un qualunque strumento metodologico richiede, come è ovvio, che vi sia adeguatezza tra lo scopo che si vuole raggiungere ed il mezzo usato per perseguirlo; ma la valutazione di tale adeguatezza rimanda a sua volta, nel caso specifico, da un lato alla chiarezza ed alla precisione con cui è formulato e definito il problema che interessa, dall'altro alla conoscenza non superficiale delle

Maria Grazia Sandrini, *Filosofia dei metodi induttivi e logica della ricerca*, ISBN 978-88-6453-034-5 (print) ISBN 978-88-6453-037-6 (online)

potenzialità dello strumento metodologico stesso. La carenza in uno o in ambedue gli aspetti non può che riflettersi negativamente sull'interpretazione dei risultati degli esperimenti effettuati.

Peraltro, sembra abbastanza diffusa la tendenza a ritenere che le tecniche statistiche, come in genere ogni strumento matematico, siano una sorta di meccanismo pressoché automatico di inferenza, neutrale rispetto a compromissioni epistemologiche, ontologiche e filosofiche. Il presente lavoro dovrebbe aver reso manifesto, viceversa, come tale convinzione sia profondamente errata, e come ogni metodologia sia intimamente connessa con una concezione epistemologica. Il fatto che, con il trascorrere degli anni e con gli sviluppi più propriamente tecnici dei metodi stessi, sia andato in gran parte perduto il senso immediato di tale originaria teoria della conoscenza rende solo l'uso dei test più ambiguo e la loro interpretazione più soggetta a rischio di errore.

È facile che l'insieme delle tecniche metodologiche in uso si trascini dietro una concezione del mondo in parte anacronistica con lo sviluppo attuale della scienza o in disaccordo con le concezioni epistemologiche al momento più accreditate. Un lavoro di riconsiderazione dei nessi tra metodologia ed epistemologia, anche al fine di rendere gli approcci metodologici maggiormente flessibili alle esigenze di differenti e molteplici concezioni della conoscenza, può costituire un impegno di grande interesse sia per lo statistico che per il filosofo. In ogni caso, una proficua utilizzazione dei metodi di inferenza statistico-induttiva non può prescindere da un'adeguata conoscenza della loro storia e del retaggio filosofico, dal quale il loro stesso significato risulta, tutto o in parte, condizionato.

A tale proposito, la questione dei metodi di accettazione delle teorie, che l'approccio bayesiano mette direttamente in discussione, mostra inequivocabilmente come entri qui in giuoco, e non in modo marginale, la visione filosofico-epistemologica esplicitamente o implicitamente adottata. Tuttavia, gli argomenti contro l'approccio bayesiano all'accettazione delle teorie e alla scelta tra teorie rivali, su cui si è creduto opportuno insistere, in quanto indipendenti da una presa di posizione filosofica, sono sopratutto quelli di carattere logico e formale, volti a mostrare come l'approccio probabilistico in tale tipo di questioni risulti, *quanto meno,* sterile.

A conclusione, possiamo riassumere i risultati più significativi nei seguenti punti:

1. La chiave interpretativa usata, mentre consiglia una più serena riconsiderazione dell'annosa polemica tra bayesiani e 'oggettivisti', ci ha permesso di collocare l'approccio bayesiano ed i metodi di test di ipotesi , tra loro alternativi, nel contesto di giustificazione, mentri i test di significatività, che ci appaiono erroneamente concepiti come alternativi al teorema di Bayes, come peculiari di un

contesto di scoperta, nel quale sia importante eliminare l'ipotesi di un verificarsi casuale degli eventi. Ciò permette, non tanto di far cadere una vecchia polemica, quanto di restituire ai test di significatività il loro originario significato, andato perduto a seguito dei fraintendimenti della scuola statistica americana.

