

Induzione e inferenza statistica

di SILVIO VIANELLI

Secondo la definizione di Aristotele, l'induzione è il procedimento che dai particolari porta all'«universale» e tale concetto è stato sostanzialmente ribadito da filosofi e scienziati nei secoli successivi.

Ma dopo che le scienze hanno cominciato ad usare ampiamente il procedimento induttivo, come è avvenuto nella seconda metà del '600, il problema dell'induzione, quale problema della *validità* del procedimento stesso e del diritto di usarlo, è stato posto ed affrontato criticamente con rinnovato interesse.

A parlo chiaramente fu, allora, il dubbio scettico di Hume: «Tutte le inferenze tratte dall'esperienza suppongono, come loro fondamento, che il futuro rassomiglierà al passato e che poteri simili saranno uniti a simili qualità sensibili. Se ci fosse qualche sospetto che il corso della natura potesse cambiare e che il passato non servisse di regola per il futuro, ogni esperienza diverrebbe inutile e non potrebbe dare origine ad alcuna inferenza o conclusione».

Il problema dell'induzione è stato posto in questi termini e ad esso sono state date tre soluzioni filosofiche fondamentali: 1) la soluzione oggettivistica; 2) la soluzione soggettivistica; 3) la soluzione pragmatica. Quest'ultima segna il passaggio dalla concezione necessitaristica — presupposta dalle altre due — ad una concezione probabilistica dell'induzione.

La soluzione oggettivistica è stata discussa e riproposta da vari filosofi e scienziati sino al XIX secolo, ossia fino a quando il fondamento dell'induzione è stato considerato il principio dell'uniformità delle leggi di natura e tale principio non era altro che lo stesso principio di causalità.

La soluzione soggettivistica, prospettata dallo stesso Kant come risposta al dubbio di Hume sulla possibilità della generalizzazione scientifica, ha costituito una dottrina simmetrica e opposta a quella dell'uniformità naturale, ma il suo significato è lo stesso.

La giustificazione pragmatica è stata avanzata, nella filosofia contemporanea, quando si è riconosciuta la impossibilità di una giustificazione teorica, ma non si è giunti a negare la legittimità del problema cioè della richiesta di una giustificazione.

La giustificazione è stata, in questa direzione, cercata mediante un'interpretazione *probabilistica dell'induzione* e si è pervenuti alle due conclusioni seguenti: 1) l'induzione è il solo mezzo per ottenere previsioni; 2) essa è il solo metodo di autocorrezione.

Tuttavia, allo stato attuale del problema, bisogna riconoscere che una giustificazione dell'induzione non solo è impossibile, ma che il problema di essa è privo di senso, se per giustificazione s'intende la dimostrazione della validità infallibile del procedimento induttivo.

La giustificazione, infatti, se fosse raggiunta, avrebbe per suo effetto immediato la eliminazione di ogni rischio dei procedimenti induttivi e la riduzione di questi procedimenti alla certezza.

In realtà, i procedimenti scientifici e, in generale, i comportamenti e le direttive *razionali* dell'uomo, consistono nel limitare il rischio cioè nel renderlo calcolabile non potendolo eliminare.

Nel linguaggio corrente — e non di rado anche in quello scientifico — il concetto di induzione è considerato equivalente a quello di inferenza, ma, anche se talvolta possono coincidere — o, comunque essere assunti equivalenti — in generale i due concetti differiscono qualora si attribuisca al termine inferenza il suo più ampio significato.

Invero, sull'induzione vi è quasi l'unanimità nel considerarla come un procedimento logico, opposto a quello di deduzione, per cui dall'osservazione di casi particolari si sale ad affermazioni generali o, talvolta, dalla conoscenza dei fatti si risale alla conoscenza delle leggi che li regolano, mentre per non pochi filosofi il termine inferenza, equivalente ad *illazione*, dovrebbe essere usato per designare ogni forma di *deduzione*, con cui si dimostri il logico conseguire di una verità ad un'altra.

