STATISTICA, anno LXIV, n. 2, 2004

FISHER, VON MISES, JEFFREYS, NEYMAN, PEARSON E L'INFERENZA STATISTICA

Domenico Costantini

1. PREMESSA

La statistica d'oggidì, segnatamente quella italiana, appare quasi esclusivamente volta alla gestione aziendale – il controllo dei processi produttivi, l'analisi dei mercati, la valutazione della qualità, la soddisfazione dei consumatori e via di questo passo – mentre sembra aver dimenticato le sue origini che furono, di tanto in tanto lo si deve pur ricordare, intimamente legate alla fisica ed alla biologia. Senza alcun dubbio la gestione delle aziende è un importante ambito di applicazione della metodologia statistica; tuttavia concentrarsi troppo su questo tipo di applicazione la riduce ad un ruolo ancillare orbandola delle grandi potenzialità epistemologiche insite sia nelle ricerche degli autori classici, P. S. Laplace, C. F. Gauss, S. D. Poisson, ai quali si deve la nascita della statistica inferenziale come disciplina scientifica, sia in quelle degli scienziati, G. Mendel, L. Boltzmann, W. Lexis, F. Galton, K. Pearson, R. A. Fisher, che le diedero l'impronta moderna. La conseguente diminuzione d'interesse verso la concezione oggettiva della probabilità, ha fatto sì che le ricerche relative alla fondazione frequentista delle discipline statistico-probabilistiche non abbiano molto progredito dal tempo di R. von Mises. Una conseguenza di ciò è la propensione ad identificare, troppo sbrigativamente, probabilità e frequenze relative. A questo proposito vale a pena di rammentare come recentemente un autore, ci riferiamo a J. H. Gillespie non certo l'ultimo fra gli studiosi di genetica di popolazioni, abbia definito, si fa per dire, la nozione di probabilità. Riferendosi ad un locus di cui l'allele A_1 è uno dei possibili attributi, Gillespies (1998, p. 9) si esprime nei riguardi della frequenza allelica, nozione che il contesto chiaramente identifica come la probabilità di un allele, nei termini seguenti:

> Possiamo pensare alla frequenza allelica, p, in due modi diversi. Uno consiste semplicemente nell'intenderla come la frequenza relativa degli alleli A_1 tra tutti gli alleli A della popolazione. L'altro, come la probabilità che un allele scelto a caso dalla popolazione sia un allele A_1.

È difficile immaginare una definizione altrettanto sciatta e fuorviante segnatamente quando si rammenti il ruolo primario che la genetica di popolazione ha nello sviluppo della ricerca scientifica. In quella definizione infatti si sostiene l'equivalenza di due nozioni, per la verità malamente definite, che in realtà sono antitetiche: la prima identifica la probabilità con la frequenza relativa di una popolazione finita, una identificazione insensata dal momento che per questa via, a motivo del fatto che ogni aumento della numerosità della popolazione comporta un cambiamento della probabilità, è impossibile dare un significato preciso a questa nozione; la seconda è sostanzialmente la definizione classica di probabilità che col riferimento all'estrazione casuale cerca di eludere, senza riuscirvi, il ricorso al principio d'indifferenza. Si badi che, qualche riga dopo, Gillespies (1998, p. 10) invita il lettore a meditare con cura sulla definizione probabilistica di *p*.

L'interpretazione frequentista si affloscia come un castello di carte non solo quando è sbrigativamente basata su frequenze relative connesse a popolazioni finite ma, anche, quando la definizione di probabilità, pur riferendosi all'infinito, viene introdotta senza alcuna attenzione al necessario passaggio al limite. Purtroppo questa fu proprio la posizione di Fisher che, avendo identificato la probabilità con la frequenza relativa ed essendo consapevole del "dettaglio" fuggito a Gillespies, volse la sua attenzione verso popolazioni, come lui stesso ebbe a dire, potenzialmente infinite (si veda Fisher, 1922, p. 312). Ma né Fisher né altri sono in grado di precisare cosa sia una popolazione potenzialmente infinita senza fare riferimento, come fece von Mises, al collettivo cioè a una successione infinita la cui casualità – definita per mezzo del *rellosigkeitaxiom* che assicura l'invarianza dei limiti delle frequenze relative delle sue sottosuccessioni – sia introdotta prima della definizione di probabilità.

Questa lunga premessa è necessaria al fine di inquadrare nella giusta prospettiva la posizione di J. Neyman ed E. S. Pearson nei riguardi delle discipline statistico-probabilistiche. Infatti, la teoria dell'inferenza statistica che elaborarono negli anni trenta del secolo scorso può essere intesa come il tentativo di salvaguardare l'oggettività delle inferenze statistiche nel contesto della gestione delle aziende che, proprio in quel torno di tempo, si presentava come un promettente ambito di applicazione delle nozioni probabilistiche. Questo tentativo, altamente meritorio, era reso necessario dall'affacciarsi, al proscenio delle discipline statistico-probabilistiche, di un modo di intendere la probabilità drasticamente contrapposto a quello che fino a quel tempo aveva improntato le inferenze statistiche. Come è ben noto, o almeno dovrebbe esserlo, sto riferendomi ai contributi antifrequentisti di J. M. Keynes, F. P. Ramsey e B. de Finetti, segnatamente alle tesi di quest'ultimo il cui obiettivo dichiarato, la lotta all'oggettivismo condotta in nome di un probabilismo personalistico (de Finetti, 1931), era volta principalmente a combattere l'oggettività delle inferenze statistiche.

Per ragioni che indicherò nel seguito, purtroppo quel tentativo è sostanzialmente fallito; anche se molti statistici seguono formalmente le idee di Neyman e Pearson, nella sostanza hanno abbandonato l'oggettivismo in particolare quello frequentistico per un soggettivismo tanto più pericoloso in quanto non avvertito. La situazione attuale della statistica che ho dianzi deplorato è la miglior conferma di questa affermazione.

Sono fermamente convinto che un'efficace difesa dell'oggettivismo imponga di tornare su una questione di capitale importanza per le discipline statistico-probabilistiche, cioè a dire sul ruolo di queste discipline nelle scienze naturali, questione che necessariamente implica un'attenta analisi del rapporto che intercorre fra probabilità e frequenze relative. Naturalmente quest'analisi non attrae quegli statistici che avendo posto l'azienda al centro del loro interesse, hanno di fatto accolto un'interpretazione personalistica della probabilità. A costoro non ho nulla da dire: si muovono in un contesto troppo lontano da quello cui intendo riferirmi; quel che segue è rivolto a chi sostanzialmente accetta un'interpretazione oggettiva della probabilità e della statistica inferenziale.

Fortunatamente, non tutti gli statistici hanno abbracciato il soggettivismo. Vi sono ancora studiosi che pensano all'inferenza statistica in termini oggettivi e, sia pure senza entrare nei dettagli, condividono le posizioni di Fisher relativamente alla probabilità. Ma è nei dettagli che si nasconde il demonio. Infatti, soffermandosi un poco su di essi, ci si convince di come la sbrigativa identificazione fra probabilità e frequenza relativa sia perniciosa ai fini della difesa di una visione oggettivista della statistica inferenziale. Il cortese invito di Italo Scardovi – proposto nel numero 2, 1999, della rivista *STATISTICA* da lui diretta – non è certamente l'occasione più adatta ad affrontare il cruciale dilemma dell'oggettività o della soggettività delle inferenze statistiche. Tuttavia, e cercherò di mostrarlo in quel che segue, si possono percepire echi di quel dibattito anche operando il "ripensamento critico, al di fuori delle mode correnti, della teoria dell'inferenza statistica secondo Jerzy Neyman ed Egon Pearson" cui ci ha invitato Scardovi. E cosa v'è di più lontano dalle mode correnti di un confronto fra le posizioni di questi due autori, da un lato, e quelle di Fisher, von Mises e H. Jeffreys, dall'altro? Questo è proprio ciò che mi appresto a fare nel presente lavoro che vuole essere un omaggio ad Amato Herzel e Alighiero Naddeo, omaggio che non intende ridursi ad un atto formale di ossequio e rispetto per due colleghi che ci hanno lasciato, bensì sostanziarsi come un "ripensamento" critico delle posizioni cui si ispiravano, operato nella convinzione che, proprio a ragione "della loro intelligenza e della loro cultura" che Scardovi sottolinea, è ciò che da me si sarebbero attesi.

2. SAGGI DI SIGNIFICATIVITÀ E DI IPOTESI

Il ripensamento di cui ho testé detto prende le mosse dalla teoria dei saggi (*tests*) particolarmente da quella che, con l'intento di contrapporla alla bayesiana, viene comunemente qualificata frequentista o oggettivista. Come cercherò di mostrare: il primo aggettivo è fuori luogo, è nulla più di giaculatoria che dovrebbe servire a salvare l'anima a statistici che ormai frequentisti non lo sono più, per lo meno quando questo termine venga preso sul serio; il secondo è del tutto fuorviante dal momento che non si vede la ragione per la quale l'approccio bayesiano alla stima debba necessariamente ispirarsi ad una filosofia soggettivista mentre non possa basarsi su una filosofia oggettivista come è stato dal Settecento, all'Ottocento e in molti casi anche nel secolo scorso, si pensi a C. Gini, a K. Pearson e a Jeffreys. In realtà la maggior parte degli statistici che si qualificano con quei due aggettivi lo fanno in quanto si ritengono seguaci dell'empirismo – correttamente intesa, questa posizione filosofica è la dottrina che non ammette principi o idee innate mentre sostiene che tutta la conoscenza deriva soltanto dal contatto con le cose, vale a dire dallo sperimentare la realtà che sta attorno a noi – perché nutrono la convinzione, ad un tempo ingenua e rozza, per la quale bisogna considerare null'altro che le frequenze relative sbrigativamente identificate coi fatti i soli cui è concesso parlare. Ma – lo ebbe a dire oltre cinquant'anni fa Giuseppe Pompilj (1948) servendosi della battuta del Padre dei *Sei personaggi in cerca d'autore*, "un fatto è come un sacco: vuoto non sta in piedi" – lasciati a sé stessi i fatti tacciono e tutti coloro che se ne dimenticano cadono in un soggettivismo, come ho appena detto, in tanto più pericoloso in quanto meno avvertito. Quando Pompilj formulava l'avvertimento appena ricordato aveva in mente le critiche giniane alla teoria dei saggi e si potrebbe pensare che quella dei fatti come sacchi vuoti fosse una posizione legata alla laplaciana teoria delle probabilità inverse. Ma così non è. Anche autori a noi più vicini, lontani dal soggettivismo ma di matrice bayesiana, hanno ripreso questa critica con grande vigore (Jaynes). Gli è che la pretesa squisitamente filosofica di lasciar parlare i fatti è stata demolita a livello epistemologico col mostrare che, qualora non vengano situati in un preciso contesto teorico che dia loro voce, i fatti sono muti.

Pertanto quegli statistici che argomentando solo sulle frequenze relative sono convinti di non introdurre nelle inferenze elementi estranei e fuorvianti, sono in realtà gli epigoni di una filosofia che ha fatto il suo tempo. Magari inconsapevolmente seguendo Fisher, vale a dire lo statistico più legato a quell'empirismo di cui ho appena detto, quegli statistici pretendono che la verosimiglianza sia lo strumento concettuale atto a trattare con accadimenti incerti perché solo questa nozione sarebbe in grado di misurare il nostro ordine di preferenza fra le possibili popolazioni cui siamo interessati (Fisher, 1958, p. 10).