2. I metodi 'giustificativi', così come i metodi di stima, riguardano ipotesi parametriche, ben determinate. Ipotesi 'qualitative'(come «il farmaco è efficace») non sono direttamente *controllabili* con la metodologia a disposizione. Esse possono però essere sottoposte ai test di significatività che potranno permettere di escludere l'ipotesi della mera accidentalità dei risultati sperimentali. Tuttavia, l'accettazione di questo tipo di ipotesi resta un punto molto delicato della metodologia statistica.
3. La concezione 'totalitaria' dell'induttivismo bayesiano si scontra con difficoltà sia tecniche che epistemologiche quando pretendono di sottoporre le teorie scientifiche al vaglio della formula di Bayes. Le teorie scientifiche si differenziano da altre formulazioni scientifiche, come le leggi o le generalizzazioni, per la loro struttura logico-formale e concettuale, che ne fa dei veri e propri 'linguaggi' autonomi. Tale struttura complessa sembra rendere insensata, a meno di farne dei simboli vuoti, la questione della loro accettabilità in termini di alta probabilità e di controllo empirico, sebbene la loro compatibilità con specifiche 'osservazioni' empiriche resti una condizione necessaria di scientificità.

Ciò sembra suggerire l'opportunità di porre un limite alla sensata applicabilità dei metodi di inferenza statistica induttiva nell'ambito dell'inferenza inversa.

Tale applicabilità resta certamente, all'*interno* di un sistema teorico entro cui si definisce il senso della ricerca, per quegli insiemi di ipotesi che soddisfano le condizioni del calcolo delle probabilità o dei test sperimentali.

Come ho già detto altrove[1], lo sviluppo della scienza autorizza a concepire l'attività scientifica lungo due direttrici differenti: la prima, verso l'elaborazione di linguaggi sempre più astrattivi e generalizzati, sempre meno 'empirici' e speculativi e viceversa sempre più «propositivi» e legati ad un'idea di ragione teoretica e pratica insieme; la seconda, verso la ricerca di criteri sempre più potenti di coerenza logica e di compatibilità sperimentale in relazione a linguaggi teorici costituiti. L'indubbia fertilità della metodologia induttiva si connette unicamente con questa seconda direttrice dell'attività scientifica: nel contesto di un sistema teorico, o

[1] M.G. Sandrini, *Probabilità e induzione*, cit.

logico-linguistico, assunto, essa aiuta sia ad individuare quegli aspetti del proprio campo di ricerca significativi e meritevoli di essere approfonditi, sia di determinare le ipotesi con esso compatibili e più fertili sotto aspetti specificati.

Ma la logica della costituzione del linguaggi, in cui rientra la formulazione e l'elaborazione delle teorie scientifiche, non coincide con quella che opera all'interno di linguaggi già costituiti. Ciò non significa, si badi bene, la rinuncia ad uno spirito rigorosamente scientifico, o l'apertura verso l'«anything goes» di feyrabendiana memoria; bensì il rifiuto che tutto lo spirito scientifico debba essere racchiuso in criteri empiristici di razionalità o in concezioni realistiche della scienza, che fanno di questa l'ancella della verità. Significa, invece, reclamare una estensione del concetto di razionalità e della funzione della scienza, che tenga conto degli aspetti positivi impliciti negli esiti relativistici della ricerca filosofica contemporanea, e che, fondandosi su una possibile riunificazione della sfera pratica con quella teoretica, come suggerito da Carnap, identifichi nell'attività scientifica, oltre che la caratteristica che distingue l'uomo dalle altre specie biologiche, lo strumento con cui l'uomo possa tentare di dirigere liberamente la propria esistenza e la propria storia.

Bibliografia

Bacon, F. *Novum Organum*, London 1620.

Bayes, Th., *An Essay towards Solving a Problem in tbe Dottrine of Chances*, «Philosophical Transactions of the Royal Society», 1764.

Bernoulli, J., *Ars Conjectandi*, Basel 1713.

Boole, G., *An Investigation of the Laws of Thougth*, London 1854; trad. it., *Indagine sulle leggi del pensiero,* Torino 1976.

Carnap, R., *Empiricism, semantics and ontology,* «Revue internationale de philosophie»*,* 11, 1950; trad. it. in A. Pasquinelli (a cura), *Neopositivismo,* Torino 1978.

Carnap, R., *Philosophical Foundations of Physics,* New York 1966*;* trad. it. *Fondamenti filosofici della fisica,* Milano 1966.

Carnap, R., *The Logical Foundations of Probability*, 1950 e 1962, rist. 1967.

Carnap, R., *Analiticità, significanza, induzione,* Bologna 1971.

de Finetti, B., *Teoria delle probabilità, ,* Torino 1970, 2 voll.

de Finetti, B., *Logica dell'incerto,* Milano 1989.

de Finetti, B., *L'invenzione della verità,* Milano 2006.