A noi sembra, comunque, che il concetto di inferenza, oltre che coincidere, in certi casi, con quello di induzione, assuma un ben più ampio significato se in esso si comprende anche quello di deduzione, vale a dire se per inferenza, in senso lato, si intende un processo mentale operativo per cui si passa, da certi dati o giudizi, a nuovi giudizi od a conclusioni che derivano dai primi. Pertanto, nella sua più ampia accezione, *l'inferenza deve essere intesa come una qualsivoglia forma di argomentazione che riguardi un processo di induzione oppure un procedimento di deduzione*; nel primo caso si può parlare di *inferenza induttiva*, nel secondo di *inferenza deduttiva*.

Per lo più, tuttavia, nella ricerca scientifica ed operativa, quando si usa il solo termine inferenza, si vuole intendere l'inferenza induttiva, la quale, non di rado, viene specificata come *inferenza statistica* o *inferenza probabilistica*.

Comunque, in generale, l'inferenza statistica può essere anche deduttiva, purché si tratti in ogni caso di un processo, semplice o composto, di induzione o di deduzione.

ne, con cui si dimostra la conseguenza logica di un'espressione formata da dati statistici ad un'altra costituita dal risultato dell'analisi o dell'elaborazione dei dati, oppure la conseguenza logica di un modello teorico che conduca alla specificazione di chiare proprietà ed espressioni statistiche.

Dal nostro punto di vista, pertanto, l'inferenza statistica non può significare *soltanto* un procedimento di generalizzazione dei risultati ottenuti attraverso una rilevazione parziale per *campioni* — limitata cioè alla considerazione di alcune unità o casi singoli di un fenomeno — alla totalità delle unità o dei casi del fatto o fenomeno considerato, *sulla base di determinate ipotesi plausibili*.

È questo, senza dubbio, un procedimento di inferenza statistica che riguarda più propriamente le grandi indagini campionarie su fenomeni economici e sociali, e che si propone di risalire da talune proprietà accertate in campioni reali, tratti da popolazioni finite, alle corrispondenti proprietà delle popolazioni stesse.

E lo è anche quando, sulla base di campioni ottenuti da piani razionali di sperimentazione, si vogliono saggiare gli effetti di certi trattamenti fisici oppure chimici o biologici, ed anche allorquando, mediante di campioni di prodotti industriali, si vuole controllare la qualità della produzione ecc.

Ma soltanto quando l'inferenza statistica riguarda campioni probabilistici — reali o sperimentali — ed il processo induttivo è prevalentemente probabilistico si potrà parlare di inferenza induttiva probabilistica.

Tuttavia, l'inferenza statistica, in senso lato, comprende ben altri problemi che possono riguardare taluni fattori che influiscono su determinati caratteri od avvenimenti, oppure le relazioni che intercorrono tra due o più fenomeni, e persino veri e propri problemi di previsione.

Per altro, nel suo ampio significato deduttivo, l'inferenza probabilistica comprende tutto il calcolo delle probabilità e tutti quei modelli probabilistici che insieme ai modelli deterministici costituiscono i fondamenti dell'attuale teoria statistica.

Ma, come abbiamo premesso, l'inferenza induttiva statistica non è sempre inferenza probabilistica, per altro, non sempre l'inferenza probabilistica, per quanto corretta, può condurre alla formulazione di vere e proprie leggi o relazioni scientifiche.

Così si può pervenire sovente all'accertamento di *leggi statistiche*, mediante inferenze induttive — prescindendo da qualsivoglia concezione probabilistica — sulla base di dati osservati, che sono qualcosa di più di semplici relazioni empiriche valide soltanto entro limiti di tempo e di spazio considerati.

Si tratta di leggi che riguardano *domini* più ampi di quelli osservati pur non presentando i presupposti necessari per essere considerati quale leggi scientifiche di carattere generale.

