

IL CONTROLLO DELLE IPOTESI STATISTICHE: TEORIE A CONFRONTO

Paola Monari

1. INTRODUZIONE

Nella loro intensa attività di studiosi, Amato Herzel e Alighiero Naddeo hanno offerto agli statistici italiani un solido punto di riferimento metodologico e un costante richiamo alla riflessione critica. Tra il 1960 e il 1990, la statistica italiana ha visto una fervida ripresa dei più originali contributi di Scuola anglosassone in tema di inferenza; e proprio a questo argomento, i nostri Autori hanno dedicato gran parte delle loro ricerche, con particolare enfasi alla teoria dei test d'ipotesi secondo l'impostazione di Neyman e Pearson, e con la sicura convinzione che quella teoria fosse risolutiva delle ambiguità e delle contraddizioni intrinseche in ogni altra diversa proposta.

Non di meno, chiamato a un ripensamento critico sullo stato dell'arte, Naddeo ha scritto: "Avrei desiderato inquadrare in modo compatto le varie metodologie inferenziali della scuola classica ed approfondirne la ratio sottostante. Ma al momento di concretare quest'intenzione per i problemi specifici della stima e della verifica di ipotesi mi sono trovato di fronte a una pluralità di criteri di ottimizzazione difficilmente unificabili. E quando ho voluto esplorare la filosofia ispiratrice di questi criteri ho dovuto ripiegare su parole come 'ragionevole', 'giustificabile' che non avrei mai voluto usare perché non mostrano alcun carattere di necessità logica." (A. Naddeo, *Impostazione dei problemi di inferenza statistica classica*, 1978).

Con una scelta più pragmatica, Amato Herzel ha preferito invece sfuggire al coinvolgimento logico e filosofico sull'inferenza statistica e si è collocato nel più sicuro contesto della statistica matematica e del campionamento probabilistico, premessa ineliminabile delle metodologie inferenziali classiche, ricca di sviluppi formali, ma essenzialmente estranea al vivace – a volte troppo vivace – dibattito sull'induzione.

Oggi, tuttavia, ripensando all'insegnamento di questi illustri Maestri della statistica italiana, e sollecitati dall'iniziativa della rivista "Statistica" e del suo Direttore, il desiderio di riprendere alcune riflessioni sopravanza il timore di avventurarsi su un terreno assai insidioso.

2. IL PRINCIPIO DI FALSIFICAZIONE

Nell'evoluzione del pensiero moderno sul ragionamento induttivo e sul trattamento dell'incertezza, l'atteggiamento scientifico verso le metodologie statistiche ha da sempre perseguito con alterna fortuna due obiettivi fondamentali, presenti anche nel pensiero comune: la individuazione delle asserzioni "false", per eliminare tutto ciò che è inessenziale alla descrizione di un sistema fenomenico complesso, e la scelta, tra le asserzioni verosimili, di quelle più congruenti con il sistema. Questa distinzione si fonda su un principio di non contraddittorietà, secondo il quale un sistema si definisce non contraddittorio quando è in grado di distinguere, tra le asserzioni possibili, quelle compatibili col sistema da quelle non compatibili. La dicotomia apparentemente semplice tra i due momenti inferenziali si fa spesso sfumata, anche per la difficoltà di formulare le asserzioni che spiegano il paradigma scientifico di riferimento, soprattutto quando esse sono di tipo statistico.

Di fronte a un problema conoscitivo, il ricercatore tenta qualche soluzione: una teoria, un'asserzione, un'ipotesi statistica. Enunciati sempre provvisori, da sottoporre a severi controlli orientati alla falsificazione dell'enunciato stesso. Se, come scrive Popper, ciò che "...spinge il teorico a cercare una teoria migliore è quasi sempre la falsificazione sperimentale di una teoria fino a quel momento accettata o corroborata" (K. Popper, *Logik der Forschung*, 1934), il metodo per "prova ed errore" sembra assumere i tratti caratteristici del metodo scientifico. Ma già nelle parole di Popper emerge un primo motivo di ambiguità tra processo di formulazione di una teoria, che è momento di scoperta non riconducibile a canoni codificati, e processo di controllo, riconducibile a rigide procedure di confutazione.

Nella prima fase di fondazione di una teoria per approssimazioni successive, più asserti, tutti compatibili con la teoria, possono essere messi in competizione tra loro attraverso opportuni criteri di scelta. Il problema metodologico sta nella demarcazione tra rifiuto di un enunciato perché "falso" o rifiuto di un enunciato perché "poco probabile" nel confronto con gli altri. Il criterio di riconoscimento tra falsificazione ed esclusione, a mio parere, sta soltanto nelle procedure inferenziali adottate.