Senza proseguire in questa direzione, chiedo a questi "frequentisti" o "oggettivisti" che dir si voglia: stanno davvero le cose come voi dite? È proprio vero che un oggettivista deve servirsi solo della verosimiglianza? È proprio vero che un frequentista non può coinvolgere nelle sue inferenze le probabilità iniziali? La risposta a queste domande è nettamente negativa. Tutti coloro che si ritengono frequentisti e oggettivisti dovrebbero conoscere l'opera di von Mises, il più autorevole sostenitore dell'oggettivismo frequentista, l'autore che con maggior coerenza e rigore ha formulato e difeso questo indirizzo e che, alla luce della sua definizione di probabilità, ha analizzato la statistica infererenziale con preclara onestà intellettuale. Il risultato di questa analisi è disponibile in uno splendido ma, purtroppo, poco noto trattato: mi riferisco a *Mathematical Theory of Probability and Statistics*, curato e completato da Hilda Geiringer per i tipi dell'*Academic Press Inc.* di New York, che apparve nel 1964, oltre dieci anni dopo la morte del suo autore. In questo trattato, con grande limpidezza espositiva, von Mises espose i fondamenti e la portata dei saggi di significatività di Lexis, K. Pearson e Fisher contrapponendoli a quelli di ipotesi di Neyman e Pearson. Le conclusioni di von Mises sono disastrose per la teoria dei saggi di ipotesi come è solitamente intesa anche se, come von Mises pose in luce, una corretta interpretazione può recuperare molto della teoria di Neyman e Pearson. Non ho lo spazio per parlare di questo recupero; d'altro canto non ne varrebbe neppure la pena dal momento che ognuno può seguirne il percorso leggendo il suddetto trattato. Mi limiterò ad attirare l'attenzione sui punti essenziali di questa analisi ai quali premetto le conclusioni di von Mises: in un saggio di ipotesi le probabilità iniziali non possono mai essere omesse; qualora non si possa pervenire ad esse, quando cioè non vi sia la possibilità di determinarle frequentisticamente, le conclusioni rifletteranno la mancanza di informazioni circa le suddette probabilità.

Una caratteristica del trattato di von Mises è la grande lucidità con cui, da un punto di vista strettamente frequentista, il suo autore affronta la statistica inferenziale. In considerazione della lucidità di cui ho appena detto, non cercherò di parafrasare i suoi argomenti ma li riporterò testualmente. Dopo aver nettamente distinto fra i saggi di significatività, posti nel capitolo IX dedicato all'analisi dei dati statistici, e i saggi di ipotesi, posti nel successivo capitolo X dedicato al problema dell'inferenza, così von Mises delinea la struttura logica di un saggio di significatività (vedi von Mises, 1964, p. 441):

> Controllare (*checking*) o saggiare (*testing*) ha la forma generale seguente: se il "risultato osservato" ha una "piccola" probabilità subordinatamente all'ipotesi assunta, respingiamo l'ipotesi.

Ovviamente questa, la cui validità è quindi ipotetica, è l'ipotesi nulla. Adeguandomi ad una abitudine, purtroppo universalmente diffusa, userò questa cattiva traduzione per denotare la nozione che Fisher aveva chiamato *null hypothesis*. A mio parere sarebbe stato molto meglio tradurre *null* con "non esistente" o "non attuale", altre due accezioni del termine inglese, mettendo con ciò in risalto il fatto che si sta trattando con un'ipotesi che, come in ogni argomentazione di natura ipotetico-deduttiva, viene avanzata al precipuo scopo di essere confutata.

Prima di affrontare i saggi di ipotesi, von Mises ritorna sulle caratteristiche di un saggio di significatività, completando la descrizione di questi saggi nel modo seguente (vedi von Mises, 1964, p. 494):

> L'analisi dei dati statistici come è stata abbozzata nel precedente capitolo procede nel modo seguente: dalle osservazioni $x_1, x_2, ..., x_n$ si deriva il valore di una funzione $x(x_1, x_2, ..., x_n)$ (ad esempio il X^2 di Pearson o il t di Student o il quoziente di Lexis); poi, con qualche assunzione probabilistica riguardante l'origine delle osservazioni, si determinano il valor medio, la varianza, e se possibile la distribuzione di $x(x_1, x_2, ..., x_n)$; infine si confrontano l'x-valore osservato e il risultato del calcolo.

Dopo di che il nostro autore fornisce un esempio relativo alle prime 2000 cifre dello sviluppo di π con l'intento di controllare, mediante il chi-quadrato, l'ipotesi che ciascuna delle 10 cifre ha la stessa probabilità di presentarsi nel suddetto sviluppo, e conclude (vedi von Mises, 1964, p. 494):

> quindi un valore osservato pari a $X_0^2 = 4,34$ (il valore effettivamente calcolato per $n = 2000$ sembrerà soddisfacente (vale a dire, una conferma dell'assunzione) mentre un valore $X^2 = 1000$ suggerirà di mettere da parte l'ipotesi.

Da quello che abbiamo detto è evidente che per von Mises un saggio di significatività prende in considerazione solamente una ipotesi. Questa constatazione è di capitale importanza: se la si dimentica o, peggio, la si banalizza si perde la possibilità di comprendere la fondamentale differenza che sussiste fra saggi di significatività e saggi di ipotesi e, soprattutto, di afferrare le ragioni che ne fanno due strumenti da impiegarsi in situazioni profondamente diverse. Su questa questione tornerò dopo aver considerato i saggi di ipotesi.

Il capitolo dedicato all'inferenza statistica, che si apre con le considerazioni dianzi viste, è suddiviso in tre parti: nella prima vengono trattati i saggi di ipotesi, nella seconda gli intervalli di fiducia (*confidence intervals*) – vale la pena di sottolineare che il principale significato del termine inglese *confidence* non è confidenza bensì fiducia – e nella terza la stima. Ora volgerò la mia attenzione ai saggi di ipotesi per considerare poi gli intervalli di fiducia. Ma prima di occuparmi di questi saggi, ricordo che von Mises affrontandoli si ricollega esplicitamente alla regola di Bayes-Laplace sostenendo a chiare lettere (von Mises, 1964, p. 495) che

> Il reale problema inferenziale comincia per davvero quando sono date due cose: un valore osservato di x e una funzione di due variabili $p_n(x|\theta)$, cioè a dire la distribuzione di x dipendente dal parametro θ. Problemi di questo tipo sono stati trattati nel capitolo VII [il cui titolo è *Inferenza probabilistica. Il metodo di Bayes*]. Ora riesponiamo il "problema di Bayes" con notazioni leggermente diverse [...] La questione è: *cosa si può dire circa θ se sono dati la formula* [la verosimiglianza calcolata mediante l'indipendenza e la costanza della probabilità] e *un valore osservato di x*? [...] Fin dall'inizio deve essere detto che non ci si può attendere alcuna risposta determinata, nessuna risposta della forma "θ uguale a 0,4" o "θ giace fra 0,4 e 0,5". È nella natura della ricerca statistica che qualunque sia l'affermazione che possiamo porre può solo avere una data probabilità (*chance*) di essere corretta.

Il termine *chance* usato da von Mises denota una nozione di probabilità il cui significato è strettamente connesso al limite della frequenza relativa in un collettivo. Prima di continuare avverto che in questa come nelle seguenti citazioni di von Mises il termine probabilità deve sempre essere inteso col significato testé precisato.

Poco dopo troviamo una importante riflessione su un'affermazione usuale fra gli statistici "frequentisti" che, pur non avendo una relazione immediata con la questione che si sta discutendo, è nondimeno di grande importanza al fine di situarla correttamente entro l'ambito delle inferenze statistiche.

Riferendosi alla distribuzione iniziale del parametro θ di cui ho testé detto, von Mises sostiene (vedi von Mises, 1964, p. 496)

> Prima di, o indipendentemente dagli *n* lanci compiuti con la moneta, qualcosa deve essere conosciuto circa il verificarsi dei vari θ-valori nell'universo delle monete ammesse dell'esperimento statistico. L'affermazione, a volte avanzata come un'obiezione, che θ non è una variabile bensì una "costante sconosciuta" avente un unico valore per la moneta che stiamo considerando, esula dalla questione. Ogni affermazione relativa alla probabilità che θ cada in un dato intervallo [...] è necessariamente un'affermazione relativa ad un universo di monete [...] con differenti θ-valori.

Fatte queste precisazioni, von Mises affronta i saggi di ipotesi (vedi von Mises, 1964, p. 505):

> Una variante del problema dell'inferenza [che come abbiamo visto von Mises considera risolvibile solo mediante la regola di Bayes-Laplace] è noto come *il problema dei saggi di ipotesi*. Di nuovo sono dati una funzione $p(x|\theta)$ [...] e una regione H di θ-valori e, di nuovo, si considera un'infinita successione di prove in cui si è osservato x. Dopo ciascuna prova [...] si avanza un'affermazione, vale a dire: se il valore osservato x cade in una specificata regione di θ-valori A, affermiamo che θ è in H, e se x cade nel complemento $\overline{A}$ (= non-A), affermiamo che θ è esterno ad H (cioè in $\overline{H}$ o non-H). Il problema è nuovo di trovare la probabilità di aver fatto un'asserzione corretta.

Ovviamente, A è la regione di accettazione, $\overline{A}$ quella di rifiuto o critica, H è l'ipotesi nulla e $\overline{H}$ è l'ipotesi alternativa. Dopo aver posto il problema in questi termini, von Mises ne mette in risalto l'aspetto, davvero cruciale, passato sotto silenzio da coloro che si occupano dei saggi di ipotesi, vale a dire (von Mises, 1964, p. 505):

> la probabilità iniziale di θ sarà descritta dalla sua funzione di distribuzione $P(\theta)$. Il fatto che in generale $P(\theta)$ non è conosciuta origina le principali difficoltà del problema.

Dopo aver impostato correttamente, bayesianamente quindi, la questione, von Mises introduce le probabilità di successo e di errore

$$\int_H P(A|\theta)dP(\theta), \quad x \in A, \ \theta \in H \tag{1}$$

$$\int_{\overline{H}} P(\overline{A}|\theta)dP(\theta), \quad x \in \overline{A}, \ \theta \in \overline{H} \tag{2}$$

$$\int_H P(\overline{A}|\theta)dP(\theta), \quad x \in \overline{A}, \ \theta \in H \tag{3}$$

$$\int_{\overline{H}} P(A|\theta)dP(\theta), \quad x \in A, \ \theta \in \overline{H} \tag{4}$$

vale a dire la somma di (1) e (2) rispettivamente quella di (3) e (4) – ovviamente, la (3) e la (4) sono le probabilità di commettere un errore di primo e secondo tipo – e afferma (von Mises, 1964, p. 509):

> In tutti i lavori che seguono la linea di pensiero di Neyman e Pearson, vengono discusse le probabilità degli errori di primo e secondo tipo, ma raramente viene presa in considerazione la probabilità globale [iniziale] $P(\theta)$. Non è corretto dire che α è la probabilità dell'errore di primo tipo: questa probabilità è $\alpha\pi_0$ [π_0 è la probabilità iniziale di $\theta = \theta_0$]; ed è sbagliato dire che la probabilità di un errore di secondo tipo dipende da θ: dipende dalla distribuzione di θ. Anche la quantità β che introdurremo fra breve non è la probabilità di un errore di secondo tipo ma il minimo limite

superiore di $P(A\mid\theta)$ per tutti i θ in $\overline{H}$. Molti autori (seguendo R. A. Fisher) desiderano evitare l'uso della distribuzione iniziale. Il fatto che "essa non sia nota" non cambia la situazione ed ometterla dalle formule non la elimina. Si può cercare di pervenire a conclusioni che non dipendono o non dipendono in modo essenziale da $P(\theta)$; altrimenti le nostre affermazioni riflettono la nostra (possibilmente incompleta) informazione relativamente a $P(\theta)$.

Pertanto un saggio di ipotesi considera sempre almeno due ipotesi che sono paragonabili unicamente sulla scorta delle loro probabilità iniziali.

Considerando per semplicità il caso in cui H sia $\theta=\theta_0$ e $\overline{H}$ sia $\theta=\theta_1$, ne consegue che qualora, per ignoranza o pregiudizio filosofico, si omettano le probabilità iniziali $P(\theta_0)$ e $P(\theta_1)$ limitando l'attenzione alle inconfrontabili verosimiglianze $P(\overline{A}\mid\theta_0)$ e $P(A\mid\theta_1)$, si finisce per chiedersi, come per celia fanno i fanciulli, se corre più veloce il treno o è più dolce lo zucchero. Sicuramente questa non fu la posizione di von Mises che si guardò bene dal seguire la politica dello struzzo essendo ben consapevole, come abbiamo appena visto, che non conoscere le probabilità iniziali non cambia la situazione in cui ci si trova mentre ometterle non le elimina. L'ammonizione che ho testé parafrasato viene dal più autorevole dei frequentisti; ne consegue che chi la ignora si pone al di fuori di un coerente e rigoroso frequentismo. Frequentismo non significa ignorare i principi della probabilità ma dare ad essi un'interpretazione; per questa ragione ho detto che è fuori luogo usare il termine frequentista al fine di qualificare gli statistici che ignorano le regole fondamentali della probabilità.