De Moivre, A., *The Doctrine of Chance,* 1° ed., London 1718; 2° , London 1738; 3°, London 1756.

De Morgan, A., *Formal Logic,* London 1847.

Feigl, H., *Induzione e empirismo*, Roma 1979.

Fisher, R.A., *The Design of Experiments*, London 1949; trad. it., *La programmazione degli esperimenti*, Pisa 1954.

Fisher, R.A., *Statistical Methods for Research Workers*, London 1936; trad. it., *Metodi statistici ad uso dei ricercatori,* Torino 1948.

Fisher, R.A., *Contributions to Mathematical Statistics*, New York 1950.

Fisher, R.A., *Statistical Methods and Scientific Inference,* London 1956.

Maria Grazia Sandrini, *Filosofia dei metodi induttivi e logica della ricerca*, ISBN 978-88-6453-034-5 (print) ISBN 978-88-6453-037-6 (online)

Galilei, G., *Opere*, Torino 1964, voll. 2.

Hempel, C., *Philosophy of Natural Science,* New Jersey 1966; trad.it. *Filosofia delle scienze naturali,* Bologna, 1968.

Hempel, C, *Fundamentals of concept formation in empirical science*, Chicago 1952; trad. it. *La formazione dei concetti e delle teoria nella scienza empirica,* Milano 1976.

Herschel, J.F.W., *Preliminary Discourse on the Study of Natural Philosophy,* London 1830.

Hintikka, J., Suppes, P., *Aspects of Inductive Logic*, Amsterdam 1966.

Hodges, J.L. jr., Lehmann, E., *I concetti fondamentali della probabilità e della statistica,* Bologna,1971, 2 voll.

Howson, C. (a cura), *Critica della ragione scientifica,* Milano 1981.

Howson C., Urbach P., *Scientific Reasoning. The Bayesian Approach*, La Salle 1989.

Hume, D., *A Treatise of Human Nature,* London 1739.

Jevons, W.S., *The Principles of Science*, London 1873 e 1877; rist. 2° ed., New York 1958.

Keynes, J.M., *A treatise on Probability*, 1° ed. 1921, New York 1962.

Kuhn, Th., *The structure of Scientific Revolutions*, Chicago 1962; trad. it., *La struttura delle rivoluzioni scientifiche,* Torino 1969.

Lakatos, I. (a cura di), *The problem of Inductive Logic,* Amsterdam 1968.

Laplace, P.S., *Théorie analytique des probabilités*, Paris 1812.

Laplace, P.S., *Essai philosophique sur les probabilités,* Paris 1814.

Mach, E., *Die Analyse der Empfindungen und das Verhältnis des Physischen zum Psychischen,* 1896; trad. it. *L'analisi delle sensazioni e il rapporto tra fisico e psichico,* Milano 1975.

Mach, E., *Erkenntnis und Irrtum*, Leipzig 1905; trad. it. *Conoscenza ed errore*, Torino 1982.

Marchioli R., Tognoli G., *Cause-effetti in medicina,* Roma 1994.

McPherson, G., *Statistics in Scientific Investigation*, New York 1990.

Mill, J.S., *System of Logic*, London 1843; trad. it., *Sistema di logica,* Roma 1968.

Monari, P., Cocchi, D., (a cura), *Probabilità e Induzione*, Bologna 1993.

Montmort, P.R., *Essay d'analyse sur les jeux de hazard*, Paris 1713.

Nagel, E., *The Structure of Science,* Harcourt 1961; trad.it., *La struttura della scienza,* Milano 1968.

Newton, I., *Philosophiae Naturalis Principia matematica*, London 1687.

Neyman, J., *First course in Probability and Statistics,* New York 1953.

Neyman, J., *The problem of inductive inference,* «Comm. Pure Appl. Meth.», 8, 1955.

Neyman, J., *«Inductive Behavior» as a Basic Concept of Phylosophy of Science,* «Revue de l'Institut Intern. de Statistique», vol 25, 1957.

Neyman, J., *Statistics and probability,* «Science», 127, 1958.

Neyman, J., *Behavioristic points of view on Mathematical Statistics,* in AA.VV., *On political economy and econometrics. Essays in honour of Oskar Lange,*Varsavia 1965.

Neyman, J., Pearson, S.E., *The Testing of Statistical Hypotheses in Relation to Probabilities a Priori*, «Proceedings of the Cambridge Philos. Society», XXIX, 1932-33.