La falsificazione è sempre di tipo ipotetico-deduttivo (se H , allora T ; $\neg T$, allora $\neg H$) e si fonda su una logica a due valori – "vero" o "falso" – in cui l'enunciato conclusivo non ha alcun contenuto di probabilità. La scelta per esclusione successiva si fonda invece su una logica della probabilità intesa a dare gradi di fiducia alle asserzioni possibili.

Quando gli enunciati sono di tipo statistico, il canone metodologico della scienza del Novecento, cresciuto sui principi del più tradizionale metodo sperimentale, ha continuato a ispirarsi al confronto tra ipotesi e dato d'osservazione, anche se le soluzioni tecniche hanno dovuto adattarsi al nuovo modo di intendere la variabilità interna ai fenomeni.

3. LA TEORIA DELLA SIGNIFICATIVITÀ

L'ideale oggettivista, storicamente assunto a garanzia di universalità della ricerca scientifica, ha suggerito le prime soluzioni tecnico-formali del problema dell'induzione su proposizioni statistiche, che si sono essenzialmente ispirate al più tradizionale schema ipotetico-deduttivo. E proprio su questo schema si è fondata la "teoria della significatività", della quale Ronald A. Fisher ha dato una compiuta teorizzazione (R.A. Fisher, 1922, 1956).

Il criterio fisheriano di rifiutare un'ipotesi sulla base della sola verosimiglianza è riconducibile all'asserzione ipotetica "se H , allora T , con probabilità p ", essendo T l'insieme dei risultati compatibili con H a cui viene associata una variabile aleatoria X , con distribuzione di probabilità $p(X|H)$. Nessun risultato dell'insieme T è formalmente in contraddizione con l'ipotesi H . La confutazione di H deve quindi avvalersi di criteri convenzionali in cui non v'è alcuna necessità logica, bensì solo opportunità empirica (I. Scardovi, 1977, 1984). Per giustificare il rifiuto della particolare ipotesi sotto controllo, Fisher propone infatti il principio di disgiunzione semplice: o si è avverato un evento raro, o l'ipotesi H non è vera.

In questa impostazione, l'ipotesi viene rifiutata quando i risultati sperimentali, E , cadono in una circoscritta regione di T a cui è assegnata una piccola probabilità, convenzionalmente definita. Ribadisce Fisher, "Quando parliamo della probabilità che un certo oggetto soddisfi una certa condizione, immaginiamo di dividere tutti gli oggetti in due classi a seconda che soddisfino, oppure no, quella condizione. Quest'ultima è la loro unica caratteristica di cui prendiamo conoscenza." (R.A. Fisher, *On the mathematical foundation of theoretical statistics*, 1922).

La teoria fisheriana della significatività si riferisce a una precisa ipotesi – l'ipotesi H_0 di completa accidentalità dei risultati osservati, all'interno di un sistema teorico di riferimento – e la controlla in una visione squisitamente falsificazionista, nel confronto con l'evidenza empirica.

A questo proposito, lo stesso Fisher riconosce la difficoltà di proporre teorie e metodi scevri da un immanente alone di oscurità e ammette che nell'inferenza statistica rimangono irrisolti paradossi fondamentali. Tuttavia, se la forma matematica è specificata, se sono ben individuati i metodi per stimare i parametri che definiscono la popolazione ipotetica e si conosce la distribuzione delle quantità (statistiche) che questi parametri assumono nei diversi campioni casuali tratti da quella popolazione, allora – dice Fisher – l'aspetto teorico del trattamento di un particolare insieme di dati è completamente chiarito (Fisher, 1922).

È solo al termine di questo processo logico che Fisher colloca il problema del controllo di un'ipotesi statistica, come percorso a ritroso per saggiare gli assunti ipotetici via via posti. Le procedure di controllo potrebbero consentire di falsificare statisticamente il modello assunto a descrivere la popolazione o, in subordine, il valore (stimato) da attribuire ai parametri di quel modello. Pertanto, l'ipotesi zero assume la forma di un'asserzione base assunta come *experimentum crucis*. Se falsificata, il percorso induttivo teso alla formulazione di una teoria scientifica procede; altrimenti si ferma, almeno provvisoriamente.

