La teoria Bayesiana è una teoria di buon senso

Pietro Muliere
Dipartimento di Economia Politica e Metodi Quantitativi
Università degli Studi di Pavia

2 Ottobre 1996

" In definitiva è solo dei fatti, dei singoli fatti, che ha senso parlare. E' solo ai fatti, ai singoli fatti che (se sono futuri, o se, comunque, ne ignoriamo l'esito) possiamo attribuire una probabilità ".

(B. de Finetti (1971)).

1 Premessa

Sul tema **Induzione** si sono scontrati schiere di filosofi raccolti in gruppi agguerritissimi: induttivisti e deduttivisti, falsificazionisti e verificazionisti, empiristi e razionalisti, convenzionalisti e pragmatisti, oggettivisti e soggettivisti e tanti altri ancora. Una allegra confusione di linguaggi, uno scontro continuo di idee. Il dibattito continua, vi attendono gli addetti alla Scienza ed altri personaggi che non sono proprio filosofi e nemmeno tanto scienziati: *gli Statistici* . Anche loro divisi in fazioni irriducibili intenti a disputar di teoria ed a risolvere problemi.

Come affrontare dunque il problema dell'induzione? Come porre in evidenza le distinzioni e le connessioni tra le varie scuole di pensiero o come analizzare i motivi che portarono alla nascita o al superamento di una concezione?

Esistono diverse risposte a tali quesiti. Un primo modo potrebbe essere quello di concentrare la nostra attenzione su una "storia interna" delle tecniche inferenziali. Un secondo approccio potrebbe consistere nell'affermare che per una corretta ricostruzione dei fondamenti dell'inferenza statistica bisognerebbe concentrarsi su fattori esterni: quali l'ambiente ed i bisogni delle istituzioni. Una terza risposta potrebbe essere quella di fondere in un'unica storia, sia la storia interna che la storia esterna.

Tuttavia il problema comune a tutti e tre gli approcci è il seguente: con quali criteri è ragionevole tracciare un confine tra storia interna e storia esterna ?

In questa relazione non affronterò questi problemi; non mi chiederò fino a che punto l'enunciato di un teorema fa parte della storia interna o se le convinzioni filosofiche che hanno portato ad una concezione fanno parte della storia esterna.

Non traccerò una storia dell'induzione nella ricerca scientifica anche perchè credo che non vi possa essere altra storia se non quella della consapevolezza che il *momento induttivo è essenziale alla conoscenza.*

Credo che cercare nella storia della scienza l'apporto concettuale venuto alle scienze dal metodo statistico, dà spesso alla Statistica i tratti di un meccanismo onnivalente. La storia della Scienza non è solo la storia delle scoperte e delle invenzioni: è la storia delle ricerche e degli errori, delle ipotesi e dei controlli, dei "totem" e dei "tabù di ogni epoca. Una storia che molto spesso è indecifrabile.

In questa relazione mi limiterò a discutere dei fondamenti dell'inferenza statistica e proverò a chiarire la logica sottostante l'impostazione bayesiana. Dopo aver presentato brevemente il problema dell'induzione e aver messo in evidenza il legame tra probabilità ed induzione, darò uno sguardo alla storia interna dell'inferenza statistica e presenterò con maggior dettaglio il meccanismo inferenziale bayesiano per metter in evidenza come l'approccio bayesiano sia, almeno a mio avviso, l'unico modo corretto per affrontare e risolvere problemi inferenziali.

2 Induzione e Probabilità

Induzione. Il problema, lo sappiamo, ottiene la sua formulazione canonica nel 1739 in "*A Treatise of Human Nature*" di Hume.

Siamo di fronte ad un bigliardo, "*la prima palla viene messa in movimento; tocca la seconda che immediatamente si mette in moto*". Questa è la situazione che Hume ci invita a considerare e così continua "*Nemmeno un uomo come Adamo nel pieno vigore intellettuale ma senza esperienza sarebbe capace di inferire il movimento della seconda palla dall'impulso impressole dalla prima....... Ma se egli avesse visto un numero sufficiente di esempi di questo tipo .... allora inferirebbe sempre senza esitazione il movimento della seconda palla* ". La conclusione di Hume è che "*allora tutti i ragionamenti concernenti la causa e l'effetto [induttivi] sono fondati sull'esperienza e tutti i ragionamenti di questo tipo sulla supposizione che ...cause simili in circostanze simili produrranno sempre gli stessi effetti*". Ma si chiede Hume " *che ragione abbiamo per supporre che ... il futuro sia conforme al passato ? . Questa conformità è una questione di fatto*" osserva Hume "*dunque, se deve essere provato, l'unico tipo di prova ammesso sarà dall'esperienza*". Ma e tocchiamo qui il punto cruciale "*la nostra esperienza nel passato non può costituire una prova di alcunchè nel futuro, se non appunto a partire dalla supposizione che vi sia rassomiglianza tra essi. Questa è perciò una supposizione che non ammette alcuna prova e che accettiamo senza prove*".

Di qui Hume trae la famosa conclusione che doveva inquietare la filosofia occidentale per due secoli : **non è la ragione che è la guida della vita, ma l'abitudine. Questa soltanto determina la mente....... a supporre la conformità del futuro al passato. Per quanto ovvio questo passo possa sembrare, la ragione non sarebbe capace di farlo in tutta l'eternità** (D. Hume,1739).

Il problema filosofico dell'induzione è allora il seguente : siamo razionalmente giustificati nel ragionare da esempi ripetuti di cui abbiamo esperienza ad esempi di cui non abbiamo avuto esperienza? La risposta di Hume è seccamente negativa. Nè, dopo Hume, chi ha raccolto la sua sfida è riuscito a far molto meglio. Io temo che questa sequenza di sconfitte è dovuta al fatto che la sfida di Hume non possa essere vinta sul terreno proposto da Hume.

Ancora oggi si assiste ad un acceso dibattito sui fondamenti del ragionamento induttivo ed ad una divisione nell'impostazione e nella risoluzione di questioni più propriamente tecniche. Analoghe divisioni si ripetono nel campo della teoria della probabilità. Si può dire, estremizzando, che ad ogni idea di probabilità corrisponda una proposta induttiva. La causa nascosta di molte discussioni sull'inferenza statistica è proprio rappresentata dal mancato riconoscimento di tale diversità di atteggiamenti sulla probabilità.

Già nella bernoulliana "*Ars Conjectandi*" (1713), nell'*Essay* Bayesiano (1763), saggi entrambi postumi, nell'*Essay Philosophique* di Laplace (1795), per non citare che i momenti più significativi, **Probabilità ed Induzione** appaiono quasi categorie inscindibili. A conferma di ciò possiamo citare l'autorevole giudizio di H. Poincarè, secondo il quale :"*tutte le volte in cui si ragiona per induzione, si fa, più o meno coscientemente, uso del calcolo delle probabilità.*

Per avvalorare questo punto di vista utilizziamo le conclusioni raggiunte da Bruno de Finetti (1938) nel Colloquio di Ginevra sulla probabilità. Scrive de Finetti:" *I casi in cui non si hanno dubbi che il risultato ottenuto in un esperimento (anche quando esso sia stato realizzato una sola volta) si ripeta in condizioni somiglianti, sono semplicemente i casi in cui, l'osservazione di eventi di tipo analogo, avendo mostrato che le frequenze per ogni tipo sono sempre state 0 oppure 1, si è portati a prevedere che anche le esperienze del nuovo tipo avranno lo stesso risultato, vale a dire che esse si svolgono in conformità con le vedute di un determinista. Nei casi di questo genere, si è dunque proprio condotti a ragionare ed ad agire come se si conservasse la fede dei deterministi nel principio di causalità e ciò attribuendo all'idea di causa il significato corrispondente alla concezione di Hume. Così la causalità non ci appare più come principio che sia necessario premettere affinchè la scienza sia possibile, essa è il risultato del processo naturale di associare le idee secondo l'influenza suggestiva di certe analogie, e si collega alla teoria della probabilità tramite la quale soltanto questo processo psicologico può essere analizzato e spiegato* ".