Come sottolinea lo stesso von Mises, si può giungere a conclusioni che non dipendono o dipendono in misura ridotta dalle probabilità iniziali, questo vale in certi casi per la nozione di potenza di un saggio, ma l'impostazione e la soluzione del problema non può prescindere da queste probabilità. In verità, per lo meno sulla base di una dichiarazione fatta da E. Pearson negli anni sessanta del secolo scorso, la mancanza di rigore del confronto fra le verosimiglianze operato senza tener conto delle probabilità iniziali, era percepito anche dai proponenti i saggi di ipotesi dal momento che (Pearson, 1962, p. 55)

> Certamente noi [Neyman e Pearson] eravamo consapevoli che le inferenze debbono usare le informazioni iniziali [...] ma dopo aver riflettuto e discusso a lungo su questo punto, siamo giunti alla conclusione, giusta o sbagliata, che la possibilità di assegnare valori numerici sicuri a queste entità era così infrequente che il nostro modo di procedere avrebbe dovuto seguire altre vie. [...] Proprio perché il problema di specificare i valori numerici di queste probabilità ci era sembrato così frequentemente insolubile, ci proponemmo di analizzare in quale senso le conclusioni tratte da un saggio potessero essere descritte come indipendenti da queste probabilità.

Questa consapevolezza non era solo di Pearson; per convincersene basta rileggere le considerazioni che Neyman fece in un lavoro dedicato ai fondamenti della statistica inferenziale (Neyman, 1952). Il problema posto dal ruolo delle probabilità iniziali nei saggi di ipotesi non fu mai banale; per primi Neyman e Pearson lo ponderarono a lungo a differenza di quel che troppo frequentemente fanno i loro seguaci: a imporre la rimozione di queste probabilità era stato un giudizio di natura pratica assolutamente privo di qualsivoglia giustificazione teorica; squisitamente pratica è infatti la decisione di eliminare le probabilità iniziali a ragione della frequente impossibilità di assegnar loro valori numerici sicuri.

Ben diversa era stata la posizione di Fisher. A suo parere (Fisher, 1956, cap. I par. 2), al fine di non negare la validità della scienza sperimentale, è necessario che dalla conoscenza di un campione si possa inferire alcunché sulla corrispondente popolazione; tuttavia – ad eccezione dei casi in cui siano disponibili probabilità fiduciali di cui parleremo nel seguito – la probabilità non è adeguata allo scopo suddetto. La nozione che assolve questo compito è la verosimiglianza che non soddisfa i principi della probabilità e, come ho già ricordato, fu introdotta da Fisher come una nozione alternativa a quella di probabilità non essendo stata ideata per valutare le probabilità delle differenti possibili popolazione da cui il campione può essere stato tratto bensì per misurare l'ordine di preferenza fra queste popolazioni. Senza alcun dubbio la posizione di Fisher è metodologicamente corretta basata com'è su un preciso retroterra filosofico: è pertanto epistemologicamente ineccepibile. Può essere condivisa o meno – non la condivido poiché la ritengo basata sull'empirismo di cui ho detto – ma le si deve riconoscere una chiarezza esemplare. Al contrario, nulla della chiarezza di Fisher si trova alla base delle scelte di Neyman e Pearson. La posizione di questi autori è confusa dal momento che mescola senza distinzioni di sorta considerazioni teoriche e pratiche. Con tutta la buona volontà di questo mondo, non sono in grado di comprendere come una scelta dovuta a ragioni meramente pratiche possa influenzare lo sviluppo teoretico al punto da giustificare, si fa per dire, un'evidente fallacia logica. Ma quel che più stupisce è che la maggior parte di coloro che seguono impostazioni alla Neyman-Pearson, oltre a non conoscere la riformulazione di von Mises, non sembrano conoscere né la tardiva confessione di Pearson e neppure i rovelli metodologici di Neyman.

Riservandomi di tornare sulla profonda differenza concettuale originata dal considerare una sola o più ipotesi, segnatamente sui differenti contesti in cui ciò può avvenire, volgo ora l'attenzione all'autore che per primo, e con grande chiarezza, vide come una trattazione logicamente corretta tassativamente escluda la possibilità di confrontare due ipotesi basandosi unicamente sulle loro verosimiglianze.

3. IL QUOZIENTE *K* DI JEFFREYS

Un altro splendido contributo all'inferenza – meglio, alla teoria della probabilità come lo stesso autore evidenzia col titolo – che pochi statistici conoscono ivi compresi i bayesiani, è il trattato *Theory of Probability* di Jeffreys; pubblicato nel 1939, ebbe tre edizioni e alcune revisioni essendo l'ultima del 1966. Per inquadrare il pensiero dell'autore, credo sia opportuno cominciare con un brano tratto dalla prefazione alla prima edizione (Jeffreys, 1961, p. ix)

> Gli statistici moderni hanno sviluppato notevoli tecniche matematiche ma per la maggior parte hanno rifiutato la nozione di probabilità di un'ipotesi e a causa di ciò si sono privati di ogni possibilità di dire con precisione cosa intendano quando decidono tra ipotesi.

Quindi per Jeffreys i seguaci dell'approccio ortodosso alla statistica – usiamo questo aggettivo che E. T. Jaynes (Jaynes, 1984, p. 44), un ideale allievo di Jeffreys, molto ragionevolmente introdusse in sostituzione dell'usuale ma scorretto classico; questo aggettivo è infatti appropriato per la statistica elaborata nel periodo classico cioè quello a cavallo fra il Settecento e l'Ottocento – rifiutando le probabilità delle ipotesi, specificamente le probabilità iniziali, si sono privati della possibilità di capire fino in fondo cosa stiano facendo. Jeffreys sostiene che ciò avviene anche quando essi operano un saggio di significatività: non condivido questa opinione e ne dirò le ragioni. Ora intendo mostrare, in un caso particolare ma significativo (Jeffreys, 1961, cap. V, par. 5.1), come le tesi di von Mises fossero già state sostenute da Jeffreys. Ricordando le idee di questo autore, avrò l'opportunità di porre in risalto come un saggio di ipotesi possa essere formulato in termini probabilisticamente ineccepibili e nel contempo che il cosiddetto fattore di Bayes non sia una scoperta recente ma fosse noto da almeno mezzo secolo (Jeffreys, 1961, p. 248). Come molto spesso accade, l'ignoranza della storia costringe a ripercorrerla.

Consideriamo dapprima un evento che possa o meno accadere; supponiamo poi che non si possa escludere alcuno dei valori della sua sconosciuta probabilità, ciò significa che la probabilità potrà essere uno qualunque dei numeri di (0,1); supponiamo, infine, che si abbiano ragioni per ritenere che questa sconosciuta probabilità sia pari a p.

Sulla scorta di un certo numero di osservazioni, diciamo $n = n_1 + n_2$, nelle quali n_1 volte l'evento si è verificato mentre n_2 volte non si è verificato, si intenda controllare se l'assunzione è corretta, in altri termini, se sia corretta l'ipotesi nulla secondo cui la sconosciuta probabilità è p. Oltre all'ipotesi nulla consideriamo anche l'ipotesi alternativa che nella fattispecie è

$$Q = q \in (0,\ 1) \wedge q \neq p$$

La distribuzione iniziale che Jeffreys propose per un controllo di questo tipo è

$$P(p) = 1/2 \text{ e } P(Q) = 1/2 \tag{5}$$

Questa scelta introduce una marcatissima asimmetria probabilistica sull'insieme di tutte le possibilità; infatti, mentre ad una sola possibilità assegna una probabilità finita pari ad 1/2, all'unione di tutte le infinite possibili alternative assegna lo stesso valore. Dopo questa scelta, Jeffyes calcola le verosimiglianze facendo riferimento allo schema bernoulliano e quindi le probabilità finali. Il rapporto fra queste è

$$K = \frac{(n+1)!}{n_1!n_2!} p^{n_1}(1-p)^{n_2}$$

Il simbolo *K* è di Jeffreys. *K* è quindi un quoziente fra probabilità finali che, a ragione dell'uguaglianza delle probabilità iniziali – ovviamente l'uguaglianza è relativa all'ipotesi nulla e all'ipotesi alternativa non certo alle le probabilità di tutte le ipotesi che concorrono alla individuazione di quest'ultima – si riduce ad un rapporto fra verosimiglianze. Il suo significato è palese: essendo state inizialmente poste sullo stesso piano l'ipotesi nulla e l'ipotesi alternativa, il quoziente *K* si limita a confrontare le verosimiglianze.

Per terminare sottolineo, nuovamente perché di estrema importanza, che il rapporto fra le verosimiglianze fondato sulla validità di un'ipotesi analoga alla (5), è l'unico modo probabilisticamente corretto atto a fornire un'interpretazione ineccepibile ai saggi di ipotesi. Alla luce di quanto abbiamo visto in questo e nel precedente paragrafo, i saggi di ipotesi si configurano dunque come un tipo particolare di stima statistica.

4. INFERENZA SCIENTIFICA E PROCEDURE D'ACCETTAZIONE

Nel corso degli anni cinquanta del secolo scorso, era ben noto alla comunità statistica l'aperto contrasto esistente fra le posizioni alla Fisher e alla Neyman-Pearson riguardo al controllo delle ipotesi. Oggi le cose sono cambiate tanto che sembra lecito dubitare se questa fondamentale differenza sia ancora avvertita. Per questa ragione mi sembra importante tornare sulle tesi di Fisher e per contrapposizione su quelle di Neyman e Pearson. Senza aggiungere nulla – questo mi costringerà a lunghe citazioni delle quali fin d'ora mi scuso; il mio sostegno alle tesi di Fisher lo esprimerò nel prossimo paragrafo servendomi di un celebre esempio storico – riporto testualmente alcuni brani di Fisher che mostrano come il contrasto fosse già chiaramente delineato fin dalla prima metà del secolo scorso.

Dapprima mi occuperò de *The Design of Experiment*, il contributo che per tanti versi rappresenta l'apologia dei saggi di significatività e che si apre col famoso esempio, oggi un po' dimenticato, della signora che afferma di saper riconoscere le tazze in cui il tè è stato versato sul latte da quelle in cui il latte è stato aggiunto al tè. Parlando dei possibili fraintendimenti dei risultati di un esperimento scientifico, Fisher afferma (Fisher, 1960, p. 16)

> si deve notare che l'ipotesi nulla non è mai verificata (*proved*) o confermata (*established*), ma può essere confutata (*disproved*), nel corso della sperimentazione.

E poco dopo (Fisher, 1960, p. 17)

> Alla nozione di un errore cosiddetto di "secondo tipo," dovuto all'aver accolto l'ipotesi nulla "quando è falsa", può essere dato un significato con riferimento ad una grandezza che deve essere stimata [si noti: stimata]. Non ha significato con riferimento ad un semplice saggio di significatività, nel quale le sole attese in grado di

essere usate vantaggiosamente (*available*) sono quelle che seguono dall'essere vera (*being true*) l'ipotesi nulla.

Da cui si evince che, fatto salvo il differente modo di intendere la stima del valore sconosciuto di una grandezza, la posizione di Fisher riguardo ai saggi di ipotesi non è tanto lontana da quelle di von Mises e Jeffreys. Nonostante la grande chiarezza di queste affermazioni, segnatamente di quella di pagina 16, prima di proseguire sembra opportuno notare che in esse potrebbe forse celarsi una contraddizione o, più verosimilmente, una svista. Infatti, mentre nel passo di pagina 16, che è certamente il più convincente, Fisher afferma senza mezzi termini che un'ipotesi nulla non potrà mai essere verificata ma solo confutata, in quello della seguente pagina 17 parla di verità dell'ipotesi nulla senza mettere in risalto che questa verità è sempre e solo ipotetica, mai effettiva. A farmi propendere per la svista sta la constatazione che in altre occasioni la natura ipotetica dell'ipotesi nulla è stata chiarissimamente messa in risalto da Fisher, ad esempio in questo passo (Fisher, 1956, p. 44)

> In generale i saggi di significatività sono basati su probabilità ipotetiche dedotte dalle loro ipotesi nulle

le cui verità, pertanto, non possono che essere ipotetiche.