Il risultato di un test di significatività porta al rifiuto o al non rifiuto dell'ipotesi zero, ma l'enunciato conclusivo non ha alcun contenuto di probabilità. Il livello di probabilità assunto come soglia per il rifiuto o il non rifiuto non pretende di dire nulla sull'ammissibilità dell'ipotesi: misura soltanto, attraverso l'errore di prima specie, la forza falsificante della procedura adottata. In questo contesto inferenziale, l'ipotesi è dunque un'ipotesi *ad hoc*, formulata per cogliere aspetti particolari, quasi sempre ridotti a modelli, della più complessa struttura teoretico-osservativa del ricercatore. Ne segue che più ipotesi zero possono essere messe in atto per controllare le varie implicazioni di una teoria nel confronto con i dati d'osservazione: alcune possono essere falsificate, altre no. L'insieme di queste risposte determina il grado di corroborazione della teoria stessa.

Rispetto alla logica classica, nelle procedure statistiche di tipo falsificazionista, il momento differenziante si ritrova nell'incertezza dell'enunciato conclusivo nei termini del principio di disgiunzione semplice. La risposta del test lascia spazio a una scelta convenzionale che diventa momento decisionale del ricercatore. Infatti, tutti i controlli di una teoria, rivolti alla sua falsificazione, devono arrestarsi a qualche asserzione base che si decide di accogliere come assunto di principio non discutibile. Ma, in termini di coerenza logica, il processo di falsificazione non ha una conclusione naturale perché ogni asserzione base potrebbe, a sua volta, essere posta sotto controllo. Secondo Popper: "La procedura si arresta quando si forma una sorta di accordo psicologico tra i ricercatori sulla falsificazione... In questo senso, le asserzioni base sono convenzioni" (Popper, 1934). Le osservazioni di Popper valgono, a maggior ragione, per il controllo o la corroborazione degli assunti fondati su leggi statistiche, dove l'intrinseca incertezza di ogni enunciato induttivo fa della prova e della disprova una scelta convenzionale (Scardovi, 1987).

4. LA TEORIA DEI TEST D'IPOTESI

Il desiderio (forse inconsapevole) di non negare una risposta al vuoto lasciato da un'ipotesi zero rifiutata, non consentiva di arrestare il ragionamento inferenziale alle sole procedure di falsificazione. L'assunto della impossibilità di risolvere il problema dell'induzione al di fuori di una sintassi bayesiana e l'illusione di disporre di procedure inferenziali idonee a discriminare tra una pluralità di ipotesi hanno offerto un primo spunto per soluzioni alternative. Neyman e Pearson (1928) hanno contrapposto così alla metodologia fisheriana una metodologia induttiva fondata sulla scelta tra ipotesi a confronto in termini di scelta conveniente, sulla base della sola evidenza empirica e hanno proposto nuove regole di controllo induttivo fondate sulla premessa che nessun test può dare evidenza alla "verità" o alla "falsità" di una singola ipotesi senza il confronto con una alternativa possibile. I due studiosi avevano ben chiara la distinzione tra problema inferenziale diretto e inverso. "Ci sono due metodi distinti di porre il problema, – scrivevano già nel 1928 – uno parte dalla popolazione P , e chiede qual è la probabilità che un campione C possa essere stato estratto da essa, e l'altro è il metodo inverso che parte da C e cerca la probabilità che P sia la popolazione campionata" (J. Neyman, E.

Pearson, *On the use and interpretation of certain test criteria for purposes of statistical inference*, 1928).

Consapevoli dell'impossibilità di dare una risposta soddisfacente alla scelta tra ipotesi senza riaccendere il dibattito sulla probabilità inversa, parzialmente sopita dalla forza dirompente delle soluzioni fisheriane, tutte concentrate solo sulla falsificazione, Neyman e Pearson propongono la loro teoria dei test di ipotesi come strumento convenzionale di comportamento per guidare la scelta solo sulla base dei rischi d'errore (misurabili) in cui il decisore può incorrere. La soglia al di sotto della quale si decide di respingere H_0 dipende dalla scelta del valore convenzionale che si intende assegnare alla probabilità di rifiutare come "falsa" un'ipotesi "vera" (errore di prima specie) e alla probabilità di accettare come "vera" un'ipotesi "falsa" (errore di seconda specie). Questa distinzione e il conseguente controllo dei due possibili rischi d'errore sono divenuti concetti fondamentali della teoria stessa, volutamente pensati per eludere il confronto tra probabilità *a posteriori* (Neyman, Pearson, 1933 b).

Per aggirare la difficoltà logica di giustificare l'accettazione di un'ipotesi, Neyman e Pearson (1928) hanno opportunamente ripreso il concetto fisheriano di verosimiglianza nel loro famoso criterio del rapporto di verosimiglianza e hanno affidato la scelta tra ipotesi alla soluzione metodologica (test) che consente di individuare l'ipotesi "più verosimile" al confronto con l'evidenza empirica (campione): ossia, l'ipotesi che, a parità di α (errore di prima specie), rende massima la potenza del test (ovvero minimizza l'errore di seconda specie β) (Neyman, Pearson, 1933 a).