Da queste poche osservazioni emerge con chiarezza la centralità del concetto di probabilità nel ragionamento induttivo. Conviene pertanto porre molta attenzione nella scelta di una o dell'altra concezione, poichè il problema essenziale è ancora oggi quello del significato o del valore da attribuire alla probabilità.

Il problema da enucleare ed analizzare è, a mio avviso, il seguente : **verificare se la teoria della probabilità possa svolgere un ruolo unificante nel ragionamento induttivo.**

3 Probabilità

Per quanto riguarda l'aspetto più propriamente matematico, le idee sulla probabilità non sono fortemente discordanti (l'unico punto in discussione è se le funzioni debbano essere finitamente additive o completamente additive). Il nucleo della discordia concerne invece il significato della nozione di probabilità e può essere sintetizzato nel ricordare che da un lato vi sono concezioni secondo cui, per elevare a rango scientifico la probabilità, si deve depurarla di tutti gli elementi soggettivi che la riguardano e la caratterizzano; dall'altro vi sono quelle concezioni che non solo non ritengono tali elementi come causa di disturbo, ma ne fanno il punto di partenza per la definizione e la conseguente teoria matematica.

Dal punto di vista filosofico, le concezioni del primo tipo (oggettive) ritengono la probabilità come elemento del mondo fisico, esistente fuori di ciascuno di noi; le concezioni del secondo tipo (soggettive) ritengono che la probabilità, esprimendo in ogni caso una opinione personale, venga ad esistenza solo nel momento in cui il giudizio del soggetto viene a contatto di date circostanze concrete, incerte. La probabilità, pertanto, acquista significato solo in funzione di un dato individuo in specificate situazioni (si veda De Morgan 1847) .
Al dibattito sui fondamenti della probabilità hanno contribuito moltissimi studiosi tra cui ricordiamo: R. von Mises, J.M. Keynes, F.P. Ramsey, B. de Finetti e R. Carnap.

L'inferenza statistica classica fa riferimento alla concezione di probabilità con un significato oggettivo e quindi indissolubilmente ancorata alla nozione di frequenza. mentre l'inferenza bayesiana trova un valido e rigoroso riferimento principalmente nella concezione soggettiva di probabilità.

Soffermiamoci brevemente sulle definizioni di probabilità frequentista e soggettivista facendo riferimento rispettivamente a Richard von Mises(1921) e a Bruno de Finetti (1931,1937).

La definizione di von Mises passa attraverso l'introduzione di due proprietà cui devono soddisfare le successioni di eventi oggetto di studio designate con il temine "**collettivo** ".

1. Le frequenze relative con cui ciascuno dei possibili risultati si presenta in un numero finito di prove tendono a dei valori limite, detti probabilità o frequenza limite, al tendere del numero delle prove all'infinito.
2. Tali frequenze limite rimangono inalterate se anzichè la successione originale se ne considera una estratta da essa in un modo qualunque ma ben precisato.

Dall'impiego di questi due assiomi discende che la probablità è una funzione semplicemente additiva sugli esiti delle prove, a valori reali in [0,1], essendo unitaria la probabilità dell'evento certo.

Secondo i soggettivisti la probabilità rappresenta una misura del grado di plausibilità che un individuo assegna ad un dato evento (incerto).

Nell'impostazione di de Finetti quando si parla di evento si intende un fatto singolo. La definizione numerica del grado di plausibilità assegnato da un individuo ad un evento ben determinato avviene mediante lo schema della scommessa.

Definizione 1 *Dato un evento E si dice probabilita' di E, $P(E) = p$, il prezzo p che un individuo è disposto a pagare (ricevere) per riscuotere (pagare) un importo monetario unitario se E si verifica e nullo se E non si verifica.*

L'alternativa della precedente definizione sta a significare che la valutazione della probabilità non deve cambiare se l'individuo scambia il suo ruolo da scommettitore a banco.

Giova subito sottolineare che lo schema della scommessa adottato per definire la probabilità svolge nella teoria il ruolo di *strumento di misura* perchè in ogni occasione è possibile valutare la probabilità di un evento riferendosi ad una scommessa ipotetica che lo coinvolga e la definizione che ne discende è operativa. Pertanto non si tratta di sapere se l'individuo in questione è o non è disposto ad accettare una scommessa su E a date condizioni ma di sapere quale quota p egli fisserebbe se fosse libero di sceglierla, ma fosse poi obbligato ad accettare su tale base qualunque scommessa sia pro sia contro il verificarsi di E.

Riguardo poi alle quote di scommessa si presenta il problema di ricercare se esse possono essere fissate arbitrariamente o se esistono e quali sono le condizioni che devono essere soddisfatte nel fissarle onde evitare *incoerenze.*

Allorchè un individuo ha valutato le probabilità di certi eventi. due casi si possono presentare: o è possibile scommettere contro di lui assicurandosi di guadagnare a colpo sicuro, oppure questa possibilità non esiste. Nel primo caso si deve evidentemente dire che la valutazione di probabilità data dall'individuo contiene una incoerenza, una contraddizione intrinseca; nell'altro caso diremo che l'individuo è coerente.

Con la condizione di **coerenza** si suppone che l'individuo non desideri stipulare scommesse che gli diano certamente una perdita.

Ebbene, è proprio da questo concetto posto come principio informatore dell'azione dell'uomo di fronte all'incertezza, che è possibile dedurre un insieme di regole cui deve essere assogettata la valutazione di probabilità. In altre impostazioni, tali regole, sono imposte come assiomi.

Altro filone di studio riguarda l'enunciazione in forma assiomatica delle proprietà sostanziali da stabilire come punto di partenza per una teoria formale della probabilità; in quest'ambito segnaliamo le ricerche di F.P. Cantelli, A. Renyi e, soprattutto A.N. Kolmogorov.
La convinzione di fondo di Kolmogorov (1933) si regge sul fatto che la teoria della probabilità può essere sviluppata assiomaticamente prescindendo dalle questioni inerenti il significato di probabilità. Kolmogorov mostrò che la teoria della probabilità si occupa di entità astratte che non necessitano, almeno a livello di sviluppo della teoria, di alcuna interpretazione.
Il punto di partenza è la considerazione di un insieme di risultati, detto spazio dei risultati, Z, e di una classe di sottoinsiemi di Z. Quando l'insieme Z è finito oppure infinito ma numerabile ogni suo sottoinsieme può considerarsi evento. Se Z è infinito non numerabile sorgono alcune difficoltà dovute al fatto che non tutti i suoi sottoinsiemi possono considerarsi eventi. Allora, in vista dell'assegnazione della probabilità, dobbiamo restringere la classe di sottoinsiemi di Z che devono essere considerati eventi. Tale nuova classe, A, deve, tuttavia, essere sufficientemente ampia da permettere di operare con le operazioni di unione, intersezione e complemento sui suoi elementi senza uscire dalla classe stessa. Una tale famiglia di sottoinsiemi di Z è detta sigma-algebra. Comunque sia fissata la classe A, i suoi elementi sono detti insiemi misurabili ed è solo a questi che sarà possibile attribuire la qualifica di eventi ed assegnare una probabilità. La coppia (Z,A) si dice spazio misurabile. La definizione di probabilità viene fissata mediante i seguenti assiomi:

1) ad ogni evento $E \in A$ si associa un numero reale $P(E) \geq 0$
2) per l'evento Z si ha : P(Z) =1
3) per ogni successione $E_1, E_2, \cdots \in Z$ di eventi a due a due incompatibili si ha:

$$P(\bigcup_{i=1}^{\infty} E_i) = \sum_{i=1}^{\infty} P(E_i)$$

4 Inferenza: classica e bayesiana

1. *Inferenza Classica.*

Gli statistici classici ritengono che il ragionamento induttivo debba dipendere solamente dai dati campionari.