Pearson, K., *The Grammar of Science*, London 1892.

Popper, K., *The Logic of Scientific Discovery,* London 1934; trad. it. *Logica della scoperta scientifica*, Torino 1970.

Reichenbach, H., *Experience and prediction,* Chicago 1938.

Reichenbach, H., *The Theory of probability. An Inquiry into the logical and mathematical Foundations of the calculus of probability,* Berkeley and Los Angeles 1949.

Reichenbach, H., *The Rise of Scientific Philosophy*, Los Angeles 1951; trad. it., *La nascita della filosofia scientifica*, Bologna 1961.

Reichenbach, H., *L'analisi filosofica della conoscenza*, Padova 1968.

Sandrini, M.G., *Probabilità e induzione. Carnap e la conferma come concetto semantico*, Milano 1991.

Scheffler, I., *The Anatomy of Inquiry,* New York 1963; trad. it., *Anatomia della ricerca*, Milano 1972.

Swinburne, R., ed., *Bayes's Theorem*, New York 2002.

Mises von, R., *Wahscheinlichlkeit, Statistik und Wahrheit,* Wien, 1928; 1° ed. inglese, *Probability, Statistics and Truth,* London 1939.

Venn, J., *The Logic of Chance,* London 1866.

Whewell, W., *Philosophy of the Inductive Science,* London 1840, rist. London 1967.

Whewell, W., *Philosophy of Discovery ,* London 1860.

Wright von G.H., *The logical problem of induction*, Oxford 1965.

Tavole Test

Tavola della v.a. t di Student per alcuni valori della sua funzione di ripartizione

n	0,60	0,70	0,80	0,90	0,95
1	0,32	0,73	1,38	3,08	6,31
2	0,29	0,62	1,06	1,89	2,92
3	0,28	O,58	0,98	1,64	2,35
4	0,27	0,57	0,94	1,53	2,13
5	0,27	0,56	0,92	1,48	2,O2
6	0,26	0,55	0,91	1,44	1,94
7	0,26	0,55	0,90	1,42	1,90
8	0,26	0,55	0,89	1,40	1,86
9	0,26	0,54	0,88	1,38	1,83
10	0,26	0,54	0,88	1,37	1,81
11	0,26	0,54	0,88	1,36	1,80
12	0,26	0,54	0,87	1,36	1,78
13	0,26	0,54	0,87	1,35	1,77
14	0,26	0,54	0,87	1,35	1,76
15	0,26	0,54	0,87	1,34	1,75
16	0,26	0,54	0,86	1,34	1,75
17	0,26	0,53	0,86	1,33	1,74
18	0,26	0,53	0,86	1,33	1,73
19	0,26	0,53	0,86	1,33	1,73
20	0,26	0,53	0,86	1,33	1,73
21	0,26	0,53	0,86	1,32	1,72
22	0,26	0,53	0,86	1,32	1,72
23	0,26	0,53	0,86	1,32	1,71
24	0,26	0,53	0,86	1,32	1,71
25	0,26	0,53	0,86	1,32	1,71
26	0,26	0,53	0,86	1,32	1,71
27	0,26	0,53	0,86	1,31	1,70
28	0,26	0,53	0,85	1,31	1,70
29	0,26	0,53	0,85	1,31	1,70
30	0,26	0,53	0,85	1,31	1,70
60	0,25	0,53	0,85	1,30	1,67
∞	0,25	0,52	0,84	1,28	1,64

Maria Grazia Sandrini, *Filosofia dei metodi induttivi e logica della ricerca*, ISBN 978-88-6453-034-5 (print) ISBN 978-88-6453-037-6 (online) © 2009 Firenze University Press

Tavola della v.a. χ^2 per alcuni valori della sua funzione di ripartizione e numero dei gradi di libertà (n)