Sul piano epistemologico, Neyman e Pearson, sono consapevoli della difficoltà di giustificare la decisione strategica del rifiuto e della conseguente accettazione delle ipotesi a confronto e, non potendo più rifarsi al principio fisheriano di disgiunzione semplice – che è solo falsificazionista – propongono una suggestiva argomentazione statistica, riassunta col termine *long run*. Non si cerca la verità della singola conclusione: si vuole piuttosto contenere al di sotto di un prefissato livello convenzionale la frequenza di errore in una lunga serie di decisioni. Una svolta pragmatica che ha portato a esaltare situazioni sperimentali od osservative ripetibili nel tempo o nello spazio così da favorire l'analogia, anche formale, con le procedure di campionamento ripetuto (Herzel, 1991). In questo ambito, ogni test statistico si risolve in una conclusione probabile che introduce un elemento di prudenza nell'accettazione dell'enunciato. Tale elemento consiste nell'affermare che la conclusione è stata raggiunta con un metodo che, nel lungo andare, porterà a respingere l'ipotesi "non più di una volta su cento", quando è vera, mentre consentirà di respingerla "abbastanza spesso" quando è falsa (I. Neyman, E.S. Pearson, 1928). Si va così sostituendo al ragionamento induttivo una sorta di comportamento induttivo.

L'argomento del *long run* potrebbe essere assunto a descrivere in termini statistici il tentativo di Peirce (1923) di giustificare l'atto induttivo: un criterio pragmatico per controllare la validità di una regola fondata sulla proporzione di decisioni che da premesse "vere" conducono a conclusioni "vere". Secondo Peirce, un'inferenza è valida se il criterio che la determina è tale da dar luogo, più spesso, a conclusioni "vere" indipendentemente dalla "verità" o "falsità" di ogni singola

conclusione. In una interpretazione estensiva, questa regola pretende di attribuire forza probante al lungo andare delle congetture umane, anche quando non è configurabile un lungo andare fenomenico.

5. ALCUNE INCOERENZE

La giustificazione induttiva di Peirce sembra quindi includere coerentemente anche il principio del *long run*: esso consente di configurare il test statistico come uno strumento che può essere applicato più volte, a parità di condizioni, in situazioni campionarie simili, nell'intento di decidere tra due ipotesi. Una giustificazione che, nell'uso, ha finito per trascendere il suo ruolo contingente e per imporsi come elemento essenziale all'interpretazione logica di ogni scelta tra ipotesi. Questa argomentazione, ripresa successivamente da Braithwaite (1955) e da Hacking (1965), sembra risolversi in un utile criterio per la scelta di un metodo induttivo tra tutti quelli ragionevoli per affrontare un determinato problema inferenziale, ma di per sé non basta a sostenere criteri induttivi mal fondati nei postulati iniziali.

Anche a proposito delle tecniche di Neyman-Pearson si discute ormai da tempo proprio sulla ragionevolezza delle premesse teoriche, che è condizione necessaria per l'uso del criterio del lungo andare. Non solo: nella teoria dei test, le ipotesi sotto controllo sono di natura "statistica", attengono alla forma distributiva o ai parametri di una popolazione, non al risultato probabile di un singolo evento, come molti tendono a credere. Pertanto lo stesso controllo a posteriori della "verità" o della "opportunità" delle conclusioni non sembra logicamente proponibile e non consentirebbe di accertare se la proporzione delle conclusioni "vere" è sufficientemente elevata, oppure no.

Lo sviluppo formale della teoria dei test d'ipotesi in termini comportamentali va dunque rivisitato, ed è proprio sulle premesse che conviene fermare l'attenzione. Per un uso corretto delle procedure di Neyman-Pearson in termini di accettazione, si deve assumere che l'una o l'altra delle ipotesi alternative sia "vera" e si deve anche escludere che esse siano entrambe "vere" o entrambe "false". Tali condizioni non sono sempre fenomenicamente motivabili e spesso sfuggono all'attenzione dello statistico: esse sono tuttavia necessarie per una giustificazione coerente dei successivi sviluppi formali nel controllo dei due rischi d'errore. E' ovvio che il mancato rifiuto di un'ipotesi non implica la sua accettazione, così come l'avvenuto rifiuto non implica necessariamente l'accettazione dell'ipotesi alternativa, a meno di non postulare che una delle due debba essere comunque "vera".