La scuola classica si avvale di due principi fondamentali: quello della **verosimiglianza** e quello della **ripetizione dell'esperimento**.

Il primo stabilisce che tutte le informazioni che un generico campione può fornire sono contenute nella funzione di verosimiglianza (probabilità dei dati condizionatamente alle ipotesi) e altresì questa, per un campione dato, misura la plausibilità dei possibili modi di essere del fenomeno.

Il secondo principio considera il campione effettivamente ottenuto alla stregua di uno dei possibili campioni che si sarebbero potuti ottenere ripetendo un gran numero di volte, nelle stesse condizioni, l'operazione di campionamento.

Per questa ragione il punto di riferimento naturale è la concezione frequentista di probabilità.

2.*Inferenza Bayesiana.*

L'inferenza Bayesiana si basa sul legame tra verosimiglianza e distribuzione finale che, insieme alle probabilità iniziali, è il fulcro del teorema di Bayes e quindi il punto fondamentale su cui si basa il processo di acquisizione delle nostre conoscenze. In sintesi, per il teorema di Bayes, si ha:

$$Prob.finale = K \times verosimiglianza \times Prob.iniziale$$

dove K e' una costante di proporzionalità.

Come si può notare in questa impostazione un'analisi completa dei problemi statistici richiede oltre alla verosimiglianza l'individuazione della distribuzione iniziale. La scelta della distribuzione iniziale è soggettiva.

La distribuzione finale esaurisce il problema dell'inferenza statistica in termini bayesiani. Possiamo dire che il tratto distintivo tra l'impostazione classica e quella bayesiana dei problemi di inferenza risiede, nella sostanza, nell'uso della regola di Bayes e nelle conseguenze che ne derivano.

Gli statistici che seguono l'impostazione classica rifiutano l'uso della regola di Bayes a causa dell'interpretazione che assegnano alla probabilità che ha, per essi, un significato oggettivo e quindi indissolubilmente ancorata alla nozione di frequenza. Da ciò discende dunque la rinuncia alla predetta regola, non potendosi sempre assegnare alla distribuzione iniziale tale carattere oggettivo, e la pretesa di concludere il ragionamento induttivo solo con il riferimento alla verosimiglianza e/o alle proprietà dello spazio campionario.

Vale la pena immediatamente sottolineare che anche nei procedimenti inferenziali classici entrano, anche se non vengono esplicitate, alcune valutazioni personali del ricercatore (vedi, ad esempio, la scelta dell'esperimento e la scelta delle procedure da adottare).

5 Le fasi dello sviluppo dell'inferenza statistica

La teoria della probabilità, i cui primi risultati si possono far risalire a Cardano (1501-1576), era subito pervenuta a formulare in modo corretto la problematica induttiva con i lavori di D. Bernoulli (1738) e di T. Bayes (1763). Bernoulli aveva introdotto il concetto di speranza morale e Bayes aveva stabilito il nesso tra probabilità e ragionamento induttivo.

Il criterio di decisione suggerito consisteva nel cercare di **massimizzare la speranza matematica del guadagno**. Se non che quel traguardo raggiunto d'impeto fin dall'inizio è stato dimenticato per due secoli. Perchè ? Proveremo ora a raccontarne la storia.

Seguendo de Finetti (1959) si possono individuare tre fasi nello sviluppo dell'inferenza statistica:

i) in una prima fase vi fu una rapida affermazione della teoria bayesiana, molta leggerezza nell'applicazione e conseguente discredito ed abbandono di tale concezione;

ii) in una seconda fase vi fu la ricerca di altre vie per affrontare ed impostare i problemi dell'inferenza statistica (sviluppo dell'impostazione classica o oggettivistica)

iii) nella terza fase affiorano le insufficienze dell'impostazione classica e vi è una revisione del giudizio sulle teorie bayesiane.

Vediamo di esaminare brevemente queste tre fasi.

1.*Prima fase.*

Nella prima fase importanza decisiva ebbero le discussioni sull'impostazione proposta da Bayes. Il punto cruciale del dibattito fu quello che viene indicato come " postulato di Bayes" o " principio di indifferenza". Il postulato secondo cui, quando non si sa nulla si deve adottare la distribuzione iniziale uniforme. Le critiche all'impostazione bayesiana hanno prevalentemente riguardato il ricorso al principio di indifferenza, considerato per molto tempo come un elemento essenziale di tale impostazione, e tali critiche hanno assunto uno sviluppo tale da far abbandonare la stessa impostazione bayesiana. Tale postulato non ha senso, se non in certi casi, ed è una frase vuota che usata in maniera deteriore portava ad usare sempre ed in modo acritico la distribuzione uniforme. Ciò ha fatto mettere in dubbio l'intero criterio. Vale la pena solamente ricordare come il passaggio dalla distribuzione iniziale a quella finale non sia una applicazione del principio di indifferenza.
Bayes aveva utilizzato la distribuzione uniforme in un caso particolare ma le formule relative a questo caso vennero applicate meccanicamente ad ogni tipo di esemplificazione, tra cui quella, discussa da Laplace, della probabilità che il sole sorga domani dato il numero di giorni da cui si è tramandato che il sole è sempre sorto.
Questo uso acritico e la debolezza logica di tale postulato hanno portato, nonostante il prestigio di Laplace, al sorgere di voci discordi ed anzichè eliminare il difetto i critici hanno pensato di sopprimere il ragionamento difettoso. Purtroppo la comunità scientifica accantonò il modello senza essere riuscita a costruirne uno migliore. Sfortunatamente così con l'acqua sporca venne buttato anche il bambino. Certo, di acqua sporca ce n'era. Sia per quanto riguarda la curva di utilità utilizzata da Bernoulli sia per l'uso disinvolto del principio di indifferenza da parte di Laplace ma invece di analizzarlo per controllare in quali classi di applicazioni potesse valere, si gridò allo scandalo di un principio che avrebbe "*prodotto conoscenza dall'ignoranza*". Così la teoria venne abbandonata. Fu, a mio avviso, una vera e propria sconfitta della razionalità scientifica.

2.*Seconda fase.*

Caduta in discredito l'impostazione Bayesiana gli statistici si impegnarono nella ricerca di procedimenti convincenti di inferenza statistica di tipo nuovo, che non facessero riferimento, almeno in via esplicita, alla formula di Bayes. La preoccupazione di giungere ad una spiegazione *oggettiva* della statistica costituisce la principale idea direttiva del grandissimo sviluppo che la statistica ha avuto in questo secolo. Possiamo parlare di un vero e proprio rinascimento di questa disciplina. Le radici di questa ripresa sono rintracciabili negli studi di Galton, Karl Pearson, Fisher, Gosset . In seguito si è sviluppata una scuola di cui si sogliono nominare, come principali esponenti, J. Neyman ed E. Pearson. La caratteristica di tutto questo movimento era l'insistenza sulla oggettività. Il teorema di Bayes veniva rifiutato nell'assunto che le probabilità iniziali non esistono. La ragione di questo atteggiamento sta nel fatto che la probabilità veniva intesa come frequenza o meglio come limite di una frequenza. Gli oggettivisti affermavano che le valutazioni personali non dovevano entrare nei problemi di incertezza. Afferma Fisher (1934)..."*si può arrivare a valide conclusioni con i dati solamente e senza assunzioni a priori.....*". Così facendo, tuttavia, gli oggettivisti si sono ridotti ad inventare caso per caso *test* per la conferma di ipotesi, arrivando ad una proliferazione di ricette, cioè ad una serie di metodi *ad hoc* per estrarre delle conclusioni da un insieme di dati, come conigli dal cilindro di un prestigiatore, ossia, più concretamente, per prendere le decisoni in base soltanto all'esito delle osservazioni che hanno fornito i dati.