Ma anche i più affinati procedimenti, deduttivi e induttivi, di inferenza probabilistica, applicati a dati statistici ottenuti mediante l'osservazione spontanea o sperimentale, possono condurre a vere e proprie relazioni scientifiche soltanto se le relazioni

sono convalidate da necessarie e sufficienti informazioni supplementari non contenute nei dati osservati.

È un fatto, comunque, che oggi l'inferenza scientifica, nelle sue varie metodologie induttive, si avvale di criteri e procedimenti che sono ben diversi da quelli, ad esempio, a suo tempo proposti da Stuart Mill.

Invero, i quattro metodi, cosiddetti di induzione sperimentale, definiti, rispettivamente, *delle differenze, delle concordanze, dei residui e delle variazioni concomitanti*, sono stati quasi completamente abbandonati — anche nell'ambito delle scienze sociali — quando si sono avvertiti gli inevitabili inconvenienti ai quali potevano condurre e soprattutto quando è apparso ben chiaro che tali metodi non consentivano, in alcun modo, di valutare il *rischio* della fallibilità dei procedimenti adottati o di offrire un qualche *grado* o livello di attendibilità o plausibilità dei risultati.

Non vi è dubbio, pertanto, che ogni qualsivoglia valutazione del *rischio di un giudizio errato* o di un *grado di attendibilità del giudizio* — sui risultati di una ricerca scientifica od operativa — deve essere espressa da una *quantità* calcolata sui dati ottenuti dall'osservazione spontanea o sperimentale — mediante un ben definito criterio fondato su precise assunzioni teoriche.

Quando tale quantità è connessa, in qualche modo, ad una particolare concezione della probabilità, si potrà parlare di inferenza probabilistica.

Comunque, in generale, ogni tipo di inferenza statistica deve essere inteso, nel suo più ampio significato, quale insieme dei due processi — induttivo e deduttivo — che possono essere chiariti qualora si distinguano i tre *momenti* dell'inferenza, che corrispondono, rispettivamente, ai tre aspetti della *statistica teorica, metodologica ed applicata* e si riguardi la metodologia nella sua ineliminabile essenza metodologica condizionata.

Bisogna riconoscere che, in effetti, le vere e proprie teorie dell'inferenza sono soltanto insiemi di ipotesi e nozioni astratte, sono cioè proposizioni assiomatiche dalle quali, per via puramente logica — sulla base di particolari assunzioni — si passa ad altre proposizioni.

Appare quindi lecito, da un certo angolo visuale, riguardare le teorie dell'inferenza statistica, nelle loro formulazioni, quali *processi prevalentemente deduttivi* in quanto, partendo da premesse ipotetiche, pervengono — secondo norme logiche —, a conclusioni astratte.

È nel momento *metodologico* che tali inferenze, dovendo riferirsi a ben definiti problemi attinenti a determinati dati statistici, per risalire a conclusioni di carattere particolare o più generale, diventano induttive. Non si può ignorare che, come abbiamo ricordato, ogni metodo discende da una o più teorie ed infatti il termine *metodo* — derivato dal greco — *in generale vuole significare il modo con cui si realizza una iniziativa partendo da determinati principi, stabiliti a priori, e procedendo logicamente.*

La teoria, dunque, stabilisce deduttivamente i principi dell'inferenza, mentre la metodologia si propone di realizzare concretamente il processo di induzione.

Oggi, quindi, le ricerche scientifiche ed operative si avvalgono di innumerevoli differenti metodi di inferenza statistica che scaturiscono dalle varie teorie e che, a seconda dei problemi e delle fasi delle ricerche, possono essere considerati compatibili o incompatibili fra loro. Come è possibile chiarire, si tratta di metodi talvolta complementari ma comunque rivolti, in generale, a risolvere problemi diversi.