Mi sono soffermato su questa svista, vera o presunta che sia, perché mi ha fornito l'opportunità di sottolineare la natura ipotetico-deduttiva dei saggi di significatività – sfuggita, come fra breve vedremo, anche a statistici di grande fama – e di contrapporla a quella dei saggi di ipotesi. Ancora una volta, con riferimento esplicito al comportamento induttivo teorizzato da Neyman, Fisher pone in evidenza il ruolo dei saggi di significatività stigmatizzando nel contempo la profonda differenza fra questi e quelli di ipotesi (Fisher, 1960, pp. 25-26)

> Nell'apprendere dall'esperienza [...] le conclusioni sono sempre provvisorie e hanno la natura di protocolli in divenire (*progress*), nell'interpretare e compendiare l'insieme dei dati di fatto (*evidence*) fin qui acquisiti. Pur essendo conveniente registrare un'ipotesi come contraddetta a qualche famigliare livello di significatività, del tipo 5% o 2% o 1%, nel corso di un'inferenza induttiva è necessario non perdere mai di vista l'esatta pregnanza (*strength*) che l'insieme dei dati di fatto (*evidence*) ha effettivamente raggiunto, oppure ignorare che ulteriori esperimenti potrebbero rafforzare o indebolire la suddetta pregnanza. La situazione è completamente differente nell'ambito delle Procedure di Accettazione in cui può accadere di dover eseguire azioni irreversibili, nelle quali, qualunque sia la decisione cui si è pervenuti, non è di alcuna importanza il fatto che essa sia stata presa sulla scorta di un insieme dei dati di fatto (*evidence*) di molta o poca pregnanza (*strength*). Tutto ciò che serve è una Regola d'Azione che deve essere automatica, svincolata da preoccupazioni connesse alla decisione individuale. Nel suo insieme alla procedura si perviene minimizzando le perdite dovute ad una decisione sbagliata [...] e per costruire con successo una siffatta procedura si richiede la precedente conoscenza della prevedibile distribuzione del materiale da collocare. Nell'ambito della ricerca pura non può che essere simulazione qualsivoglia valutazione del costo di un'errata conclusione, o del ritardo nel pervenire a conclusioni più corrette, e comunque una siffatta valutazione sarebbe inammissibile e irrilevante nel giudicare lo stato delle conoscenze scientifiche; [...]

> Simili differenze fra i contesti logici devono essere tenute presenti tutte le volte che si parli dei saggi di significatività come di "Regole d'Azione". Certamente buona parte della confusione è stata causata dal tentativo di formalizzare l'esposizione dei saggi di significatività in un quadro logico diverso da quello in cui erano stati effettivamente elaborati.

La citazione molto lunga è giustificata dalla necessità di tornare a riflettere su questo giudizio durissimo e, soprattutto, sul raffronto fra ricerca pura e procedure d'accettazione in uso nella gestione delle aziende.

Contrariamente a quello che usualmente accade, il tempo non mitigò il giudizio di Fisher; al contrario la contrapposizione divenne se possibile più acuta quando, tornando sull'argomento una ventina d'anni più tardi nel capitolo IV (Alcuni equivoci relativi ai saggi di significatività) della sua principale opera fondazionale *Statistical Methods and Scientific Inference*, Fisher torna a discutere le differenze esistenti fra i saggi di significatività e quelli di ipotesi (Fisher, 1956, pp. 75-76)

> I comuni saggi di significatività [...] sono apparsi nei primi due quarti del ventesimo secolo per giocare un ruolo centrale nell'analisi statistica. Nel quotidiano lavoro di ricerca sperimentale delle scienze naturali, sono costantemente usati per distinguere gli effetti di reale importanza in un programma di ricerca da quelli apparenti che potrebbero essere dovuti a errori compiuti nel corso del campionamento casuale [o d'altro tipo] [...]. Tra gli innumerevoli esempi che si potrebbero portare, essi sono usati per riconoscere [...] l'autenticità di un'associazione genetica (*genetic linkage*), il reale effetto della concimazione sul raccolto [...]. Le conclusioni tratte da siffatti saggi costituiscono i passi mediante i quali il ricercatore acquisisce una migliore conoscenza del suo materiale sperimentale e dei problemi che esso pone.
>
> È pure degno di nota che gli uomini i quali avvertirono la necessità di questi saggi, che per primi li hanno pensati o che poi li hanno resi matematicamente precisi, erano tutti attivamente impegnati in ricerche connesse alle scienze naturali. Più recentemente, invero, una considerevole parte degli sforzi intellettuali si sono volti verso il tentativo di spiegare, o piuttosto di reinterpretare, questi saggi sulla base di modelli molto diversi, vale a dire di immaginarli dotati di un significato connesso alla presa di decisioni nel corso di una procedura di accettazione. All'autore le differenze fra queste due situazioni sembrano molte e ampie, e non ritengo si sarebbero potute trascurare qualora gli autori di questa reiterpretazione avessero avuto una reale familiarità col lavoro delle scienze naturali o fossero stati consapevoli delle caratteristiche di un protocollo osservazionale che consente l'incremento della comprensione scientifica, [...] Per la verità, l'errata interpretazione sembra andar ben più in là di quanto ci si aspetterebbe se fosse dovuta al mero trasferimento di tecniche da un ambito di studi all'altro.

Come si vede, lo sprezzante giudizio di Fisher, accentuato dall'ultima frase, giunge sino a negare a Neyman e Pearson, e con loro a tutti gli statistici che ne seguirono le idee, qualunque dimestichezza tanto con la scienza quanto con la ricerca scientifica. Un tagliente giudizio che stigmatizza la pretesa di tanti statistici di discettare intorno ai saggi di significatività senza avere alcuna conoscenza né della ricerca scientifica né del modo con cui questa procede.

Ho ampiamente citato Fisher per mostrare quanto fosse radicale il suo dissenso da Neyman e Pearson relativamente al significato e ai compiti della metodologia inferenziale. Tuttavia, malgrado quello che abbiamo letto, vi sono ancora molti statistici che non colgono la fondamentale differenza fra un saggio di significatività ed uno di ipotesi. Un esempio recente di questo tetragono rifiuto di vedere le cose come stanno, è offerto da E. L. Lehmann (1993) secondo cui

> nonostante le fondamentali differenze filosofiche, nei loro principali aspetti pratici le due teorie [di Fisher e Neyman e Pearson] sono complementari piuttosto che contraddittorie ed è possibile un approccio unificato che combini le loro migliori caratteristiche

che nel contempo mostra tanto la scarsa importanza che molti statistici attribuiscono alla riflessione sui fondamenti – con l'insensibilità alle fallacie logiche che questa trascuratezza porta con sé – quanto un dato storico che non verrà mai abbastanza evidenziato, vale a dire che anche famosi statistici, forse proprio per le ragioni segnalate da Fisher, non sono stati in grado di cogliere il significato delle sue tesi. Al contrario, sono pienamente convinto delle profonde, inconciliabili differenze che sussistono fra i saggi di significatività e quelli di ipotesi e condivido con Fisher la convinzione secondo cui i saggi di significatività hanno un grande ruolo nelle scienze della natura. Sono inoltre convinto che, una volta situati nel contesto della stima, come lo stesso Fisher sembra suggerire, e resi ineccepibili, secondo le idee di Jeffreys e von Mises, i saggi di ipotesi abbiano un ruolo non trascurabile, di cui mi guardo bene dal negare l'utilità pratica, in ogni corretta gestione aziendale.

5. LO SPETTRO DELLA RADIAZIONE DI CORPO NERO

Secondo Fisher quindi nella stragrande maggioranza delle situazioni nelle quali uno scienziato viene a trovarsi nel corso del suo operare, egli dispone di una sola ipotesi atta a spiegare e a prevedere ciò che accade e accadrà nel mondo che lo circonda, appunto, quella che Fisher chiamò ipotesi nulla. Qualora lo scienziato dovesse riscontrare discordanze fra questa ipotesi e i dati d'osservazione, gli si porrà il problema di individuare un metodo atto confutare l'unica ipotesi di cui dispone al fine di avventurarsi, dopo aver bruciato i ponti alle sue spalle, nella ricerca di una nuova ipotesi che, meglio di quanto non facesse quella confutata, spieghi i fatti; una volta individuata, anch'essa potrà divenire un'ipotesi nulla e, di nuovo, diverrà attuale la sua confutazione. Questo avvicendarsi di ipotesi nulle e confutazioni è all'origine di molte rivoluzioni scientifiche: è il galileano metodo ipotetico-deduttivo modificato al fine di render conto di leggi non più deterministiche bensì statistiche. Fisher si è attenuto a questa prassi scientifica e in questa prospettiva si è mosso nella giusta direzione. In questo paragrafo, al fine di portare argomenti a favore della tesi di Fisher, intendo occuparmi della confutazione che si trova alla base della svolta intellettuale che condusse la fisica dalla meccanica newtoniana a quella quantistica.

Negli ultimi anni dell'Ottocento, alcuni fisici sperimentali che lavoravano nel *Physicalish-Technische Reichsanstalt* di Berlino, uno dei migliori se non il migliore fra i laboratori fisici di quel tempo, scoprirono che la forma dello spettro della radiazione di corpo nero – cioè la radiazione emessa da una fornace ma anche dal sole – era diversa da quella teoricamente dedotta supponendo che la legge che regola il comportamento di quella radiazione fosse la stessa che regola il comportamento delle molecole. Le caratteristiche di questa legge conseguivano, in primo luogo, dall'ipotesi che l'energia fosse una quantità continua e, in secondo luogo, dall'ipotesi che gli eventi coinvolgenti molecole fossero caratterizzati dall'indipendenza stocastica e dalla costanza della probabilità (di accomodamento). M. Planck, che lavorava a Berlino ed era in contatto con i fisici sperimentali di cui ho appena detto, rovesciò per così dire la questione nel senso che introdusse due fondamentali principi: l'energia è costituita da quantità discrete – i quanti appunto che diedero il nome alla nuova meccanica – il cui comportamento non è stocasticamente indipendente. Il significato di queste due assunzioni fu colto in tempi diversi: l'ipotesi quantistica, di cui non mi occuperò, suscitò immediatamente grande scalpore, mentre l'abbandono dell'indipendenza stocastica allora non fu colto; anzi, lo stesso Planck sembra non essersi reso conto del profondo significato di questo abbandono per lo meno altrettanto ma, forse, più importante dell'ipotesi quantistica. Per ragioni di spazio, non mi è possibile giustificare queste affermazioni, mi limiterò quindi a dire che ciò è dovuto al fatto che l'attenzione di Planck era volta verso i risuonatori (oscillatori, in termini moderni), cioè verso una sorta di molecole, bene inteso concettuali non reali, del campo elettromagnetico la cui energia aumenta col crescere del numero dei quanti da essi assorbito. Ed effettivamente Planck assegnò a risuonatori un comportamento simile a quello delle molecole di Boltzmann, vale a dire li immaginò regolati da distribuzioni statistiche fondate sull'indipendenza.

L'episodio che ho ricordato rappresenta il momento finale, la confutazione dell'ipotesi nulla, di un saggio di significatività, molto sofisticato per la verità, proprio del tipo immaginato da Fisher, cioè a dire caratterizzato da un'unica ipotesi che può solo essere confutata e dalla mancanza di qualsivoglia ipotesi alternativa. Ma per cogliere la struttura di questo saggio, che si concluse con la confutazione della meccanica classica, bisogna prendere le mosse un poco da lontano.