La questione invece della distinzione se si tratti di pensare o agire in senso induttivo, è una questione che divide in due i sostenitori della scuola classica. Da una parte vi è Fisher e dall'altra Neyman e Pearson. Neyman distingue nettamente la stima dei parametri dalla prova delle ipotesi ed afferma che nessun test statistico può consentire di rifiutare una ipotesi se non attraverso il paragone con un'altra ipotesi. Neyman e Pearson partono dall'assunto che nessun test può dare evidenza alla verità o falsità di una singola ipotesi. L'ipotesi viene respinta solo a vantaggio di un'altra più verosimile alla luce dei dati, facendo dell'atto inferenziale una regola di condotta. Neyman osserva che accettare una ipotesi non significa necessariamente affermare che quell'ipotesi è vera ma solo che si adotterà una azione connessa con quell'ipotesi in un particolare problema. Per questo motivo Neyman parla di "*comportamento induttivo*" in contrapposizione all'inferenza induttiva di Fisher e rifiuta l'impostazione bayesiana proprio sostenendo che non si tratta di fare un ragionamento induttivo.

Prima di arrivare alla terza fase mi sembra doveroso menzionare H. Jeffreys che non è un soggettivista ma che tuttavia ha sviluppato considerazioni in senso soggettivo. Egli era un logico, attribuiva alla probabilità un significato oggettivo di natura logica e non empirica. Jeffreys sosteneva che vanno prese quelle decisioni che hanno la più elevata probabilità inversa ma che in assenza di informazioni iniziali bisogna usare il principio di Bayes-Laplace della uniformità della distribuzione iniziale.

3. *Terza fase.*

La sempre più stretta connessione tra stima e decisioni non poteva non preparare ad una logica decisionale.

Wald è stato il responsabile, seppure indirettamente, della messa in discussione dell'inferenza classica ed ha contribuito a superare la prevenzione degli oggettivisti contro l'impostazione bayesiana soprattutto rendendo effettivo il concetto di vedere i problemi secondo un comportamento induttivo. Nel 1937 propone l'analisi sequenziale e la scelta di una ipotesi non è più necessariamente una questione che si pone in un momento unico, ma è piuttosto una successione di scelte. I test sequenziali, infatti, considerano di volta in volta tre possibilità: accettare l'ipotesi - rifiutare l'ipotesi - proseguire la sperimentazione fino a che sia possibile una decisione.

Nel 1950 Wald approda alla teoria delle decisioni. Uno dei risultati più importanti della sua teoria mostra che, sotto condizioni molto generali, le *decisioni ammissibili* sono quelle bayesiane vale a dire che la classe di tutte le soluzioni di Bayes e dei limiti delle soluzioni bayesiane è una classe completa cioè una classe che contiene almeno una funzione di decisione uniformemente migliore di qualunque altra non appartenente ad essa. Ciò significa che la soluzione ottimale da dare ad un problema di decisione va ricercata tra le soluzioni bayesiane. Si può dire, come conclusione, che Wald con la sua impostazione aveva incontrato tutto ciò che occorreva per obbligare al ritorno alla concezione bayesiana ma egli si limitò all'aspetto formale e non superò la sua impostazione oggettivistica.

6 Possibilità di un ragionamento induttivo coerente

Come affrontare dunque il problema dell'induzione in maniera logicamente coerente ?

Chiediamoci innanzitutto perchè ci serviamo del ragionamento induttivo qualunque ne sia la sua caratterizzazione migliore. Lo stesso Hume risconosce che il ruolo principale è quello di costituire **la guida nella vita**, e cioè di orientarci nelle nostre decisione pratiche. Forse il problema è più semplice, vale a dire, il problema della scienza è lo stesso del vivere comune e la preoccupazione di tutti è quella di **imparare dall'esperienza** ma non siamo capaci di metterci d'accordo sul modo.

Come abbiamo già detto il problema diventa allora quello di verificare se la teoria della probabilità possa svolgere un ruolo unificante nel ragionamento induttivo. In tal senso direi che il ruolo da assegnare alla teoria della probabilità nella logica induttiva consiste nell'indicare come debba modificarsi la valutazione della probabilità iniziale relativa ad eventi futuri in seguito al risultato di eventi osservati.

E' questo il senso che ha la frase "imparare dall'esperienza". Allora il ragionamento induttivo si riduce sostanzialmente al teorema delle probabilità composte o al suo corollario noto come teorema di Bayes.

Con de Finetti(1970) possiamo dire :"*Il problema del ragionamento induttivo non e' piu' un problema (....) Tutto si riduce alla nozione di probabilita' subordinata (...) ed alle considerazioni (...) sulla dipendenza stocastica per arricchimento di informazione*" e possiamo affermare che : " *...il teorema di Bayes è la chiave di volta ed il concetto informatore di ogni attività costruttiva del pensiero*".

Nella realtà ci troviamo di fronte ad un insieme di ipotesi che formano un particolare modello della situazione pratica ed una e solamente una di esse si potrà verificare; ed inoltre abbiamo a nostra disposizione dei dati che sono il risultato ottenuto nella situazione pratica. Le probabilità delle ipotesi sono le probabilità iniziali e costituiscono una fonte di informazioni. Le probabilità dei dati condizionatamente alle ipotesi sono dette verosimiglianze.
Allora il teorema di Bayes dice essenzialmente:

$$Prob(Ipotesi|Dati) = \frac{Prob(Ipotesi)Prob(Dati|Ipotesi)}{Prob(Dati)}$$

a patto che Prob(Dati) non sia nulla.

In questo teorema di immediata e facile comprensione risiede tutto il meccanismo del processo di apprendimento dall'esperienza, il quale consente di assegnare la probabilità di una

ipotesi alla luce delle informazioni fornite dai dati. Nel teorema figurano Prob(Ipotesi) e $Prob(Ipotesi|Dati)$ ed è proprio la relazione tra queste due quantità che stiamo studiando. La probabilità "finale " dell'ipotesi subordinata ai dati risulta proporzionale alla probabilità iniziale dell'ipotesi e alla cosiddetta verosimiglianza $Prob(Dati|Ipotesi)$.
E' questo il procedimento mediante il quale si **impara dall'esperienza** non nel senso che l'esperienza ci porta a confermare o respingere una opinione (che in quanto tale, non è nè vera nè falsa ma espressa mediante una distribuzione di probabilità) ma solo nel senso che l'esperienza ci fornisce un'informazione capace di arricchire la nostra opinione iniziale, ottenendo cosi una nuova distribuzione di probabilità.

6.1 Il valore di un'informazione

Una prima fondamentale osservazione è che la probabilità dipende dallo stato di informazione. Il medico dopo aver osservato il risultato di una certa analisi, può ritenere opportuno modificare la sua precedente valutazione di probabilità di malattia e giudicarla inferiore o superiore a quella precedendemente espressa. In altri termini, e più precisamente, la probabilità varia con il nostro stato di informazione, continuamente suscettibile di arricchirsi con il flusso di nuove informazioni.