La mancanza di chiarezza nelle premesse può introdurre gravi ambiguità nel metodo: si pensi a tutte le situazioni in cui l'ipotesi alternativa (semplice o composta) non si identifica con la negazione logica dell'ipotesi zero. La storia della scienza ricorda non poche ipotesi falsificate al controllo empirico, a cui non si contrappose alcuna alternativa: ipotesi che non furono immediatamente sostituite da altre ipotesi di ricerca. Lo stesso Lindley si chiede quale potrebbe essere l'alternativa al modello della relatività einsteiniana, convinto che, nella pratica della ri-

cerca, lo scienziato – come lo statistico – più spesso usa una singola ipotesi e non sempre considera le alternative (Lindley, 1990).

La teoria dei test d'ipotesi non approfondisce questi assunti preliminari, che non risultano sufficientemente discussi dagli autori. Di fatto, gli esempi riportati nei loro saggi si riconducono quasi sempre al controllo dell'ipotesi H_0 e pongono come alternativa l'ipotesi $\neg H_0$ (ossia l'ipotesi di non completa accidentalità dei risultati sperimentali, intesa come negazione logica di H_0): cosicché $p(H_0) + p(\neg H_0) = 1$. Ma è evidente che si tratta di un'alternativa convenzionale, che non si sottrae alla logica classica a due valori (“vero”-“falso”) e si risolve quindi in un artificio che consente soltanto di rinviare ogni più precisa formulazione dell'ipotesi di lavoro.

Da un attento confronto con la teoria fisheriana della significatività, emerge per altro che i tests d'ipotesi non sono mai stati pensati per l'accettazione diretta di ipotesi, in quanto non si discostano dalla logica del rifiuto: infatti, solo a seguito del rifiuto dell'ipotesi sotto controllo, essi conducono ad accettare, per esclusione, l'ipotesi alternativa sulla base di un enunciato che non assegna alcuna probabilità finale all'ipotesi “salvata”: la scelta avviene esclusivamente sul confronto tra rischi d'errore.

Queste considerazioni lasciano intravedere la difficoltà di una giustificazione epistemologica coerente di un criterio induttivo congegnato per una scelta che elude la soluzione bayesiana. E tuttavia, la teoria dei tests, negli intenti dei suoi ideatori, si proponeva appunto di aggirare pragmaticamente lo scoglio bayesiano, suggerendo un comportamento induttivo volto a garantire un'alta probabilità di respingere un'ipotesi “falsa” e di accettare un'ipotesi “vera”.

Ma una cosa è sottoporre a controllo un'ipotesi per provare il suo resistere a un'eventuale confutazione, altra cosa è decidere di accettare una particolare ipotesi compresa in una pluralità di alternative. E nessuna giustificazione fondata sul numero di successi, ossia di conclusioni “vere” tratte da premesse “vere”, potrà mai dare forza induttiva a un canone che non assicura la fondatezza delle sue premesse (Monari, 1984).

6. L'INTERPRETAZIONE BAYESIANA

La teoria di Neyman-Pearson ha finito così per perdere di generalità: estranea al canone falsificazionista, respinta dalla logica bayesiana, essa può essere ragionevolmente adottata solo quando è possibile configurare empiricamente sequenze di “esperimenti” replicati su di un'assegnata popolazione, purché le ipotesi a confronto esauriscano lo spazio delle possibilità. Quando non è possibile garantire queste condizioni, nulla si può dire sulla verità o falsità delle conclusioni.

Malgrado i tentativi di svincolare l'uso dei tests da ogni valutazione *a priori*, è proprio lo schema bayesiano a sottolineare la fragilità logica di una scelta tra alternative che prescinda dal grado di ammissibilità delle stesse.

Qual è allora il criterio di scelta razionale davanti a una pluralità di asserzioni o di ipotesi sopravvissute al processo di falsificazione? La risposta è antica, anche se le soluzioni tecniche sono recenti, e rimandano ad un uso più articolato e più critico del calcolo delle probabilità. I motivi storici che negli anni venti e trenta hanno tenuto lontano gli statistici dalle opportunità offerte dalla logica probabilistica – e, in particolare, della cosiddetta *inverse probability* – sono da tempo superati, e nessuno oggi nega il ruolo delle procedure bayesiane semplicemente richiamandosi ad una sterile disputa sul concetto di probabilità, anche se le diverse interpretazioni della probabilità generate dalle diverse situazioni conoscitive non possono non condizionare la costruzione dei modelli e la lettura dei risultati inferenziali.