Negli anni sessanta del diciannovesimo secolo J. C. Maxwell aveva mostrato che in un gas all'equilibrio, vale a dire dopo un periodo sufficientemente lungo nel quale non si è intervenuti su di esso, le sue caratteristiche macroscopiche, ad esempio la sua pressione o la sua temperatura, sono il risultato (medio) di avvenimenti microscopici, le velocità o gli urti contro le pareti del recipiente o altri ancora, relativi alle molecole di cui il gas è costituito. Tuttavia Maxwell non aveva indicato il meccanismo mediante il quale gli accadimenti microscopici conducevano alla distribuzione d'equilibrio – la celeberrima distribuzione di Maxwell, null'altro che una chi-quadrato con tre gradi di libertà – senza però dare alcuna indicazione circa il meccanismo microscopico che la genera. Se, togliendo i rispettivi tappi, mettiamo in comunicazione due bottiglie, in una delle quali è stato fatto il vuoto mentre l'altra contiene un gas, constatiamo che il gas della bottiglia piena occupa lo spazio di quella vuota a guisa che le due bottiglie conterranno gas con le medesime caratteristiche. È allora naturale chiedersi: quale fenomeno microscopico determina ciò che constatiamo a livello macroscopico?

Una decina d'anni più tardi la risposta fu data da Boltzmann che individuò nelle collisioni fra molecole la ragione che porta il gas all'equilibrio. Seguendo Boltzmann, quindi, immaginiamo la collisione fra due molecole come un avvenimento che muta l'energia cinetica delle molecole che si urtano; non si considerano urti fra un numero maggiore di molecole, tre, quattro o ancor di più, poiché si suppone il gas così rarefatto da rendere simili collisioni praticamente impossibili. Ogni molecola è caratterizzata da un data energia e quindi le molecole di un gas racchiuso in un recipiente possono essere distribuite (ripartite) in classi di molecole con la stessa energia o, meglio, la cui energia è compresa in un intervallo sufficientemente piccolo. Per semplificare il racconto considererò solo molecole monoatomiche e trascurerò l'energia potenziale; inoltre, colgo l'occasione per avvertire che nel seguito fisserò la mia attenzione sugli aspetti importanti ai nostri fini mentre ne trascurerò altri pure essenziali come ad esempio i livelli energetici in cui una particella può trovarsi. La collisione fra due molecole provoca quindi un mutamento delle frequenze delle molecole contenute per così dire nelle varie classi di energia. Più precisamente, considerando due molecole, X e Y, supponiamo che prima della collisione la molecola X abbia energia A e dopo la collisione abbia energia A', mentre Y passa dall'energia B a B'. Di nuovo per semplicità ho indicato la classe con un ben preciso valore d'energia trascurando il fatto che essa è costituita dalle molecole la cui energia è compresa nell'intervallo $A+dA$, d'altro canto, però, l'intervallo può essere scelto tanto piccola da rendere trascurabile le differenze che sussistono fra le energie delle molecole che appartengono alla medesima classe. Quindi, mentre prima delle collisione X è nella classe A dopo la collisione X sarà nella classe A'; similmente Y passerà dalla classe B alla classe B'. L'evento che ci interessa quindi è rappresentato dalla diminuzione di una unità (della numerosità) delle classi A e B e dall'aumento di una unità delle classi A' e B'.

Il periodo in cui Boltzmann opera la sua fondamentale ricerca, gli anni settanta del XIX secolo, è caratterizzato dalla supremazia intellettuale del determinismo. Si è cioè convinti che tutti i fenomeni naturali siano governati da rigide leggi che determinano tutto ciò che accade proprio nel modo in cui accade e non in un altro. Questa atmosfera originò aspre critiche alle ricerche di Bolztmann incentrate sull'uso di nozioni imprecise quali allora erano considerate quelle statistico-probabilistiche. Si deve anche tener conto del fatto che in quel periodo la definizione di probabilità universalmente accolta era la classica di Laplace; lo stesso Boltzmann, in uno dei suoi primi lavori dedicati alla teoria cinetica dei gas, aveva fatto riferimento a questa definizione. Era quindi largamente estranea alla cultura del tempo l'identificazione fra probabilità e frequenza relativa tipica delle ricerche di Boltzmann che, d'altronde, furono fra quelle che maggiormente spinsero ad abbandonare la definizione classica. Infatti le argomentazioni di Boltzmann erano sviluppate in termini di numeri di molecole, cioè di frazioni di popolazioni di molecole, con determinate caratteristiche di energia, ed era pertanto naturale pensare che le probabilità concernessero non già singole molecole bensì popolazioni di molecole: la probabilità del cambiamento di energia da A ad A' è la frazione delle molecole del sistema che passano dall'energia A all'energia A'.

Tornando alle collisioni, Boltzmann, sulla base di un'analisi meccanico-geometrica dell'urto, dedusse il numero di urti che conducono dall'energia *A* ad *A'* e da *B* a *B'* e dimostrò che i loro effetti sulla distribuzione delle (classi di) energie è tale che l'entropia aumenta fin quando la distribuzione delle molecole raggiunge quella di Maxwell. Questo è in sostanza il contenuto del celeberrimo teorema *H* di Boltzmann secondo cui l'entropia è una funzione non decrescente del tempo o, meglio, questo è ciò che Boltzmann credeva di aver dimostrato. Le discussioni cui diede origine il risultato di Boltzmann chiarirono in seguito che l'aumento dell'entropia non è certo ma solo estremamente probabile.

Un'ipotesi essenziale al fine della determinazione dei mutamenti che avvengono nella numerosità delle classi a seguito delle collisioni, e quindi ai fini della dimostrazione del teorema *H*, è l'ipotesi del caos molecolare, nota anche come *Stosszahlansatz*, secondo cui il comportamento delle molecole è stocasticamente indipendente. L'ipotesi del caos molecolare fu uno dei principali argomenti di discussione della termodinamica statistica. Il dibattito verteva sulla natura di questa ipotesi, se cioè fosse o meno compatibile col determinismo che si supponeva regolasse le collisioni molecolari. L'opinione che alla fine ebbe il sopravvento riconosce che con questa ipotesi si abbandona il determinismo, vale a dire con essa si introduce nella teoria cinetica dei gas una componente squisitamente probabilistica. Le argomentazioni di Boltzmann, fondate su un'analisi in parte deterministica e in parte statistica, suscitarono forti perplessità senza riuscire a convincere i suoi contemporanei circa la possibilità di fondare su di esse la termodinamica.

Boltzmann ragionava sull'evoluzione temporale delle frequenze di un'unica popolazione, quella delle molecole di un gas contenuto in un recipiente, ma lo sviluppo formale della termodinamica statistica divenne molto più semplice introducendo ipotesi che riguardavano la probabilità che un sistema avesse certe caratteristiche, fosse in un dato stato come anche si dice. Questo passo fu compiuto da J. W. Gibbs. Come è stato chiaramente posto in risalto da A. Baracca (1980, p. 213), nel 1902 Gibbs elaborò

> una teoria statistica formale che prescinde completamente dalla considerazione degli urti fra le molecole e mira a riprodurre direttamente le relazioni termodinamiche fondamentali a partire da una ristretta base assiomatica statistica. I corpi sono ovviamente, per Gibbs, composti di molecole, ma il formalismo statistico che egli introduce prescinde dal loro specifico comportamento dinamico.

Le ipotesi di Gibbs, sostanzialmente la validità di certe distribuzioni di probabilità, erano state da lui proposte senza alcun tentativo di giustificarne la validità ma solo sulla base della constatazione che queste distribuzioni consentivano di dedurre ciò che interessava. Quelle distribuzioni concernevano le probabilità di classi di sistemi di molecole: nel caso in cui i sistemi siano perfettamente isolati, la distribuzione uniforme, si suppone cioè che i sistemi della classe siano equiprobabili: è il caso detto microcanonico, l'unico di cui mi occuperò dal momento che rappresenta il più semplice dei (tre) casi presi in considerazione del fisico americano.

Prima di procedere, tuttavia, vale la pena di soffermarsi brevemente sul ruolo dell'energia supponendo che le classi d'energia possibili siano quelle con energia minima, per comodità indichiamo questa classe con 0, con energia immediatamente superiore, 1, e poi 2, 3, 4 e così di seguito. Va da sé che, in assenza di vincoli sull'energia, cioè a dire quando l'energia del sistema può essere qualsivoglia, le possibili distribuzioni delle particelle sulle classi saranno: quella in cui tutte le particelle sono nella classe 0, nella classe 1, ..., nella classe n, la distribuzione in cui tutte le particelle sono nella classe 0 ad eccezione di una che è in 1, quella in cui sono tutte in 0 ad eccezione di due che sono in 1 e così di seguito. Si vede però immediatamente che, qualora l'energia del sistema abbia un valore fissato, non tutte le distribuzioni saranno possibili dal momento che in certi casi la somma delle energie delle molecole potrebbe essere diversa da quel valore.

Trascurando casi più complessi – la sostanza della questione che stiamo per presentare non dipende né dal numero delle molecole né da quello delle classi d'energia, tanto vale quindi considerare il caso più semplice – consideriamo una popolazione o sistema, come dicono i fisici, costituito da due molecole, di nuovo X e Y, e da due stati d'energia o celle come vengono anche detti, diciamo 0 e 1, e occupiamoci dei vari modi possibili di distribuire le molecole del nostro sistema sulle celle (d'energia) avvertendo che ogni possibile distribuzione rappresenterà uno dei possibili stati del sistema. Qualora con $X \in 0$ si indichi che la molecola X è nelle cella 0, le possibilità sono le seguenti: $X \in 0 \cap Y \in 0$, $X \in 0 \cap Y \in 1$, $X \in 1 \cap Y \in 0$ e $X \in 1 \cap Y \in 1$. Ora consideriamo un'assunzione di equiprobabilità che i fisici chiamano statistica di Maxwell-Boltzmann: consiste nell'ipotizzare l'equiprobabilità di tutti possibili stati del sistema; ciò significa assegnare a ciascuno degli stati appena individuati probabilità pari a 1/4. L'ipotesi nulla di cui ci occupiamo è quindi la statistica di Maxwell-Boltzmann.

Un momento di riflessione consente di esplicitare il legame tra la statistica di Maxwell-Boltzmann, da un lato, e, dall'altro, l'indipendenza e la costanza della probabilità. Se supponiamo che l'assegnazione delle molecole agli stati d'energia avvenga in modo sequenziale, cioè prima assegniamo X ad una cella e poi Y ad un'altra cella, allora la probabilità di uno dei possibili modi di distribuirsi delle molecole nelle celle sarà il prodotto delle probabilità di assegnare le due molecole alle due celle, ad esempio,

$$P(X \in 1 \cap Y \in 1) = P(X \in 1)P(Y \in 1 | X \in 1) \tag{6}$$

Questo vale ovviamente per tutte e quattro le possibilità, ed è immediato constatare che per arrivare all'equiprobabilità i due termini del prodotto a sinistra della (6) debbono essere entrambi pari ad 1/2. Questo significa:

$$P(Y \in 1) = P(Y \in 1 | X \in 1) \tag{7}$$

e

$$P(X \in 1) = P(Y \in 1) = 1/2 \tag{8}$$

Ora notiamo che mentre la (8) sembra naturale, almeno quando la si attribuisca alla casualità l'accomodarsi di una molecola in una cella, la (7) è chiaramente un'ipotesi d'indipendenza. Quindi l'equiprobabilità della statistica di Maxwell-Boltzmann si fonda in modo essenziale sull'indipendenza stocastica degli accomodamenti delle molecole o, come a volte abbreviando si dice, sull'indipendenza delle molecole.