Precisiamo che non esiste un valore di una informazione in quanto tale: esiste il valore che essa ha per un individuo che sulla base di essa può scegliere fra certe azioni o decisioni a sua disposizione.

Per limitarci, in un primo momento, a chiarire in che modo ogni nuova informazione modifica le nostre valutazioni di probabilità, riferiamoci ai diagrammi di Eulero-Venn.

Pensiamo, ad un tabellone contro cui viene scoccata una freccia, ed il successo consista nel colpire un punto entro la zona E. Per comodità supponiamo uniforme la densità di probabilità, nel senso che aree uguali abbiano probabilità uguali.

P(E) è dunque l'area della zona E. Abbiamo inoltre delimitato anche un'altra zona indicata con H che ha punti in comune con E. Se veniamo informati che la freccia ha anche colpito un punto della regione H non attribuiremo più al successo P(E) bensì:

$$P(E|H) = \frac{P(E \cap H)}{P(H)}$$

ossia il rapporto tra l'area di H contenuta in E e l'area totale di H.

E' questo il senso in cui la probabilità è relativa: relativa allo stato attuale di informazione relativo all'evento H. Parlando della probabilità di E (senza altre specificazioni) e indicandola con P(E), va sottinteso che ci riferiamo nel valutarla, al nostro attuale stato di conoscenze. Anche simmetricamente si ha:

$$P(H|E) = \frac{P(E \cap H)}{P(E)}$$

da cui:

$$P(E|H) = \frac{P(E)P(H|E)}{P(H)}$$

ossia il Teorema di Bayes.

Come si vede il teorema di Bayes di riduce a dire che al cambiare dell'informazione si modifica la probabilità di E in quanto rimane solo la parte di area contenuta in H rapportata all'area di H.

La conclusione può apparire deludente per chi ama cose astruse disdegnando quelle semplici, ovvie intuitive. E' una diminuzione pensare, con Bertrand, che il calcolo delle probabilità è nient'altro che il **il buon senso ridotto a calcolo** ?

E' forse una diminuzione dire che la formula di Bayes è ovvia ? No, certo. Allora con Savage (1962) possiamo dire che **the Bayesian theory is a common sense theory**.

6.2 Una esemplificazione pratica

Un esempio di carattere medico chiarirà il ruolo del teorema di Bayes e metterà in evidenza l'importanza delle probabilità iniziali.

E' noto che la sindrome di immuno deficienza acquisita (AIDS) è causata dal virus HIV. Sia H^+ l'ipotesi : un determinato individuo è HIV-siero positivo ". e supponiamo di poter eseguire un esame clinico (test ELISA) in grado di riconoscere se l'individuo è HIV-siero positivo. Indichiamo con $TEST^+$ l'evento: "l'esame eseguito sul soggetto ha dato esito positivo". Supponiamo sia noto che $TEST^+$ si verifica sempre nel caso in cui il soggetto sia HIV-siero positivo e nel β per cento dei casi con soggetti HIV-siero negativi (indichiamo con H^-). In base a queste informazioni possiamo assegnare le seguenti probabilità:

$$P(Test^+|H^+) = 1$$

$$P(Test^+|H^-) = \beta$$

Queste probabilità sono le verosimiglianze ed è proprio da tali quantità che la teoria classica di Neyman e Pearson pretende di fornire le proprie regole di comportamento. Tale teoria ci dice che, accettando l'ipotesi che l'individuo è HIV-siero positivo sulla base del risultato del test si corre il rischio di accettare una ipotesi (l'individuo è HIV-siero positivo) che potrebbe essere falsa. Il pericolo sta proprio nell'interpretare l'esperimento che dà luogo al risultato del test come una conferma di H^+. Se così fosse l'individuo dovrebbe agire come se fosse siero positivo sottoponendosi a cure anche se non fosse stato contagiato.

Viceversa l'interpretazione dell'esperimento è molto soddisfacente qualora si considerino le probabilità a priori delle ipotesi e si applica il meccanismo bayesiano.

In particolare se si suppone che nella popolazione vi è l' α per cento di persone HIV-siero positive ed il nostro individuo è un qualunque individuo della popolazione data si ricaverà mediante il teorema di Bayes:

$$P(H^+|Test^-) = \frac{P(H^+)P(Test^+|H^+)}{P(H^+)P(Test^+|H^+) + P(H^-)P(Test^+|H^-)}$$

da cui:

$$P(H^+|Test^+) = \frac{\alpha}{\alpha + (1-\alpha)\beta}.$$

Fissiamo $\beta = 0.015$ e supponiamo che da indagini svolte si sappia che in Italia vi è un individuo HIV-siero positivo ogni mille ossia:

$$\alpha = \frac{1}{1000}$$

allora si deduce immediatamente che

$$P(H^+|Test^+) = 0,0625586.$$

Questo valore fornisce la rilevanza diagnostica del test ELISA per qualunque individuo appartenente alla popolazione Italiana (di cui non si sappia se ha o no infezione da HIV). : in tal caso potremmo dire che solo il 6.25 per cento degli individui per cui il test ELISA ha dato luogo ad una diagnosi positiva risultano veramente HIV siero positive.

Supponiamo ora di essere a conoscenza che l'individuo in considerazione è un individuo tossicodipendente di Milano. In tal caso si ha:

$$\alpha = \frac{1}{3}$$

da cui

$$P(H^+|Test^+) = 0,970874$$

In questa situazione lo stesso risultato significa che nel 97 per cento dei casi quando il test ELISA fornisce una risposta positiva l'individuo è HIV-siero positivo.
Si può quindi osservare che le probabilità ottenute tramite l'impostazione bayesiana sono in perfetto accordo anche con l'interpretazione che si dà in pratica ai test clinici.

Questo semplice esempio mette in evidenza la forte dipendenza che esiste tra le valutazioni statistiche e le probabilità iniziali e non solo dai risultati di un esperimento.

In conclusione per risolvere i problemi basta avanzare considerazioni conformi al buon senso.

7 Il problema centrale dell'inferenza Bayesiana

Per poter apprezzare a pieno la logica e la natura distintiva del paradigma bayesiano è necessario considerare, a mio avviso. il problema della previsione di fatti futuri.
Nel 1920 Karl Pearson, a proposito di quello che egli chiamava "*il problema fondamentale della statistica pratica*" così si esprimeva:
Un evento si è verificato p volte su n = p+q prove. e non abbiamo alcuna conoscenza a priori della frequenza dell'evento nella popolazione totale delle sue manifestazioni. Qual è la probabilità che esso si verifichi r volte su m = r+s prove ulteriori ?.

Il passo riportato da Karl Pearson ci permette di affermare che usualmente il ricercatore fissa come obiettivo la valutazione della probabilità di un evento collegato a certi fatti futuri subordinatamente alle determinazioni di certe osservazioni.

Gli statistici si trovano sovente ad affrontare questo problema.

Ad esempio quando $(x_1, x_2, \ldots, x_n)$ è un insieme di n osservazioni di una successione $\{X_n\}$ e si vuole prevedere $y = x_{n+1}$ un futuro valore della successione.

La soluzione Bayesiana è quella di descrivere la connessione tra y e $(x_1, x_2, \ldots, x_n)$ mediante la probabilità $p(y|x_1, x_2, \ldots, x_n)$ ossia:

$$P(X_{n+1} = y|X_1 = x_1, \ldots, X_n = x_n) \tag{1}$$

Se la legge di probabilità della successione dei numeri aleatori $\{X_n\}$ è assegnata la determinazione della (1) segue gli usuali procedimenti del calcolo delle probabilità, vale a dire:

$$P(X_{n+1} = y|X_1 = x_1, \ldots, X_n = x_n) = \frac{P(X_1 = x_1, \ldots, X_n = x_n, X_{n+1} = y)}{P(X_1 = x_1, \ldots, X_n = x_n)}$$

a patto che $P(X_1 = x_1, \ldots, X_n = x_n)$ non sia nulla.