Se si esaminano attentamente le proprietà, i criteri e i procedimenti che costituiscono i diversi *metodi* di inferenza, è facile accertare che si tratta di *metodi condizionati*, qualora si attribuisca al termine *condizione* il significato che gli è proprio nella concezione filosofica — ossia campo di possibilità al di fuori del quale non si ha l'oggetto condizionato — oppure nel linguaggio scientifico, *quale limitazione imposta a enti o grandezze perché si verifichi un dato evento o abbia senso un determinato risultato*.

Il momento applicativo dell'inferenza statistica si inizia con la *decisione* riguardante la scelta del metodo da impiegare per risolvere un determinato problema concreto.

I problemi di inferenza statistica — com'è noto — sono illimitati, ma praticamente, in gran parte, vengono ancora raggruppati — benché in modo discutibile — in due grandi categorie: *problemi di stime di parametri incogniti e problemi di verifica, o prova, di ipotesi*.

Ciascuna di queste categorie comprende un vasto campo di problemi di assai varia natura, comunque il momento applicativo dell'inferenza statistica prosegue operativamente con l'impiego del metodo scelto e si conclude con la *decisione* definitiva attinente ai valori stimati o alle ipotesi saggiate.

E poiché ogni teoria assume, in astratto, determinate ipotesi e condizioni, prescindendo da dati statistici ottenuti da osservazioni, empiriche o sperimentali, *tutte le teorie — non viziate da errori di logica — debbono essere considerate valide nel contesto delle ipotesi e condizioni assunte*.

Non ha senso, pertanto, quando si confrontano teorie fondate su differenti assunzioni, concludere che una di esse debba essere considerata, in ogni caso, la migliore o comunque preferibile alle altre.

Certamente alcune teorie presentano fra loro un contrasto insanabile, ma, nella grande maggioranza, le teorie dell'inferenza possono conciliarsi o, comunque, coesistere.

Soltanto se si può dimostrare che le ipotesi assunte da una teoria sono assurde o che le condizioni poste sono irrealistiche, la teoria cade in difetto e non ha senso porla a confronto con altre teorie razionali convalidabili dai fatti.

In realtà è nella fase dell'*applicazione* che molti metodi si dimostrano incompatibili fra loro, poiché in tale fase diventano *praticamente e razionalmente validi, di volta in volta, soltanto quei metodi che soddisfano le particolari esigenze e condizioni dei pro-*

blemi concreti che sono chiamati a risolvere, mentre tutti gli altri metodi — pur essendo perfettamente validi entro i limiti delle assunzioni da cui derivano — divengono incompatibili con i primi.

2. Per addivenire alla scoperta di relazioni od all'approntamento di previsioni come, più in generale, per qualsiasi attendibile processo di inferenza scientifica, la statistica non si limita a trarre informazioni dai dati osservati e dalle loro elaborazioni, bensì si avvale inevitabilmente anche di *informazioni supplementari od ausiliarie*, qualitative e quantitative, che non sono contenute nelle osservazioni o ricavate da esse.

Queste informazioni sono supplementari nella più ampia accezione del termine, ossia quando si considera supplementare tutto ciò che serve, o può servire, di supplemento inteso quest'ultimo come aggiunta fatta a qualche cosa — in particolare, ad altra informazione — che si rivela insufficiente per determinate finalità.

Ovviamente le informazioni supplementari, od ausiliarie, sono di vari tipi e possono essere classificate in diversi modi.

Così, ad esempio, una classificazione, fra le tante, semplice ma fondamentale, è quella che distingue tali informazioni, rispettivamente, nelle tre seguenti categorie.

Una prima categoria è costituita da informazioni fornite da altri dati o parametri statistici ricavati da precedenti rilevazioni o indagini oppure da particolari ricerche sperimentali o relazioni comunque accertate. Queste informazioni possono essere definite *oggettive*.