Passando al caso generale, consideriamo quindi un sistema costituito da N molecole e d celle caratterizzate dall'energia ε_j, $j = 1, ..., d$, una descrizione individuale del sistema sarà

$$D = X_1 \in j_1 \cap ... X_n \in j_n \cap ... \cap X_N \in j_N\,, \qquad j_n \in \{1,...,d\} \tag{9}$$

Il nome della (9) le viene dal fatto che essa fornisce una descrizione particolareggiata delle molecole del sistema nel senso che per ogni molecola specifica la cella nella quale si trova. Come abbiamo visto, la statistica di Maxwell-Boltzmann assegna la stessa probabilità a tutte le descrizioni individuali del sistema, vale a dire, qualunque sia D,

$$P(D) = d^{-N} \tag{10}$$

dal momento che d^N è il numero delle possibili descrizioni individuali del sistema. Dalla (10) segue la distribuzione multinomiale sui vettori d'occupazione del sistema cioè sui vettori che descrivono quante molecole sono in ciascuna cella. Infatti, dato un siffatto vettore $\mathbf{N} = (N_1,...,N_r,...,N_d)$, $\sum_{r=1}^{d} N_r = N$, tenendo conto che vi sono $\dfrac{N!}{N_1!...N_d!}$ descrizioni individuale compatibili con $\mathbf{N}$, abbiamo

$$P_{MB}(\mathbf{N}) = \frac{N!}{N_1!...N_d!} d^{-N} \tag{11}$$

Ciò che abbiamo visto è in sostanza il passaggio dall'ipotesi nulla, la (10), alla distribuzione campionaria, la (11). A questo punto la metodologia statistica introduce i livelli di significatività con tutto quel che segue. Ma il contesto nel quale lavorano i fisici, caratterizzato da popolazioni con miriadi di miriadi di unità – è dell'ordine di 10^{23} il numero di molecole di un litro di ossigeno alla pressione di una atmosfera e alla temperatura di zero gradi centigradi – cambia qualitativamente la questione. La (11) è una distribuzione così concentrata attorno alla sua moda che numeri di occupazione diversi da quello modale hanno probabilità trascurabili.

Quindi una volta individuato il massimo della (11) abbiamo anche individuato la distribuzione nelle varie celle d'energia delle molecole del gas in esame. Tuttavia la determinazione della moda non può essere fatta mediante l'analisi della (11) poiché è necessario tener presente il vincolo imposto dall'energia totale del sistema. Il problema non è quindi di calcolare un massimo bensì di calcolare il massimo della (11) vincolato, ovviamente, al numero delle molecole del sistema, cioè $\sum_{j=1}^{d} N_j = N$, e poi alla sua energia totale, cioè $\sum_{j=1}^{d} \varepsilon_j N_j = E$ essendo ε_j l'energia comune a tutte le molecole della j-esima cella. La ricerca di questo massimo è tutt'altro che semplice a meno di supporre che tanto il numero delle particelle quanto quello delle celle siano talmente elevati da rendere applicabile sia l'approssimazione di Stirling per i fattoriali sia il metodo dei moltiplicatori di Lagrange. Se le cose stanno in questo modo la risposta diviene semplice: il vettore di occupazione per il quale la distribuzione di probabilità (11) raggiunge il massimo è quello i cui numeri d'occupazione, espressi come frequenze relative, sono dati dall'espressione

$$\frac{1}{\exp(a + b\varepsilon_j)}, \quad j = 0, 1, 2, ..., d \tag{12}$$

in cui a e b sono costanti dipendenti dal potenziale chimico e dalla temperatura che ora non è di alcun interesse precisare. La (12) è detta la legge di distribuzione di Maxwell-Boltzmann perchè, come abbiamo visto, si riferisce ad un vettore di occupazione quindi ad una distribuzione delle molecole sulle celle del sistema. Vale la pena di sottolineare che la (12) non individua una probabilità né tantomeno una distribuzione di probabilità bensì una distribuzione di molecole. A partire dalla (12) si deriva l'equazione di stato dei gas perfetti e con essa tutte le proprietà interessanti dei gas (di molecole) rarefatti, come il libero cammino medio delle molecole, la viscosità, la conducibilità termica e così di seguito.

Fin qui, con le semplificazioni che ho detto, abbiamo sostanzialmente seguito le argomentazioni di Boltzmann e Gibbs, siamo cioè tra la fine dell'Ottocento e l'inizio del Novecento. Ma come abbiamo visto, negli ultimi anni dell'Ottocento, i ricercatori dell'Istituto imperiale di Berlino scoprirono che la (12) non era in grado di rendere ragione di ciò che l'osservazione della radiazione di un corpo nero consentiva di costatare. Un poco più precisamente, il comportamento di questa radiazione dedotto dalla (12) è conforme a ciò che avviene alle basse frequenze ma si distanzia notevolmente da quel che accade alle alte frequenze; tra l'altro prevede la catastrofe ultravioletta, cioè a dire l'emissione tende all'infinito con la frequenza, che gli esperimenti indicano non avvenire. Quindi la (12) è incapace di descrivere ciò che l'osservazione mostra e, inoltre, prevede eventi che non accadono; per quel che concerne la radiazione di corpo nero, la (12) è in contrasto con la realtà e conseguentemente lo è anche la (10) da cui la distribuzione di Maxwell-Boltzmann viene dedotta; detto altrimenti, l'ipotesi (nulla) (10) fu confutata dalle osservazioni. Il fatto notevole fu che la comunità scientifica prese atto di questa confutazione senza che alcuno avesse ancora formulata una qualsivoglia ipotesi alternativa in grado di rendere ragione dello spettro della radiazione di corpo nero. Dalla confutazione della (12), e quindi della (10), e solo da questa confutazione, prese le mosse la ricerca di Planck; l'individuazione dell'ipotesi quantistica lo mise in grado, seguendo Boltzmann, di scrivere l'entropia come una sommatoria che, una volta massimizzata, coi vincoli già visti, lo condusse alla legge di distribuzione di Planck, vale a dire alle frequenze relative espresse dalla formula

$$\frac{1}{\exp(a+b\varepsilon_j)-1}, \quad j = 0, 1, 2, ..., d \tag{13}$$

che si riferisce ai bosoni – i fotoni sono particelle di questo tipo – ed è l'analogo della (12) quando, ovviamente, si tenga conto delle differenze che sussistono fra bosoni e molecole. Dalla (13) si deduce la forma della distribuzione dell'energia della radiazione di corpo nero che ha un andamento in perfetto accordo con quello che l'esperienza mostra.

A questo punto è necessario notare che Planck non si occupò, per lo meno esplicitamente, di formulare un'ipotesi probabilistica da sostituire alla (10); il grande fisico tedesco lavorò sull'ipotesi quantistica, la cui natura non è probabilistica, ma da sola l'ipotesi dell'energia discreta non è sufficiente a dedurre la (13); era necessaria un'altra ipotesi che Planck aveva implicitamente assunto: il comportamento dei quanti d'energia non è stocasticamente indipendente; ciò diverrà chiaro oltre vent'anni dopo.

Ciò che ho detto ha posto chiaramente in luce come la comunità scientifica non abbia avuto bisogno di alcuna ipotesi alternativa; l'ipotesi nulla era stata confutata, ciò fu sufficiente per spingere alla ricerca di un'ipotesi che potesse spiegare quello che accadeva. Questa ricerca andò avanti per circa vent'anni; in tutto quel periodo non si formulò alcuna ipotesi alternativa a quella (nulla) confutata dagli sperimentatori dell'Istituto imperiale di Berlino. E non poteva essere altrimenti perché la dipendenza stocastica, implicitamente introdotta da Planck, avrebbe potuto diventare palese solo nel momento in cui si fossero esplicitate le caratteristiche probabilistiche dei quanti di energia. Questa esplicitazione avvenne oltre due decenni dopo la pubblicazione in cui Plank rese nota la sua legge di distribuzione, cioè nel 1924 quando S. N. Bose e A. Einstein idearono la distribuzione probabilistica che poi sarebbe stata chiamata la statistica di Bose-Einstein; vale la pena di notare che questa statistica ipotizza l'equiprobabilità dei vettori di occupazione del sistema e, pertanto, è in aperto contrasto con la (10). La statistica di Bose-Einstein, che vista alla luce della teoria di Neyman e Pearson è un'ipotesi alternativa, venne quindi concepita molti anni dopo la confutazione della (10). A rischio d'essere noioso lo ripeto ancora una volta: l'ipotesi nulla venne confutata e pertanto abbandonata ben prima che si fosse in grado di formulare un'ipotesi alternativa. Questa caratteristica tipica di ogni argomentazione ipotetico-deduttiva frequentemente avanzate nelle scienze naturali, è un inoppugnabile dato di fatto anche se la stragrande maggioranza degli statistici, come Lehmann nel lavoro citato in precedenza, dà per scontata l'assunzione dei saggi di significatività fra quelli di ipotesi. Al di là d'ogni dubbio, l'esempio che abbiamo visto mostra che questa assunzione non è possibile; le dure parole di Fisher che ho dianzi ricordato sottolineano questa impossibilità.

6. GLI INTERVALLI DI FIDUCIA

Nel venire agli intervalli di fiducia, torno sulla posizione di Neyman nei confronti della stima statistica cui ho accennato parlando di E. Pearson nel secondo paragrafo. Nel già citato volume in cui Neyman con grande impegno affronta i fondamenti della statistica inferenziale, particolarmente nel IV capitolo (Neyman, 1952, pp. 165-269) dedicato alla stima, viene posto l'accento sui casi in cui può essere usato l'approccio bayesiano, sui problemi che pone la scelta della distribuzione iniziale e sulle possibilità di ovviare, mediante la teoria degli intervalli di fiducia, alla supposta inesistenza di siffatte distribuzioni. È opportuno notare che questo volume, pur essendo stato pubblicato solo nel 1952, è basato su lezioni e conferenze tenute quindici anni prima e, sia pure come pubblicazione ciclostilata, era apparso fin dal 1938 (Neyman, 1952, p. v). Pertanto, a differenza di quelle già ricordate di E. Pearson, si tratta di idee messe a punto proprio nel periodo in cui Neyman stava elaborando la sua teoria degli intervalli di fiducia.

Venendo alla stima statistica, noto che per il nostro autore l'impossibilità di pervenire alle probabilità iniziali è una conseguenza della sua filosofia frequentista della probabilità. A questo proposito val la pena di ricordare il seguente passo che chiarisce la posizione di Neyman (Neyman, 1952, p. 194)

> Tutte le volte che le condizioni di un particolare problema implicano che il parametro sconosciuto è una variabile casuale con una specifica distribuzione a priori, la formula di Bayes fornisce una soluzione chiara e rigorosa al problema della stima; questa soluzione [...] ha una semplice interpretazione in termini di successi [ottenuti] nello stimare il parametro sconosciuto; quando le condizioni del problema pratico considerato non implicano la distribuzione a priori del parametro [...], allora è ancora "legittimo" usare la formula di Bayes; tuttavia, nonostante il teorema di Bernstein-v.Mises e i tentativi di stimare la distribuzione a priori servendosi dell'esperienza acquisita, una siffatta applicazione della formula di Bayes ha una dubbia interpretazione frequentista.

Come mostra questa citazione, la situazione è ben lontana da quella che di solito viene presentata nei testi di statistica inferenziale; per questa ragione è indispensabile una distinzione ché, altrimenti, si rischia di fare d'ogni erba un fascio.

Neyman, in primo luogo, sostiene che l'uso della regola di Bayes fornisce un'ottima soluzione al problema della stima quando si abbia la possibilità di interpretare la probabilità iniziale in termini frequentisti; questa posizione fu anche di Fisher. In secondo luogo, e qui la differenza con Fisher è profonda, che l'uso della regola è sempre legittimo anche quando la distribuzione iniziale non sia riconducile ad una frequenza relativa, ad esempio quando la distribuzione iniziale sia determinata sulla base del principio d'indifferenza magari riformulato in termini moderni. In terzo luogo, che qualora si dia l'eventualità testé vista, Neyman non pone in dubbio la liceità dell'uso della regola bensì l'interpretazione frequentista del risultato cioè a dire della distribuzione finale. Come abbiamo visto questa è sostanzialmente la posizione di von Mises anche se per la verità questo autore pone sempre in risalto come l'approccio alla stima non possa essere altro che probabilistico. Quindi per Neyman le difficoltà, che abbiamo visto essere di natura interpretativa e quindi filosofica, sorgono quando non sia possibile ancorare la distribuzione iniziale alla frequenza relativa. Ed ecco allora il metodo che, negli anni trenta del secolo scorso, Neyman sviluppò per arrivare ad affermazioni probabilistiche, frequentisticamente intese, circa il valore sconosciuto θ di una grandezza, che divenne noto come il metodo dei limiti o intervalli di fiducia. L'impostazione che Neyman diede al problema, pur essendo ineccepibile, presenta qualche oscurità sia nel frequentismo che guida la sua deduzione sia, e soprattutto, in una ingiustificata commistione di due interpretazioni, una probabilistica e l'altra operazionale, dei risultati raggiunti. Per questa ragione seguirò ancora von Mises la cui lucidità di esposizione è sempre notevole. Sostiene infatti questo autore (von Mises, 1964, p. 537):

> si suppone nota la distribuzione $p(x|\theta)$ [x denota le osservazioni sperimentali], dipendente da un parametro θ, mentre è sconosciuta la distribuzione [iniziale] $p(\theta)$. Di nuovo, dovremo formulare asserzioni circa θ cercare di trovare la probabilità che queste siano corrette.
> Se entrambe $p(x|\theta)$ e $p(\theta)$ sono densità, la densità della probabilità del verificarsi di definiti valori di x e θ è il prodotto
>
> $$p(x|\theta)p(\theta) \tag{14}$$

in cui, come ognuno vede, il problema è posto in termini probabilisticamente ineccepibili. Von Mises considera poi una regione R dello spazio prodotto di tutti i possibili valori di x e θ e prosegue (vedi [12] p. 537)

> Nell'infinita successione degli esperimenti, ciascun valore osservato x è collegato a un definito valore di θ, a ciascuna prova corrisponde un dato punto nel piano dei valori di x e θ. La probabilità (frequenza limite) dei punti che cadono nella regione R è, per (14),
>
> $$P(\mathrm{R}) = \iint p(x|\theta)p(\theta)dxd\theta$$
>
> [l'integrazione è relativa alla regione summenzionata] e, in generale, questa quantità può essere calcolata solo se si conosce $p(\theta)$. Vedremo tuttavia che esistono regioni speciali R tali che $P(R)$ può essere trovata indipendentemente da ogni conoscenza o assunzione relativa a $p(\theta)$.