E' immediato osservare che per poter valutare la probabilità di eventi futuri sulla base di eventi passati (noti) occorrerà stabilire un legame logico di dipendenza tra tutti gli eventi (passati e futuri). Naturalmente i tipi di dipendenza tra enti aleatori sono infiniti ed il loro esame non potrà mai essere esaurito. Fortunatamente le situazioni in cui lo statistico è chiamato ad operare si lasciano spesso descrivere da schemi abbastanza semplici.

Giova osservare che uno schema che negherebbe l'essenza del ragionamento induttivo è quello di indipendenza stocastica. Infatti in condizioni di indipendenza stocastica si ha:

$$P(X_{n+1} = y|X_1 = x_1, \ldots, X_n = x_n) = P(X_{n+1} = y)$$

e quindi non si apprenderebbe nulla dall'esperienza passata.

Al di là dell'indipendenza si annoverano forme di dipendenza più o meno strette che possono variamente caratterizzare la dipendenza della previsione da quanto si è osservato. Alcuni esempi sono: la scambiabilità, la dipendenza markoviana, la stazionarietà, l'indipendenza degli incrementi, ecc...

Due esempi ci permetteranno di discutere e mettere in luce i punti salienti dell'impostazione bayesiana letta in una ottica previsiva.

Esempio 1 : Il lancio di una moneta.

Immaginiamo di giocare a testa e croce con una moneta.

Indichiamo con X_i la variabile che assume il valore $x_i = 1$ se si manifesta testa nell'iesimo lancio e $x_i = 0$ se si manifesta croce. Si supponga che si siano assegnate le probabilità $P(X_1 = x_1, \ldots, X_n = x_n)$ per ogni n =1,2,3,4,.....

Si è così in grado di determinare la probabilità che nell'(n+1)iesimo lancio si verifichi testa, nell'ipotesi che negli n precedenti si siano manifestate, nell'ordine k teste e (n-k) croci. Indicato con $(X_1 = x_1, \ldots, X_n = x_n)$ l'esito sperimentale sopra richiamato, la probabilità richiesta ha l'espressione:

$$P(X_{n+1} = y|X_1 = x_1, \ldots, X_n = x_n) = \frac{P(X_1 = x_1, \ldots, X_n = x_n, X_{n+1} = y)}{P(X_1 = x_1, \ldots, X_n = x_n)}$$

Tale espressione costituisce la risposta al problema posto da K. Pearson nel caso particolare in cui r = m = 1.

Esempio 2 : Estrazione da un'urna.

Immaginiamo di trovarci a pescare, successivamente e con reimmisione dopo ogni estrazione, palline da un'urna contenente, secondo una composizione non nota, palline bianche e non bianche. Indichiamo con X_i la variabile che assume $x_i = 1$ se si verifica l'evento: estrazione di una pallina bianca all'iesima prova e $x_i = 0$ altrimenti. Supponiamo di aver assegnato le probabilità congiunte di n variabili con n =1,2,3,4,.....

Con tali elementi siamo in grado di rispondere al problema di K. Pearson.

In entrambi gli esempi per poter calcolare la distribuzione predittiva è necessario specificare le probabilità congiunte: $P(X_1 = x_1, \ldots, X_n = x_n)$.

Usualmente gli Statistici assumono, in entrambi i casi, che esista un parametro θ, tale che, dato θ gli elementi della successione siano indipendenti ed identicamente distribuiti, cioè:

$$P(X_1 = x_1, \ldots, X_n = x_n|\theta) = \prod_{i=1}^{n} p(x_i|\theta)$$

Pertanto se $H(\theta)$ è la distribuzione del parametro θ si ha:

$$P(X_1 = x_1, \ldots, X_n = x_n) = \int \prod_{i=1}^{n} p(x_i|\theta) dH(\theta) \tag{2}$$

da cui è immediato dedurre, se H ammette densità h, che :

$$P(X_{n+1} = y|X_1 = x_1, \ldots, X_n = x_n) = \int p(y|\theta)h(\theta|x_1, \ldots, x_n)d\theta$$

con

$$h(\theta|x_1, \ldots, x_n) = \frac{h(\theta)\prod_{i=1}^{n} p(x_i|\theta)}{\int h(\theta)\prod_{i=1}^{n} p(x_i|\theta)d\theta} \tag{3}$$

Nella (3) $h(\theta|x_1, \ldots, x_n)$ è la densità finale del parametro mentre $h(\theta)$ indica la densità iniziale su θ e $\prod_{i=1}^{n} p(x_i|\theta)$ indica la verosimiglianza.

In questa logica, per ogni scelta di θ selezioniamo una funzione di densità che regola il meccanismo aleatorio sui dati sperimentali.

Ci troviamo pertanto di fronte ad una doppia incertezza:

a) intorno al valore di $\theta \in \Theta$;

b) intorno al dato sperimentale regolato dalla funzione di densità $p(x|\theta)$, una volta scelto θ.

Tale impostazione va sotto il nome di **impostazione ipotetica** e pone, come scopo primario dell'induzione statistica, l'individuazione del meccanismo aleatorio che genera i dati e, generalmente, non pone problemi circa la natura delle ipotesi associate al modello.

Osserviamo invece che problemi vi sono e sono inerenti la natura del parametro θ.

Infatti nel caso dell'urna θ ha effettivamente un significato fisico: dichiarare che un'urna contiene una proporzione di palline bianche è un fatto oggettivo direttamente verificabile. In linguaggio statistico, ogni precisata composizione prende il nome di ipotesi e, di conseguenza, si parlerà di ipotesi e di probabilità subordinata di una ipotesi. In tale situazione è dunque lecito parlare di probabilità di una ipotesi e di probabilità di un evento subordinato ad una ipotesi proprio grazie al fatto che le ipotesi sono eventi, ovvero fatti osservabili. Ne consegue che la determinazione di una ipotesi subordinata all'osservazione dell'esito di un dato numero di osservazioni può costituire l'oggetto principale della ricerca statistica. In tale situazione l'adozione dell'impostazione ipotetica non pone problemi sul piano logico.

Nel caso dell'urna detta $H(\theta)$ la funzione di ripartizione iniziale dell'ipotesi possiamo determinare $H(\theta|x_1,\ldots,x_n)$ la funzione di ripartizione finale dell'ipotesi. Se supponiamo, ad esempio, che la distribuzione iniziale per θ sia una Beta di parametri p e q allora la distribuzione finale è ancora una Beta di parametri $(p+\sum_{i=1}^{n} x_i)$ e $(n+q-\sum_{i=1}^{n} x_i)$.

E' immediato allora dedurre che :

$$P(X_{n+1}=1|X_1=x_1,\ldots,X_n=x_n)=\frac{p+\sum_{i=1}^{n} x_i}{n+p+q}$$

Se poi la distribuzione iniziale per θ è una distribuzione uniforme, ovvero p = q = 1 otteniamo che :

$$P(X_{n+1}=1|X_1=x_1,\ldots,X_n=x_n)=\frac{\sum_{i=1}^{n} x_i+1}{n+2}$$

Vale la pena osservare che se avessimo ottenuti tutti successi la probabilità di un ulteriore successo sarebbe:

$$P(X_{n+1}=1|X_1=x_1,\ldots,X_n=x_n)=\frac{n+1}{n+2}.$$

Questa è la regola di successione che Laplace (1812) ha utilizzato questa formula per calcolare la probabilità che il sole sorgerà domani dato che è sempre sorto.