Un'altra categoria è formata da ipotesi, opinioni, credenze e nozioni teoriche assunte dallo statistico in vari modi e che possono risultare anche dalla combinazione di opinioni e credenze di due o più persone. Queste informazioni sono evidentemente *soggettive*.

Non di rado, tuttavia, nella ricerca scientifica ed operativa si presenta l'opportunità di combinare necessarie informazioni supplementari oggettive con informazioni soggettive per ottenere informazioni che possono essere chiamate *miste*.

Spesso le informazioni supplementari vanno sotto il nome di informazioni a priori, o iniziali, ma in realtà queste espressioni, a rigore, non sono sempre le più appropriate in quanto, pur risultando, per lo più, informazioni precedenti a quelle costituite dai dati osservati di un campione o di una popolazione, esse, in realtà, vengono quasi sempre considerate soltanto dopo che, stabilite le finalità delle ricerche, sono state prese in esame le osservazioni.

Riteniamo, comunque, che il termine *supplementare*, oltre che più appropriato di quello *a priori*, sia anche di più ampio significato poiché, talvolta, le informazioni ausiliarie vengono ricavate mediante nuove rilevazioni od indagini, durante qualche fase del processo induttivo-deduttivo della ricerca e, quindi, non sono, ovviamente, note a priori.

Pertanto, se si considera che l'inferenza scientifica, in generale, necessita di credenze a priori, mentre sulla base dei soli dati osservati non è possibile accertare la ve-

rità di un'ipotesi, è facile rendersi conto che le *inferenze statistiche*, di qualsivoglia natura, non possono prescindere da informazioni supplementari.

Anche le più semplici, ma razionali, elaborazioni dei dati osservati presuppongono assunzioni a priori, implicite od esplicite, nei riguardi, ad esempio, della casualità o dell'indipendenza dei dati oppure dei modelli di talune distribuzioni di frequenze o di distribuzioni di fondamentali parametri incogniti.

Eppure vi è ancora qualcuno che ritiene, erroneamente, che la necessità di espressioni di credenze a priori non sia una necessità di tutta l'inferenza scientifica, ma piuttosto soltanto un elemento della inferenza Bayesiana.

Tale convinzione deriva, fra l'altro, dall'errata idea che l'assunzione di un particolare modello non equivalga ad una credenza a priori, ma il concetto che assimila l'assunzione di un modello, in qualsivoglia tipo di ricerca, ad una distribuzione a priori dell'approccio Bayesiano, diventa chiara allorquando si consideri che ogni modello può essere immaginato come un caso particolare di un modello più complesso o di una famiglia di modelli.

Così, ad esempio, l'assunzione di una curva normale può considerarsi derivante da una famiglia di curve normali di ordine r definita dal parametro r per il quale si assume a priori il valore 2.

Ma, più in generale, ancorché si prescinda da inferenze Bayesiane, informazioni supplementari possono essere usate per stimare — con metodi adeguati, sotto determinate condizioni — i parametri incogniti di modelli statistici.

Invero, le informazioni a priori — nel loro più ampio significato — non implicano necessariamente soltanto approcci Bayesiani, ma riguardano, ad esempio, anche equazioni e disequaglianze quali stime supplementari di parametri incogniti.

Così grande importanza assumono spesso le informazioni supplementari per quanto attiene, in particolare, la robustezza di stimatori minimax, il metodo dei minimi quadrati ristretti, gli stimatori di minimi quadrati a due stadi e alcuni tipi di stimatori, detti misti, che consentono anche che l'informazione a priori sia in parte distorta.

È appena il caso di ricordare poi che, indipendentemente da approcci Bayesiani, informazioni supplementari a priori vengono da tempo utilizzate nelle indagini campionarie probabilistiche. È noto, infatti, che la conoscenza di una variabile altamente correlata con quella riguardata da una rilevazione campionaria può consentire di migliorare la probabilità di scelta delle unità di campionamento e la costruzione di strati della popolazione sino a ridurre lo stimatore della varianza di campionamento dei parametri incogniti.