Come è noto, dato un insieme esaustivo e disgiuntivo di ipotesi, tutte compatibili con una medesima evidenza empirica E , ciascuna dotata di un proprio grado iniziale di plausibilità misurabile in termini di probabilità, ad ogni ipotesi – $H_1, H_2 \dots, H_i, \dots$ – si può associare una probabilità *a posteriori* $p(H_i|E)$ che ne stimi il grado di ammissibilità nel confronto con l'evidenza, in funzione delle sue componenti: la probabilità iniziale $p(H_i)$ e la probabilità condizionata $p(E|H_i)$, quest'ultima sostituibile, a certe condizioni, con una verosimiglianza. Si può quindi costruire una sequenza ordinata di ipotesi in termini di probabilità "finali", così che, in assenza di altri elementi decisionali, l'asserzione più probabile possa assumere il ruolo di spiegazione provvisoria e semplificata di un fenomeno o di un suo aspetto.

La consuetudine degli statistici a trattare con disinvoltura verosimiglianze del tipo $p(E|H_i)$ (soprattutto quando H_i si identifica con H_0) ha portato a stimare le probabilità *a posteriori* non direttamente, ma tramite la mediazione delle probabilità iniziali $p(H_i)$.

Questa via ha consentito di raggiungere risultati brillanti nei problemi di stima dei parametri incogniti di un modello, pur lasciando aperta la discussione sulla scelta della distribuzione iniziale; ha trovato invece difficoltà nella formalizzazione della scelta tra ipotesi. Se, nei termini più generali del ragionamento bayesiano, i due problemi convergono in un unico momento induttivo, nella formulazione tecnica seguono percorsi diversi. La differenza fondamentale è che nella stima di un parametro la probabilità iniziale è diffusa e si identifica in una distribuzione $p(\theta)$, così da generare, combinata con la verosimiglianza, una distribuzione di probabilità finale; mentre nella scelta tra ipotesi, la probabilità iniziale è tutta concentrata sulle sole alternative possibili (ad esempio, $H_1: \theta = \theta_1$ e $H_2: \theta = \theta_2$).

La questione non è irrilevante perché riapre il dibattito sulla valutazione quantitativa di $p(H)$, che modifica la propria struttura trasformandosi da distribuzione a misura puntuale.

Il tentativo di eludere l'ostacolo riducendo le procedure bayesiane per la scelta tra ipotesi a problemi di inferenza predittiva da accadimenti noti a eventi attesi (B. de Finetti, 1971; D. V. Lindley, 1990) modifica sostanzialmente l'obiettivo dell'induzione, confondendo all'interno della ricerca scientifica finalità descrittive e finalità esplicative, e trova le sue radici nella regola di successione di Laplace. Prevedere un evento non ancora accaduto sulla base delle informazioni acquisite è cosa diversa dall'attribuire gradi di ammissibilità a ipotesi alternative (I. Scardovi, 1983).

Il successo di un'inferenza predittiva può corroborare un'ipotesi, ma non consente di ordinare un insieme di alternative in termini di probabilità finali. Non rimane quindi che affrontare la stima di $p(H_0 | E)$ attraverso la valutazione delle sue componenti raccogliendo la provocazione che Lindley lancia agli statistici bayesiani quando scrive: "Noi siamo come i geometristi che studiano lo spazio e mai lo misurano. Dove sono gli osservatori tra noi? Tanti lavori sono stati fatti sulla misura delle distanze. Dov'è l'equivalente nelle probabilità?" E ancora: "Si tratta di chiarire se le procedure bayesiane vogliono dedicarsi alla probabilità come grado di fiducia e quindi diventare strumento per l'analisi delle opinioni oppure, lavorando in ambito statistico, vogliono occuparsi di analisi dei dati. In questo caso, devono cimentarsi coi dati e tentare di misurare le probabilità delle asserzioni implicite nel modello" (Lindley, *The 1988 Wald Memorial Lectures: The Present Position in Bayesian Statistics*, 1990).

Da tempo la probabilità permea tutti i settori della ricerca chiamati a darsi criteri di induzione o di decisione, ed è entrata a pieno diritto nelle più sofisticate strutture inferenziali dove ad ogni incertezza deve corrispondere una misura. Non si comprende perché sussista ancora tanta reticenza a considerare la probabilità di un evento o di un'ipotesi della stessa natura epistemica della stima di una lunghezza o di una velocità e perché la misura della probabilità deve scatenare più controversie filosofiche o metodologiche della misura di una massa, di un'onda, di un suono, di una qualunque percezione sensoriale.