$F(\chi^2)$ / n	0.01	0.05	0.50	0.90	0.95	0.99
2	0,020	0,103	1,386	4,605	5,991	9,210
3	0,115	0,352	2,366	6,251	7,815	11,341
4	0,297	0,711	3,357	7,779	9,488	13,277
5	0,554	1,145	4,351	9,236	11,070	15,086
6	0,872	1,635	5,348	10,645	12,592	16,812
7	1,239	2,167	6,346	12,017	14,067	18,475
8	1,646	2,733	7,344	13,362	15,507	20,090
9	2,088	3,325	8,343	14,684	16,919	21,666
10	2,558	3,940	9,342	15,987	18,307	23,209
11	3,053	4,575	10,341	17,275	19,675	24,725
12	3,571	5,226	11,340	18,549	21,026	26,217
13	4,107	5,892	12,340	19,812	22,362	27,688
14	4,660	6,571	13,339	21,064	23,685	29,141
15	5,229	7,261	14,339	22,307	24,996	30,578
16	5,812	7,962	15,338	23,542	26,296	32,000
17	6,408	8,672	16,338	24,769	27,587	33,409
18	7,015	9,390	17,338	25,989	28,869	34,805
19	7,633	10,117	18,338	27,204	30,144	36,191
20	8,260	10,851	19,337	28,412	31,410	37,566
21	8,897	11,591	20,337	29,615	32,671	38,932
22	9,542	12,338	21,337	30,813	33,924	40,289
23	10,196	13,091	22,337	32,007	35,172	41,638
24	10,856	13,848	23,337	33,196	36,415	42,980
25	11,524	14,611	24,337	34,382	37,652	44,314
26	12,198	15,379	25,336	35,563	38,885	45,642
27	12,879	16,151	26,336	36,741	40,113	46,963
28	13,565	16,928	27,336	37,916	41,337	48,278
29	14,256	17,708	28,336	39,087	42,557	49,588
30	14,953	18,493	29,336	40,256	43,773	50,892

MANUALI

Biomedica

Branchi R., *Le impronte nel paziente totalmente edentulo*

Rossetti R., *Manuale di batteriologia clinica. Dalla teoria alla pratica in laboratorio*

Rucci L., *Testo Atlante di embriologia clinica della Laringe. La chirurgia conservativa compartimentale della regione glottica*

Scienze

Bart J.C.J., *Polymer Additive Analytics. Industrial Practice and Case Studies*

Caramelli D., *Antropologia molecolare. Manuale di base*

Scialpi A., Mengoni A. (a cura di), *La PCR e le sue varianti. Quaderno di laboratorio*

Simonetta M.A., *Short history of Biology from the Origins to the 20th Century*

Spinicci R., *Elementi di chimica*

Scienze Sociali

Ciampi F., *Fondamenti di economia e gestione delle imprese*

Maggino F., *L'analisi dei dati nell'indagine statistica. Volume 1. La realizzazione dell'indagine e l'analisi preliminare dei dati*

Maggino F., *L'analisi dei dati nell'indagine statistica. Volume 2. L'esplorazione dei dati e la validazione dei risultati*

Magliulo A., *Elementi di economia del turismo*

Visentini L., Bertoldi, M., *Conoscere le organizzazioni. Una guida alle prospettive analitiche e alle pratiche gestionali*

Scienze Tecnologiche

Borri C., Pastò S., *Lezioni di ingegneria del vento*

Borri C., Betti M., Marino E., *Lectures on Solid Mechanics*

Gulli R., *Struttura e costruzione / Structure and Construction*

Policicchio F., *Lineamenti di infrastrutture ferroviarie*

Umanistica

Bertini F., *Risorse, conflitti, continenti e nazioni. Dalla rivoluzione industriale alle guerre irachene, dal Risorgimento alla conferma della Costituzione repubblicana*

Bombi A.S., Pinto G., Cannoni E., *Pictorial Assessment of Interpersonal Relationships (PAIR). An analytic system for understanding children's drawings*

Borello E., Mannori S., *Teoria e tecnica delle comunicazioni di massa*

Brandi L., Salvadori B., *Dal suono alla parola. Percezione e produzione del linguaggio tra neurolinguistica e psicolinguistica*

Marcialis N., *Introduzione alla lingua paleoslava*

Michelazzo F., *Nuovi itinerari alla scoperta del greco antico. Le strutture fondamentali della lingua greca: fonetica, morfologia, sintassi, semantica, pragmatica*

Peruzzi A., *Il significato inesistente. Lezioni sulla semantica*
Peruzzi A., *Modelli della spiegazione scientifica*
Sandrini M.G., *Filosofia dei metodi induttivi e logica della ricerca*
Trisciuzzi L., Zappaterra T., Bichi L., *Tenersi per mano. Disabilità e formazione del sé nell'autobiografia*

Finito di stampare presso