Questo concetto, naturalmente, deve essere distinto da quello che attiene all'uso di informazioni supplementari, quali variabili strumentali, per conseguire l'identificabilità nella stima di equazioni strutturali; è, tuttavia, analogo, invece, all'uso di una variabile concomitante nell'analisi della covarianza, in quanto consente di ridurre la variabilità dei residui.

In generale poi, come abbiamo accennato, i modelli statistici, sia di distribuzioni di parametri incogniti sia di distribuzioni statistiche di frequenza, sono rappresentazioni formali di credenze che costituiscono importanti informazioni supplementari comunemente definite *a priori*.

Giova avvertire, tuttavia, che le distribuzioni a priori assunte, ad esempio, dall'inferenza Bayesiana talvolta sono i risultati di confronti fra vari esperti, rappresentano cioè cosiddette *distribuzioni del consenso* scaturite dalle combinazioni delle opinioni di due o più persone.

Queste ultime, dunque, conseguite dopo la *contrattazione*, possono anche essere riguardate, per un certo aspetto, come distribuzioni a posteriori rispetto alle opinioni iniziali dei singoli esperti.

Non di rado, in effetti, le stesse informazioni supplementari possono essere considerate a priori od a posteriori a seconda dei punti di vista dai quali vengono riguardate.

Pertanto, è opportuno che le distribuzioni assunte a priori, nell'inferenza Bayesiana, vengano invece chiamate distribuzioni iniziali — come suggerisce, ad esempio, il De Finetti — e, analogamente, le distribuzioni che, nella predetta inferenza, vengono dette a posteriori siano chiamate distribuzioni finali. Ciò perchè prima di assumere una distribuzione definitivamente iniziale, in generale, deve essere compiuta una analisi — che possiamo dire prefinale — su vari tipi di distribuzioni a priori, esaminando i possibili risultati conseguibili di volta in volta, con il teorema di Bayes, sulla base dei dati osservati e del comportamento di tali risultati alla luce di convenienti informazioni supplementari, oggettive o miste.

Queste informazioni contribuiscono dunque alla scelta di una distribuzione iniziale di un approccio Bayesiano, ma, più in generale, informazioni supplementari, non soltanto a priori, sono necessarie per una razionale costruzione di modelli statistici, mentre informazioni a posteriori occorrono poi sempre per un'esatta interpretazione del concreto significato di qualsivoglia ricerca statistica che non si limiti a pure descrizioni formali.

A proposito delle ricerche statistiche, in generale, giova forse tener presente che nella pratica quotidiana si suole fare una netta distinzione fra ricerche condotte su campioni statistici tratti da popolazioni, reali o teoriche, finite o infinite, e ricerche attinenti ad intere popolazioni.

Tuttavia l'esigenza di informazioni supplementari è sempre evidente in entrambi i tipi di ricerche.

Così, ad esempio, riguardo alla conformità dei dati osservati — di un campione o di una popolazione — ad un modello teorico, è ovvio che la pura e semplice applicazione di qualche test può essere, in molti casi, necessaria ma, in generale, non è mai sufficiente per assicurare la reale validità del modello poiché, in base ai soli dati osservati non è mai possibile — come abbiamo detto — accertare la validità di un'ipotesi.

Per confortare il risultato di un test occorrono informazioni supplementari, qua-

litative e quantitative che dovrebbero essere di conoscenza del ricercatore. E se le più *giudiziose* informazioni del ricercatore sono necessarie e sufficienti per accertare o respingere decisamente la validità di un modello, l'impiego del test è del tutto inutile.

Non vi è dubbio che per le inferenze probabilistiche, di qualsiasi tipo, ricavate da campioni, si presenta la necessità di appropriate informazioni supplementari che non servono per le ricerche svolte su popolazioni.