Se la probabilità è una misura della nostra incertezza, come qualunque altra misura è soggetta ad errore. Poiché tutte le teorie della misura implicano a fronte di un valore misurato un'entità misurabile, non vi è motivo per non estendere queste considerazioni alla probabilità individuale di un asserto H , distinguendo tra valore incognito $\pi(H)$ e sua stima $p(H)$. Non è necessario approfondire che cosa si intende per $\pi(H)$: i problemi epistemologici sono gli stessi della definizione di lunghezza, di forza, di benessere, di velocità (P. Monari, 1992). E non vale obiettare che $\pi(H)$ è personale e soggettiva e non configura in alcun modo una valutazione impersonale attribuibile a tutte le persone razionali: è sufficiente che le probabilità $p(H)$ siano coerenti con la struttura teorica esplicativa del fenomeno osservato, e quindi misurabili con criteri oggettivamente trasferibili (qui la oggettività non attiene alla formazione della probabilità, bensì alla intersoggettività dei criteri di formazione e di misura) come avviene per tante altre determinazioni "soggettive". Si pensi solo alla misura delle percezioni sensoriali, entrata nella scienza ormai da decenni, e oggi costituita in corpo metodologico autonomo con una propria teoria degli errori.

7. CONCLUSIONI

Nel tempo, la discussione sui procedimenti inferenziali ha seguito sentieri diversi che solo raramente hanno saputo trovare momenti di incontro coerenti. Le stesse soluzioni bayesiane, che pure hanno offerto una chiave interpretativa e razionale del ragionamento induttivo, non sempre sono riuscite, nel momento della formalizzazione statistica, a risolversi in criteri pragmaticamente efficaci. Di qui la

storica controversia che ha finito per mettere in discussione il modo stesso di intendere l'induzione, le sue giustificazioni, le sue finalità. Il dibattito sui principi fondamentali dell'inferenza statistica induttiva rimane quindi inconcluso.

Le regole dell'inferenza finora codificate sono regole procedurali sempre relative a un sistema linguistico. Logiche alternative possono quindi coesistere, in quanto riferite a sistemi di regole di inferenza fondati su linguaggi differenti per struttura logica. L'aspirazione ad un modello inferenziale unificante si scontra con la pluralità dei percorsi induttivi praticati dalla ricerca scientifica: ogni procedimento induttivo è materia di fatto – ci ricorda Hume – e pertanto non sussiste altro criterio per scegliere tra diverse soluzioni inferenziali coerentemente fondate, che non sia il “successo” alla prova semantica dei fatti: ancora una volta un criterio empirico del tutto estraneo alla formalizzazione logica o matematica.

Se le controversie continuano ad animare il dibattito è appunto per la difficoltà di confutare qualunque teoria sull'inferenza induttiva sintatticamente coerente. Si potrebbe pensare di invalidare una teoria dell'inferenza sulla base delle contraddizioni intuitive o empiriche costruite su situazioni estreme; ma la storia della scienza ci insegna che teorie consolidate convivono felicemente con i loro paradossi.

Ogni confronto tra le diverse teorie dell'inferenza statistica è soprattutto un confronto tra gli assiomi iniziali (giustificabili solo in termini intuitivi) e l'insieme convenzionale delle relazioni che connettono il mondo dei fatti al mondo delle idee: ossia è un confronto tra modelli.

Ogni teoria induttiva ha quindi un campo di validità strettamente limitato alla struttura convenzionale su cui si fonda: quando viene impropriamente estesa al di fuori di quell'ambito, introduce elementi di incoerenza e genera componenti di ambiguità logicamente non giustificabili. Ad esempio, è un non senso metodologico pretendere di rileggere in chiave bayesiana la teoria della significatività, o pretendere di confrontare, in presenza di alternative coerenti, scelte inferenziali fondate su un test neymaniano oppure su probabilità bayesiane.

Queste semplici considerazioni, dovrebbe dissuadere anche gli statistici dalla tendenza a trasferire in strutture inferenziali, di per sé coerenti, sintassi non appropriate, con la pretesa di ordinare la molteplicità delle soluzioni inferenziali secondo gerarchie di valori (o di criteri) tra loro non commensurabili.

*Dipartimento di Scienze Statistiche
Università di Bologna*

PAOLA MONARI

RIFERIMENTI BIBLIOGRAFICI

- R.B. BRAITHWAITE (1955), *Scientific explanation*, Cambridge University Press.
 B. DE FINETTI (1971), *Probabilità di una teoria e probabilità dei fatti*, “Studi di probabilità, statistica e ricerca operativa in onore di G.Pompij”, Roma.
 R.A. FISHER (1922), *On the Mathematical Foundations of Theoretical statistics*, “Philosophical Transactions of the Royal Statistical Society”, A.
 R.A. FISHER (1930), *Inverse Probability*, “Proc. Cambridge Philosophical Society”, 26.
 R.A. FISHER (1956), *Statistical Methods and Scientific Inference*, Oliver and Boyd, London.