Sarebbe istruttivo seguire almeno alcune delle derivazioni che consentono nei vari casi di pervenire alla conoscenza della probabilità di R ma questo mi porterebbe troppo lontano. Mi limito pertanto a ricordare un'argomentazione intuitiva fornita da von Mises che fa riferimento ad una semplice rappresentazione geometrica dello spazio prodotto definito dai valori di x e θ. In questo caso, alla probabilità che interessa si perviene considerando, dapprima, per ogni valore di θ, un intervallo di valori di x delimitato dai punti $x_1(\theta)$ e $x_2(\theta)$, integrando rispetto a x su questo intervallo si perviene alla verosimiglianza α che, dato un valore del parametro sconosciuto, stabilisce la probabilità con cui il valore di x sarà interno al suddetto intervallo. Dopo di che, di nuovo integrando su tutti i valori di θ, quindi in corrispondenza di tutte le possibili verosimiglianze il cui valore è α, si perviene alla probabilità della regione *R* che risulta pari ad α. È opportuno attirare l'attenzione sul significato frequentistico di questa doppia integrazione che, vista alla luce dei due collettivi antro i quali sono definiti x e θ, consente di ottenere un risultato che non dipende dalla distribuzione iniziale $p(\theta)$. Vengo allora alle conclusioni di von Mises (1964, p. 539):

> se, a seguito di ogni osservazione x, affermiamo che θ giace nell'intervallo compreso fra $\theta_1(x)$ e $\theta_2(x)$, in cui $\theta_1(x)$ e $\theta_2(x)$ sono il minimo e il massimo dei valori di θ la cui ascissa x è nella regione [*R*], abbiamo una probabilità pari ad α di essere corretti qualunque possa essere la probabilità iniziale. Se, ad esempio, α = 0,9, siamo sicuri che a lungo andare [*in the long run*] il 90% delle nostre asserzioni saranno corrette.

in cui il riferimento al lungo andare è un'imprescindibile necessità. Dopo aver posto in risalto l'interpretazione frequentista su cui poggia il metodo degli intervalli di fiducia di Neyman, von Mises rende espliciti tanto il significato quanto la portata della derivazione di Neyman (von Mises, 1964, pp. 539 e 540)

> Si è visto che si può pervenire ad una probabilità di successo elevata quanto vogliamo. D'altro canto, le asserzioni per le quali vale questa probabilità di successo sono strettamente stabilite. Nel problema di saggiare un'ipotesi, dopo ciascuna osservazione di un valore di x intendevamo sostenere che θ giaceva in un *intervallo precedentemente determinato*, lo *stesso* intervallo qualunque fosse il valore x osservato. Ora, seguendo il metodo degli intervalli di fiducia, non siamo liberi di scegliere questi intervalli. Nel nostro originale problema inferenziale, intendevamo ancora di più: consideravamo solo quei casi in cui certi $x = x_1$ erano stati osservati e sostenevamo che θ giacesse in un dato intervallo, subordinatamente allo specifico valore x_1. L'elevata probabilità di successo del metodo degli intervalli di fiducia è raggiunta a scapito della libertà di formulare le asserzioni.

Infatti, in corrispondenza di ogni gruppo di osservazioni x esiste un intervallo che per un dato θ ha il valore α fissato ma questo intervallo è diverso per ogni x. Ed è proprio questo dato di fatto la ragione della imprescindibile necessità di cui ho dianzi detto come è semplice costatare per lo meno quando ci si ponga nella prospettiva dalla quale von Mises vide la questione. La probabilità, α, di aver avanzato un'asserzione corretta non è valida per ogni gruppo o per un numero finito di gruppi siffatti bensì per l'intera infinita successione $x_1, x_2, ..., x_n, ...$ di tutti i gruppi di osservazioni che potrebbero essere compiuti a seguito di un'infinità di esperimenti, vale a dire il collettivo nel senso in cui lo intendeva von Mises. Solo un'infinità di osservazioni, casuali sempre nel senso di von Mises, origina un collettivo in cui per così dire si situa la densità $p(x,\theta)$, la (14), ed è solo con riferimento a questo collettivo che ha senso parlare della probabilità della regione *R*. Detto ancora più chiaramente, è solo la concezione frequentista della probabilità elaborata da von Mises che consente di determinare la probabilità della regione *R* a prescindere da ogni conoscenza o assunzione relativa alla distribuzione iniziale. Attribuire la probabilità α ad un unico gruppo di osservazioni sperimentali riapre, in sostanza, il vecchio e molto dibattuto problema del caso singolo col quale s'è invano scontrata l'interpretazione frequentista della probabilità. α non è la probabilità di un termine del collettivo, quindi di un'asserzione, bensì il limite della frequenza relativa delle asserzioni corrette relativamente alle infinite volte in cui l'asserzione avrebbe potuto essere compiuta. Insomma: α è una probabilità frequentista cioè a dire è sensata solo quando venga definita entro un collettivo, con tutto quello che questa definizione si porta appresso. Sono chiari allora i limiti, sia di interpretazione sia di portata, delle inferenze volte all'individuazione di intervalli di fiducia: il ricorso alle probabilità iniziali può essere evitato purché si sia consapevoli di questi limiti. Questi limiti, sia pure relativi a momenti diversi dell'inferenza, sono della medesima natura di quelli individuati dal teorema di Bernstein-von Mises che lo stesso Neyman ricorda. In sostanza, in entrambi i casi si tratta della necessità di operare un numero molto elevato di osservazioni; infatti, per il teorema suddetto, affinché la probabilità iniziale possa essere trascurata deve tendere all'infinito la numerosità del campione, parimenti per gli intervalli di fiducia la probabilità iniziale può essere trascurata se il numero delle stime tende all'infinito. Ed è solo in questa eventualità che ha senso ricorrere alla probabilità della regione *R*; quando non si da questa eventualità quella probabilità è vuota di senso. Ma – mi chiedo – quanti autori, segnatamente quanti di quelli di lingua italiana nei loro libri di testo rendono consapevoli i loro studenti dei suddetti limiti? Pochi, davvero pochi, purtroppo.

7. LE PROBABILITÀ FIDUCIALI

Fin qui le idee di Neyman. Tuttavia, con pochissime eccezioni, gli intervalli di fiducia non vengono presentati lungo le linee che ho esposto, bensì dedotti, si fa per dire, servendosi di un'argomentazione che, non so se intenzionalmente o meno, è analoga a quella usata da Fisher per la sua teoria delle probabilità fiduciali. A questo proposito vale la pena di ricordare che immediatamente dopo la formulazione fisheriana, per primo e senza timori reverenziali, Neyman denunciò la confusione e il gravissimo errore logico su cui erano basate le suddette probabilità. Riferendosi al lavoro di Fisher e a quello di P. V. Sutkhatme che ne seguì le orme, Neyman afferma che (Neyman, 1952, p. 213)

> dal punto di vista dell'usuale teoria delle probabilità la soluzione di Fisher-Sukhatme [la probabilità fiduciale] è sbagliata. L'errore consiste nell'aver confuso la legge della probabilità assoluta [...] con quella della probabilità relativa [...] delle stesse variabili

in cui probabilità assoluta sta per verosimiglianza e probabilità relativa per probabilità finale.

Anche Gini, in termini sostanzialmente analoghi, aveva denunciato lo stesso errore. Le sue critiche furono poi riprese da G. Pompilj che diede ad esse una chiarezza adamantina – assente tanto nelle argomentazioni di Gini quanto in quelle di Neyman – riformulandole in termini di variabili casuali e delle rispettive probabilità. Purtroppo queste denuncie sono passate come acqua sul marmo: nessuno sembra più rammentarle; per questo motivo mi pare opportuno tornarvi con qualche dettaglio.

Riferendosi esplicitamente agli intervalli di fiducia, Pompilj sostiene (Pompilj, 1948, p. 20)

> Chi avesse seguito la strada da noi percorsa, non sarebbe potuto cadere nell'errore sopra lamentato, perché da essa risulta evidente il salto logico che si deve fare per trasformare questa semplice subinversione in un'inversione. Ma in realtà si sono seguite altre argomentazioni che, a causa di una certa incompletezza di simbolismo, hanno condotto direttamente, se non inevitabilmente, all'errore! Vale la pena di percorrere tale strada per mettere bene in luce il punto delicato della questione.

Una nota a piè di pagina, dopo l'ultima parola della citazione, avverte (Pompilj, 1948, p. 20 nota 12)

> Mi risulta che un'osservazione analoga è stata sviluppata dal Prof. Gini in una comunicazione del Seminario di Statistica dell'Università di Roma, in data 9-VII-1946 e non ancora pubblicata.

Dopo questa chiarissima premessa in cui Pompilj, come già aveva fatto Jeffreys ma non Neyman, individua nella incompletezza del simbolismo una delle ragioni dell'errore, torna a considerare un esempio fatto in precedenza dicendo

> Si è gia detto [...] che, fissato il valore x della media della v.c. genitrice, la media y del campione soddisfa con probabilità $\Theta = 0,997$ alla disuguaglianza
>
> $$x - 3\frac{\mu}{\sqrt{N}} \leq y \leq x + 3\frac{\mu}{\sqrt{N}} \tag{15}$$
>
> Sottraendo dai tre membri della (15) la quantità $x + y$ e moltiplicando il risultato per -1, si ottiene
>
> $$y - 3\frac{\mu}{\sqrt{N}} \leq x \leq y + 3\frac{\mu}{\sqrt{N}} \tag{16}$$
>
> relazione quest'ultima ancora valida con probabilità Θ, essendo stata ottenuta dalla (15) con semplici trasformazioni algebriche.

In quanti testi compaiono queste "semplici trasformazioni algebriche" come ironicamente dice Pompilj? Semplice trasformazione sì ma viziata da un errore a ragione del fatto che, come nota Pompilj, i simboli x e y hanno funzioni diverse. Infatti x denota il valor medio noto, o supposto tale, della popolazione genitrice, cioè la popolazione da cui sarà tratto il campione il cui valor medio incognito sarà y. Pertanto la (15) (Pompilj, 1948, p. 21)

> andava scritta più correttamente così
>
> $$x - 3\frac{\mu}{\sqrt{N}} \leq \Psi \leq x + 3\frac{\mu}{\sqrt{N}} \tag{17}$$

se poi si vuole scrivere dentro la formula stessa anche il fatto che essa è valida con probabilità Θ, si potrebbe introdurre il simbolo $[\Xi]\Theta$ e scrivere:

$$x-3\frac{\mu}{\sqrt{N}} \leq [\Psi]0{,}997 \leq x+3\frac{\mu}{\sqrt{N}}$$

Se invertiamo ora analiticamente quest'ultima formula si arriva solo alla relazione:

$$[\Psi]0{,}997-3\frac{\mu}{\sqrt{N}} \leq x \leq [\Psi]0{,}997+3\frac{\mu}{\sqrt{N}}$$

e non già alla

$$y-3\frac{\mu}{\sqrt{N}} \leq [\Xi]0{,}997 \leq y+3\frac{\mu}{\sqrt{N}}$$

che corrisponde alla maniera in cui in tanti testi di statistica inferenziale viene letta la (16). In nota Pompilj aggiunge (1948, p. 21 nota 13)

> In altre parole una successione di passaggi formali non può mai condurre all'inversione statistica della relazione originaria, ma solo ad una diversa espressione; come ha fatto osservare all'Anderson il prof. Gini nella discussione che [...] è seguita all'articolo del prof. Anderson.