Nell'esempio della moneta invece non si può pensare ad un insieme di ipotesi circa la presunta irregolarità della stessa ed ad un collegamento tra la probabilità di un evento e ciascuna di tale ipotesi. In simili circostanze non sarebbe corretto partire, per la risoluzione del problema induttivo, dall'impostazione ipotetica. In questo caso quando il parametro non ha significato fisico, sarebbe meglio chiedersi: avendo osservato un certo numero di risultati passati, cosa possiamo dire sull'esito di un futuro lancio? Per rispondere dobbiamo risolvere un problema predittivo.

In definitiva ci si può servire delle tecniche fondate sui modelli ipotetici solo quando le ipotesi corrispondono a fatti osservabili, vale a dire ad eventi per cui ha senso parlare di probabilità.

Con Roberts (1965) possiamo dire che occorre liberarsi dalla struttura delle ipotesi e porre l'attenzione sulle distribuzioni predittive. Pertanto da un punto di vista più generale riteniamo che l'attenzione debba essere concentrata sull'espressione della probabilità di eventi che riguardano le osservazioni future condizionate alle informazioni acquisite tramite le osservazioni del fenomeno al quale l'evento è legato. In tale situazione solo il passato ed il futuro di risultati sperimentali sono coinvolti, senza qualsiasi intermediazione di alcun modello. Non è necessario pertanto ricorrere all'impostazione ipotetica a patto che si sia potuta assegnare la legge del processo.

Ciò purtroppo non è sempre agevole.
Esiste uno schema, la scambiabilità, che rende semplice tale compito ed entro tale schema vi è una stretta connessione tra problema impostato in modo ipotetico e problema previsivo. Il concetto di scambiabilità fu introdotto da de Finetti (1930, 1937) con il preciso scopo di giustificare il metodo di induzione statistica. La scambiabilità infatti è lo strumento per presentare l'induzione come una naturale via di ragionare sulle probabilità di fatti osservabili evitando pseudoenti ed oscurità di carattere metafisico quali sono spesso i parametri che compaiono nell'impostazione ipotetica.

Definizione 2 *Una successione infinita di variabili aleatorie X_n si dice scambiabile, e equivalentemente le v.a. della successione verranno dette scambiabili, se, per ogni n, la funzione di ripartizione di $X_1, X_2, \ldots, X_n$ non dipende dall'ordine secondo il quale gli indici 1,2,...,n appaiono, ma soltanto da n*

Vogliamo sottolineare che se le v.a. della successione esprimono il risultato di una singola prova di un esperimento che si ritiene possa essere ripetuto un numero infinito di volte, allora affermare che la successione è scambiabile corrisponde all'ipotesi inferenziale per la quale non si ritiene importante ai fini dell'analisi dell'esperimento l'ordine secondo il quale i risultati appariranno. O equivalentemente vengono ritenuti egualmente probabili tutti i risultati su n prove che differiscono tra loro soltanto per l'ordine. La scambiabilita' della successione implica per la predittiva:

$$P(X_{n+1} = y|X_1 = x_1, \ldots, X_n = x_n) = P(X_{n+1} = y|X_1 = x_{i1}, \ldots, X_n = x_{in})$$

dove $i_1, i_2, \ldots, i_n$ è una permutazione di $1, 2, \ldots, n$.
Emerge chiaramente che la scambiabilità non è una proprietà intrinseca degli elementi della successione ma piuttosto uno schema teorico che il soggetto impone, in relazione ai propri convincimenti e alle proprie informazioni.

Il risultato principale che si dimostra per successioni di v.a. scambiabili è il seguente **teorema di rappresentazione** dovuto a de Finetti (1931, 1937):

Teorema 1 *Una successione di v.a. è scambiabile se e solo se esiste una funzione di ripartizione aleatoria F tale che, condizionatamente a F, le v.a. della successione sono indipendenti ed identicamente distribuite con funzione di ripartizione F .*

Il teorema di rappresentazione permette di giustificare l'impostazione bayesiana che considera successioni di v.a. indipendenti ed identicamente distribuite condizionalmente ad un ente aleatorio Θ. L'elemento aleatorio Θ diventa in questo caso la stessa funzione di ripartizione aleatoria F.

Assumendo la scambiabilità della successione si può allora dimostrare che deve esistere una funzione di ripartizione $H(\theta)$ tale che valga per le probabilità congiunte la seguente espressione:

$$P(X_1 = x_1, \dots, X_n = x_n) = \int \prod_{i=1}^{n} p(x_i|\theta)dH(\theta)$$

Come si può notare è la stessa espressione ottenuta nella (2) dove però avevamo scelto il modello e la distribuzione iniziale del parametro. Inoltre de Finetti ha dimostrato che $H(\theta)$ è la funzione di ripartizione limite verso cui converge all'aumentare del numero delle prove il processo delle funzioni di ripartizioni empiriche. La scambiabilità allora implica l'esistenza di un modello e di una distribuzione iniziale.

Si puoò facilmente dedurre la distribuzione predittiva dalla sola assunzione di scambiabilità:

$$P(X_{n+1} = 1|X_1, \dots, X_n) = \frac{\int p(y|\theta) \prod_{i=1}^{n} p(x_i|\theta)dH(\theta)}{\int \prod_{i=1}^{n} p(x_i|\theta)dH(\theta)} \tag{4}$$

Negli esempi considerati la successione $\{X_n\}$ è una sucecssione infinita di v.a. a valori in (0,1) ed è facile specificando l' espressione (4) dedurre che :

$$P(X_{n+1} = 1|X_1, \dots, X - n) = \frac{\int_0^1 \theta^y (1-\theta)^{1-y} \theta^{\sum_{i=1}^{n} x_i} (1-\theta)^{(n-\sum_{i=1}^{n} x_i)} dH(\theta)}{\int_0^1 \theta^{\sum_{i=1}^{n} x_i} (1-\theta)^{(n-\sum_{i=1}^{n} x_i)} dH(\theta)}$$

e pertanto:

$$P(X_{n+1} = 1|X_1, \dots, X_n) = \int_0^1 \theta^y (1-\theta)^{1-y} dH(\theta|x_1, \dots, x_n)$$

come si vede tutto dipende dalla distribuzione condizionale di Θ date $(X_1, \dots, X_n)$ ossia dalla distribuzione finale.

Se ora vogliamo conoscere la probabilità di un futuro successo quando conosciamo gli esiti di n prove passate otteniamo:

$$P(X_{n+1} = 1|X_1 = x_1, \dots, X_n = x_n) = E(\Theta|x_1, \dots, x_n)$$

Analogamente si ha:

$$P(X_{n+1} = 0|X_1 = x_1, \dots, X_n = x_n) = 1 - E(\Theta|x_1, \dots, x_n)$$

Possiamo chiederci : è possibile partendo da una legge di probabilità del processo delle osservazioni pervenire alla costruzione di un modello ipotetico ?

La risposta è affermativa quando si richieda alla legge del processo di soddisfare il requisito della scambiabilità. Si assuma, ad esempio, inizialmente la legge di probabilità del processo nel modo seguente:

$$P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n) = B(\sum_{i=1}^{n} x_i + 1, n + 1 - \sum_{i=1}^{n} x_i)$$

E' immediato verificare, in questo caso, che la legge di probabilità è scambiabile perchè è invariante alle permutazioni degli argomenti. Allora per la prima parte del teorema di rappresentazione di de Finetti esisterà certamente una legge di probabilità $H(\theta)$ che riproduce la distribuzione del processo . E' facile verificare che, in tale circostanza, tale distribuzione iniziale è una distribuzione uniforme su (0,1). In conclusione si può dire che in questo esempio supporre la legge di probabilità scambiabile sui fatti osservabili equivale ad impostare il problema in forma ipotetica con una distribuzione iniziale uniforme.