- C. GINI (1943), *I test di significatività*, "Atti della VII Riunione Scientifica della SIS", Roma.
- D. V. LINDLEY (1990), *The 1988 Wald Memorial Lectures: The Present Position in Bayesian Statistics*, "Statistical Sciences", Vol. 5, n. 1.
- I. HACKING (1965), *Logic of statistical inference*, Cambridge University Press.
- A. HERZEL (1991), *Inferenza su popolazioni finite*, "Atti del Convegno della SIS: Sviluppi metodologici nei diversi approcci all'inferenza statistica", Vol. II, Cagliari.
- D.V. LINDLEY (1992), *Is our View of Bayesian Statistics too Narrow?*, "Bayesian Statistics" 4, Proceedings of the Fourth Valencia International Meeting, Clarendon Press, Oxford.
- P. MONARI (1984), *Long run e tests d'ipotesi*, "Statistica", n. 3.
- P. MONARI (1992), *Misurazione e scale. Un riesame critico*, Atti della XXXVI Riunione Scientifica della SIS.
- P. MONARI (1993), *Falsification or choice among alternatives: the unsolved dilemma of hypothesis testing*, "Journal of the Statistical Society", 2, n. 3.
- A. NADDEO (1978), *Impostazione dei problemi di inferenza statistica classica*, "Atti del Convegno: I fondamenti dell'inferenza statistica", Dipartimento Statistico dell'Università di Firenze.
- J. NEYMAN (1955), *The problem of inductive inference*, "Communications on pure and applied mathematics", VIII.
- J. NEYMAN, E.S. PEARSON (1928), *On the use and interpretation of certain test criteria for purposes of statistical inference*, "Biometrika", XX, A.
- J. NEYMAN, E.S. PEARSON (1933a) *On the problem of the most efficient tests of statistical hypotheses*, "Philosophical Transactions of the Royal Statistical Society", A.
- J. NEYMAN, E.S. PEARSON (1933b), *The testing of statistical hypotheses in relation to probabilities a priori*, "Proceedings of Cambridge Philosophical Society".
- C.S. PEIRCE (1923), *Chance, love, and logic* (ed. M.R. Cohen), New York.
- K. POPPER (1934), *Logik der Forschung*, traduzione italiana: *Logica della scoperta scientifica*, Einaudi, Torino, 1970.
- I. SCARDOVI (1977), *L'inferenza induttiva nella ricerca scientifica*, "Atti del Convegno: I fondamenti dell'inferenza statistica", Dipartimento Statistico dell'Università di Firenze.
- I. SCARDOVI (1983), *Induzione statistica e scienza sperimentale*, "Statistica", n. 4.
- I. SCARDOVI (1984), *A proposito di conoscenza e strategia*, "Statistica", n. 1.
- I. SCARDOVI (1987), *La legge statistica*, "Atti del Congresso: Temi e prospettive della logica e della filosofia della scienza contemporanee" Vol. II, CLUEB, Bologna.
- I. SCARDOVI, P. MONARI (1984), *Statistical induction: probable knowledge or optimal strategy*, "Epistemologia", Special Issue.

RIASSUNTO

Il controllo delle ipotesi statistiche: teorie a confronto

Vengono proposte alcune riflessioni critiche sulle teorie dominanti per il controllo delle ipotesi statistiche. La ricerca scientifica ha sempre perseguito due obiettivi fondamentali: la individuazione delle asserzioni false e la scelta, tra le assunzioni verosimili, di quelle più congruenti coi fatti. La demarcazione metodologica tra rifiuto di un assunto statistico perché "falso" o l'esclusione perché "poco probabile" nel confronto con gli altri riguarda le premesse logiche delle procedure inferenziali. Nel primo gruppo si possono collocare i metodi proposti da Fisher e da Neyman e Pearson; nel secondo, le tecniche bayesiane. Per altro, ogni soluzione ha una validità limitata dal sistema di convenzioni preliminari su cui si fonda. In tal senso, teorie diverse possono convivere ed essere utilizzate per risolvere concreti problemi inferenziali.

SUMMARY

The control of statistical hypotheses: a comparison between theories

The paper presents some brief notes regarding the theory of hypothesis testing and the characteristics of inductive procedures of statistics by critically rethinking two basic themes: identifying false assumptions and selecting, amongst the likely assertions, those which are most consistent with a given system. The methodological demarcation between rejection of a statistical statement, because it is “false”, or its exclusion, because it is “less probable”, lies in the fundamental premises of inferential procedures. In the first class we find the methods proposed by Fisher and Neyman and Pearson; in the second one, the Bayesian techniques. Any particular solution has a limit of validity strictly bounded by the conventional procedural rules on which it is based. In this sense, different theories can be used for solving real inferential problems.