L'articolo menzionato nella citazione è del 1947; non mette conto citarlo.

È invece di grande importanza rammentare che, con altrettanta chiarezza, Jeffreys aveva denunciato il grave errore logico, del tutto analogo a quello testé visto, che caratterizza l'argomento fiduciale di Fisher applicato in particolare alla derivazione di Student. Riferendosi ad un passaggio matematico, dello stesso tipo delle semplici trasformazioni algebriche di Pompilj, Jeffreys dice (Jeffreys, 1961, p. 381)

> Tuttavia, la trasformazione matematica [di Fisher] che sembra del tutto innocente copre il passaggio dai dati x e σ [il valor medio e lo scarto quadratico medio della popolazione] ai dati $\bar{x}$ e s [il valor medio e lo scarto quadratico medio del campione] [...] che la notazione usata non è adeguata ad esprimere.

Le considerazioni di Neyman, Jeffreys, Gini e Pompilj evidenziano un dato di fatto troppo frequentemente dimenticato: nessuna manipolazione matematica è in grado di farci passare da una verosimiglianza a una probabilità finale. Come Gini aveva lucidamente notato, un'inversione statistica non può essere compiuta mediante una successione di passaggi formali. La regola di Bayes-Laplace, per il tramite della probabilità iniziale, è la sola via logicamente corretta di passare da una verosimiglianza a una probabilità finale. Altra strada non esiste ad eccezione del caso particolarissimo in cui la distribuzione iniziale sia uniforme, come è noto almeno dal 1939, l'anno in cui fu pubblicato il trattato di Jeffreys da cui ho tratto la citazione di poc'anzi.

8. CONCLUSIONE

Non c'è molto da aggiungere a quanto abbiamo visto se non sottolineare come alla base dell'epistemologia di Neyman e Pearson vi fosse un profondo convincimento frequentista che si estrinsecava in una fiducia, tutto sommato irrazionale, in quel che accade a lungo andare. Come ha esplicitamente detto lo stesso Neyman (1952, p. 232)

> Le teoria degli intervalli di fiducia è stata costruita per fornire una soluzione ai problemi di stima che avesse un chiara interpretazione frequentista, caratteristica del punto di vista classico

in cui l'uso dell'aggettivo "classico" è improprio come ho già avuto modo di notare. Tuttavia questa è una fiducia che, nella migliore delle ipotesi, si può nutrire nelle eventualità in cui l'inferenza debba essere operata numerose volte, ad esempio e come fu precisato dallo stesso Neyman, quattro volte alla settimana (Neyman, 1952, p. 213), ovviamente per un numero elevato di settimane, in modo che il lungo andare sia lungo per davvero. In sintesi: le tesi di Neyman e Pearson, sono accettabili – anche se ciò non deve indurre a dimenticare la differenza che sussiste fra un'assunzione teorica e un comportamento pratico – solo a livello pratico e con riferimento a certe ben precise applicazioni che, di solito, sono tipiche della gestione di un'azienda. Dopo questo riconoscimento è però doveroso ricordare che la posizione fondazionale basata sul lungo andare era e resta molto debole per almeno tre ragioni. In primo luogo, il lungo andare è inscindibile dall'interpretazione operazionale di Neyman connessa ad una regola di comportamento, si tratta cioè del famoso comportamento induttivo che, come dovrebbe essere noto, ha aperto più problemi di quanti ne abbia chiusi. In secondo luogo, la scarsa attenzione prestata alla definizione di probabilità tipica di questo approccio presta il fianco alle critiche di coloro che rifiutano un'interpretazione oggettiva della probabilità. È infatti fin troppo facile obiettare: come sono connessi il lungo periodo e il limite della frequenza relativa? Detto altrimenti, quanto deve essere lungo il lungo periodo affinché le ipotesi dei neymaniani possano trovare riscontro nella realtà? A rigore, il lungo andare è l'infinito che, come tutti sappiamo, è piuttosto lungo. Noto per inciso che Neyman, pur mostrando di conoscere bene i lavori di von Mises, non fa riferimento ai collettivi e pertanto né alla casualità della successione né al limite senza i quali qualunque definizione frequentista, operazionale o meno, è semplicemente vana. E forse questo non è un caso giacché nella realtà non si danno collettivi. Infine, quando la metodologia neymaniana basata sul comportamento induttivo non venne più invocata per rendere ragione di ciò che accadeva nel lungo periodo bensì usata per controlli che, nelle eventualità migliori, erano eseguiti una decina di volte, divenne palese che la regola neymaniana poteva essere giustificato solo sulla base di convinzioni personali. Questo fu proprio il caso di Leonard Savage che parlando del suo *The Foundations of Statistics* ammette quanto segue (Pearson, 1962, p. 9)

> Sebbene questo libro attiri l'attenzione sui meriti del concetto di probabilità soggettiva (o personale), non è stato scritto per anticipare cambiamenti radicali nella pratica statistica. L'idea era piuttosto che la probabilità soggettiva avrebbe condotto ad una migliore giustificazione della statistica, come era allora praticata, senza avere alcuna immediata conseguenza pratica. Tuttavia, da allora è diventato sempre più chiaro che il concetto di probabilità soggettiva è capace di suggerire e unificare importanti progressi nella pratica statistica.

vale a dire, avendo preso atto che un'interpretazione soggettivistica della probabilità era in grado di rendere ragione di ciò che di fatto avveniva nella prassi statistica, Savage e molti altri con lui buttarono via il presunto oggettivismo, relativo a periodi che ormai erano divenuti fin troppo brevi, per l'estremismo filosofico di de Finetti. Il cerchio s'era chiuso. L'abbandono del rigore fondazionale di von Mises aveva condotto ad applicare la statistica inferenziale in contesti che non avevano nulla da spartire né coi collettivi né con la loro casualità. Per conservare l'oggettività di queste applicazioni l'unica via d'uscita sarebbe stata il ricorso a probabilità iniziali basate su conoscenze intersoggettive elaborate in accordo con una riformulazione logicamente ineccepibile del principio d'indifferenza. Il pregiudizio empirista che Neyman e Pearson condividevano con Fisher, impedì loro di percorrere questa strada. Fu quindi facile per la critica definettiana screditare tanto il loro gratuito oggettivismo quanto, nel caso dei saggi di ipotesi, le loro fallacie logiche. Nei fatti il fallimento del tentativo di giustificare le inferenze statistiche sulla base del frequentismo era già presente, sia pure *in nuce*, nel ricorso al comportamento induttivo. Nondimeno confidando in un oggettivismo ad un tempo confuso e risibile, le tecniche di Neyman e Pearson continuano ad essere largamente usate, segnatamente da coloro che non si prendono la briga di meditare su quello che stanno facendo. Tuttavia con la prepotente rinascita del bayesianesimo, in particolare soggettivistico, la sconfitta dell'ambiguo frequentismo di Neyman e Pearson divenne palese. Sulle macerie che questa sconfitta aveva lasciato dietro di sé, si instaurò un nuovo dogmatismo ben più pericoloso del vecchio legato al frequentismo: il dogmatismo di de Finetti.

Certamente Amato Herzel e Alighiero Naddeo si batterono con vigore a favore dell'oggettività della statistica inferenziale; tuttavia, come ho cercato di mostrare, la metodologia neymaniana poteva condurre in una direzione soggettivistica; anzi, col senno di poi, si può dire che si trattava di una deriva obbligata. Avendo imboccato la strada suggerita da Neyman e Pearson non v'era alcuna speranza di vincere la battaglia per l'oggettivismo. Ma le battaglie perse non sono inutili. A mio parere questa sconfitta insegna che la difesa dell'oggettivismo deve essere compiuta seguendo linee strategiche diverse da quelle adottate da Neyman e Pearson. Questa difesa deve mettersi lungo la strada indicata da Fisher, cioè quella suggerita dalle scienze della natura; ma questo sarà l'argomento di un prossimo lavoro.

DOMENICO COSTANTINI

Avvertenza. *Sono mie tutte le traduzioni dall'inglese.*

RIFERIMENTI BIBLIOGRAFICI

A. BARACCA, (1980), *Manuale critico di meccanica statistica*, CULC, Catania.
B. DE FINETTI, (1931), *Probabilismo*, "Logos", 14, Napoli.
R. A. FISHER, (1922), *On the mathematical foundations of theoretical statistics*, "Philosophical Transactions of the Royal Society of London", Serie A, 222, pp. 309-368.
R. A. FISHER, (1956), *Statistical Methods and Scientific Inference*, Oliver and Boyd, Edinburgh and London.
R. A. FISHER, (1958), *Statistical Methods for Research Workers*, (1925) Oliver and Boyd, Edinburgh and London.
R. A. FISHER, (1960), *The Design of Experiments*, (1935), Oliver and Boyd, Edinburgh and London.
J. H. GILLESPIES, (1998), *Population Genetics*, The Johns Hopkins University Press, Baltimore and London.
E. T. JAYNES, (1984), *The intuitive inadeguacy of classical statistics*, "Epistemologia", VII, pp. 43-73.
E. T. JAYNES, *Probability Theory: The Logic of Science*, http//bayes.wustl. edu/etj/prob.html
H. JEFFREYS, (1961), *Theory of Probability*, (1939), Clarendon Press, Oxford.
E. L. LEHMANN, (1993), *The Fisher, Neyman-Pearson theories of testing hypotheses: One theory or two?*, Journal of the American Statistical Association, 88, pp. 1242-1249.
R. VON MISES, (1964), *Mathematical Theory of Probability and Statistics*, Academic Press, New York.
J. NEYMAN, (1952), *Lectures and Conferences on Mathematical Statistics and Probability*, second edition, Graduate school, U. S. Department of Agriculture, Washington.
E. S. PEARSON, (1962), *Prepared contributions: Professor E. S. Pearson*, in L. J. SAVAGE and other contributors (eds.), *The Foundation of Statistical Inferences*, Methuen & Co Ltd, London.
G. POMPILJ, (1948), *Teorie statistiche della significatività e conformità dei risultati sperimentali agli schemi teorici*, "Statistica", riprodotto in "Matematica, probabilità e statistica in rapporto alle scienze empiriche nell'orizzonte di ricerca di Giuseppe Pompilj" (A cura di E. D'Arcangelo) Dip. Statistica, Probabilità e Statistiche applicate, Roma 1993.
G. POMPILJ, (1951), *Lineamenti di una teoria della persuasione*, "Archimede", riprodotto in "Matematica, probabilità e statistica in rapporto alle scienze empiriche nell'orizzonte di ricerca di Giuseppe Pompilj" (A cura di E. D'Arcangelo) Dip. Statistica, Probabilità e Statistiche applicate, Roma 1993.
L. J. SAVAGE, (1962) *Subjective probability and statistical practice*, in L. J. SAVAGE and other contributors (eds.), *The Foundation of Statistical Inferences*, Methuen & Co Ltd, London.

RIASSUNTO

Fisher, von Mises, Jeffreys, Neyman, Pearson e l'inferenza statistica

Si mettono a confronto i fondamenti dell'inferenza statistica di Neyman-Pearson, di Fisher e della stima bayesiana di von Mises e Jeffreys.

SUMMARY

Fisher, von Mises, Jeffreys, Neyman, Pearson and statistical inferences

In the paper are compared the foundations of the statistical inferences of Neyman-Pearson, of Fisher and the Bayesian estimation of von Mises and Jeffreys.