8 La convivenza pacifica

Sono molti gli studiosi che auspicano una pacifica convivenza tra bayesiani e frequentisti. Se si trattasse di guerre di religione approverei questo atteggiamento ma purtroppo la discussione verte su problemi logici. Pertanto *poichè il contrasto ha natura logica, ciò non è sostanzialmente possibile*. Si tratta allora di sviluppare con coerenza il punto di vista adottato e di definire con chiarezza anche i livelli di flessibilità rispetto a schemi ideali che non sempre sono perfettamente realizzabili.

9 Conclusioni

Mi preme sottolineare che il dibattito sui fondamenti del ragionamento induttivo è molto ricco e non può essere ridotto alla contrapposizione tra frequentisti e bayesiani. Entrambe le impostazioni presentano una notevole articolazione interna e vi sono moltissimi problemi aperti che investono entrambe le metodologie. Esistono inoltre altri stimolanti punti di vista alternativi.

Abbiamo indicato in precedenza come un elemento essenziale dell'inferenza statistica sia l'uso della probabilità per rappresentare l'incertezza. La questione ha qui natura filosofica e si connette evidentemente alla concezione adottata per la probabilità. I sospetti manifestatesi verso il punto di vista soggettivitico erano caratterizzati dal timore che, aderendo ad esso, non fosse possibile pervenire ad una teoria matematica della probabilità. L'opera pioneristica in questo campo di Bruno de Finetti ha individuato la possibilità di dare assetto logico-formale alle leggi che regolano la probabilità partendo dal suo contenuto soggettivo ed ha mostrato che molti dubbi, che ne hanno caratterizzato (e caratterizzano) l'applicazione, svaniscono se ci si attiene scrupolosamente al punto di vista soggettivistico.

I problemi che coinvolgono la statistica vertono, nella relazione abbiamo citato solo i principali, su quale possa essere l'ipotesi "vera" nell'insieme delle alternative possibili o su quale possa essere il risultato della prossima osservazione. Ho provato a porre in evidenza come l'unica risposta veramente adeguata è di fornire una **distribuzione di probabilità** sulle possibili alternative. Il compito della Statistica è di aiutare a costruire tale risposta. Non mi sembrano accettabili risposte non collegabili a questo tipo di logica, altrimenti, come ha scritto Tukey (1961), si rischia di dare *risposte precise a un problema sbagliato*. Per ovvi motivi di spazio ho tralasciato altri argomenti che sono spesso oggetto di polemica ma che considero di minor rilievo. Vorrei concludere dicendo che tutte le procedure inferenziali ad un certo punto fanno leva su elementi esterni al modello e che hanno un carattere in una certa misura soggettivo. Nessun paradigma può pretendere di essere **oggettivo** l'importante che le assunzioni e gli elementi soggettivi siano esposti e non, come dice Good (1976) *nascosti sotto il tappeto*. Vale la pena infine sottolineare che, molto spesso, anche le tecniche bayesiane vengono considerate come espedienti meramente formali ed in tal caso non sono più degne di fede di ogni altro strumento (o metodo ad hoc) presente nella statistica oggettivistica.

Bibliografia

BAYES, T (1763), " An essay towards solving a problem in the doctrine of chances" Pubblicato postumo su *Phil. Trans. Roy.Soc. London*, 53, 296-325. Ripreso in *Biometrika* ,45, (1958), 293-315 , con una nota biografica di G.A. Barnard.

BERNOULLI, D (1738). " Specimen Theoriae Novae de mensura sortis" *Commentarii Academiae Scientiarum Imperialis Petropolitanae*. 5. 175-192. (traduzione dal latino in inglese di L. Sommer: Exposition of a New Theory on the Measrument of Risk , *Econometrica* ,22, (1954). 23-35.

DE FINETTI, B (1931). " Sul significato soggettivo della probabilità " *Fundamenta Mathematicae*, vol 17, 298-329.

DE FINETTI, B (1937), " La prevision: ses lois logiques, ses sources subjectives" *Ann. Inst. H. Poincarè*, tome VII, 1- 68.

DE FINETTI, B (1938), " Resoconto critico del colloquio di Ginevra intorno alla teoria della probabilita"' *Giornale dell'Istituto Italiano degli Attuari*. 9, 3-42.

DE FINETTI, B (1938), " Sur la condition d'equivalence partielle" *Act.Sc.Ind.*,739, Herman, Paris.

DE FINETTI, B (1959), " La probabilità e la statistica nei rapporti con l'induzione, secondo i diversi punti di vista " *Atti del Corso CIME su Induzione e Statistica*. Varenna, 1- 115.

DE FINETTI, B (1970), *Teoria della probabilità. Sintesi introduttiva con appendice critica*, Einaudi, Torino.

DE FINETTI, B (1971), " Probabilità di una teoria e probabilità dei fatti " *Studi di Probabilità, Statistica e Ricerca Operativa in onore di Giuseppe Pompilj*, Oderisi, Gubbio, 86-101.

DE FINETTI, B (1973), " Bayesianesimo: il suo ruolo unificante per i fondamenti e le applicazioni della Statistica " *Bulletin of the International Statistical Institute*, Proceedings of the 39th session,vol. 4,349-368.

DE FINETTI, B (1977), " Inferenza statistica Bayesiana" Relazione al Convegno su " I fondamenti dell'inferenza statistica", Firenze.

DE MORGAN, A (1847), *Formal Logic*, Taylor e Walten, London

FISHER ,R.A. (1955), "Statistical Methods and Scientific Induction" *J. Roy.Statist.Soc.B*

GOOD, I.J. (1976), " The Bayesian influence, or to sweep subjectivism under the carpet " *The Foundation of Probability Theory, Statistical Inference and Statistical Theories of Science* (W.L. Harter e C.A. Hooker eds.) Reidel, Dordrecht.

KOLMOGOROV, A (1950). *Foundations of the theory of Probability*. Chelsea Publ. Co. , New York.

JEFFREYS,H. (1939), *Theory of Probability*, Clarendon Press. Oxford.

LAPLACE, P.S. (1795), *Essai Philosophique sur les Probabilites*. trad.itatialan in " Opere", UTET ; Torino, 1967.

LAPLACE, P.S. (1812), *Theorie Analytique des Probabilites*. Courcier , Paris, trad.italiana in " Opere " , UTET ; Torino. 1967.

LINDLEY, D.V. (1953), " Statistical Inference " *J. Roy.Statist.Soc.B* 15, 30-76.

LINDLEY, D.V. (1990), " The present position in Bayesian Statistics" *Statistical Science*, 5, 44-89.

POINCARE, H. (1902), *La Science et l'Hypothese*, Flammarion, Paris.

ROBERTS, H.V. (1965), " Probabilistic prediction " *J. Amer. Statist. Assoc.* 60, 50-62.

SAVAGE, L.J. (1962), " Bayesian Statistics " in *Recent Developments in Information and Decision Processes* (R.E. Machol e P. Gray eds.) Macmillan and Co. N. Y. 161-194

SAVAGE, L.J. (1954), *The Foundations of Statistics* J. Wiley, New York.

STIGLER, S.M. (1986), *The History of Statistics* Harvard, M.A. University Press

TUKEY, J.W. (1961), " The future of data analysis " *Ann. Math. Statist.* 33, 1-67.

WALD, A. (1950), *Statistical Decision Functions*. J.Wiley, New York.

VON MISES, R (1928), *Probability, Statistics and Truth*. Ristampato nel 1957, Macmillan, London.