Ma, com'è noto, la moderna analisi dei dati, comunque, si propone, fra l'altro, di estrarre dalle osservazioni di fenomeni collettivi — intese come popolazioni statistiche e non come campioni — talune strutture, distinte, dal punto di vista matematico, rispettivamente, in strutture *algebriche*, *d'ordine* e *topologiche*.

Se non che, nelle applicazioni concrete, le strutture da matematiche diventano più propriamente statistiche per cui debbono assumere ben definiti particolari significati e soddisfare determinate condizioni.

Stando così le cose, le strutture individuate, pur essendo tipiche *espressioni di statistica matematica*, debbono essere interpretate nelle loro accezioni fenomeniche secondo la natura e le caratteristiche dei dati osservati.

Quindi, per le interpretazioni delle strutture accertate statisticamente, i dati osservati non sono sufficienti, bensì sono necessarie informazioni supplementari, come occorrono per una valida interpretazione del significato concreto di qualsivoglia risultato di altre ricerche statistiche.

Un importante campo di ricerche, nel quale le informazioni supplementari sono assolutamente indispensabili, è quello che concerne le previsioni di fenomeni socio-economici.

A questo riguardo, anzi, alcuni autori fanno una distinzione, per quanto discutibile, fra *metodi* di previsione puramente statistici e *sistemi* di previsione.

I metodi statistici di previsione sarebbero quelli basati essenzialmente sull'analisi di serie temporali e su informazioni del passato e che trasformano i dati osservati in informazioni del futuro in modo prevalentemente meccanico.

I sistemi di previsione, invece, comprenderebbero tutte le informazioni raccolte, per vari aspetti, dalle persone che progettano i sistemi e da quelle che sono interessate alle previsioni e che forniscono le informazioni attinenti alle decisioni operative.

È abbastanza naturale isolarsi ad analizzare una serie storica di osservazioni per costruire ed applicare una teoria elaborando particolari metodi condizionati di previsione, ma assumere che, in realtà, un metodo che determina «il miglior adattamento» ad una serie di dati del passato sia il miglior metodo da impiegare in un sistema di previsioni socio-economiche è assai pericoloso, quasi sempre ingiustificabile e spesso assurdo.

In generale, i dati delle serie storiche sono talmente autocorrelati che si presenta la tendenza al superadattamento di certe realizzazioni del passato a scapito del futuro, non ancora realizzato, e la semplice estrapolazione di modelli, per quanto razionali ed

affinati, ignora informazioni esterne ai dati, ossia informazioni supplementari le quali, più che al passato, debbono riguardare il presente e il futuro.

I metodi di previsione fondati soltanto su analisi, semplici o complesse, di serie storiche sono dunque, ovviamente, inadeguati e pertanto debbono essere associati ad informazioni che consentano di correggere gli andamenti previsti alla luce dei soli dati del passato.

In realtà i metodi basati sulle sole analisi di serie storiche trascurano completamente l'importanza fondamentale che assumono, in pratica, gli individui che in un concreto sistema di previsione *comunicano fra loro e con il programmato metodo di previsione*.

Gli individui, in effetti, possono avere fonti di informazioni che vanno molto al di là dei puri dati storici. Così, ad esempio, essi potranno sapere se un'impresa concorrente è venuta a mancare, se è stato introdotto nel mercato un nuovo prodotto competitivo, se taluni provvedimenti di politica economica o sociale diventeranno operativi dopo un certo periodo di tempo, e così via.

Gli effetti di questi eventi sono, senza dubbio, molto difficili da valutare esattamente e per descriverne l'incertezza si assumono, talvolta, distribuzioni di probabilità.

Comunque, in queste ed in molte altre circostanze non si potrà certamente aspettare che un metodo statistico consenta di prevedere tali avvenimenti usando soltanto dati storici. È necessario piuttosto che gli individui possano comunicare le loro informazioni all'elaboratore, che ha in programma il metodo, e che questo a sua volta comunichi l'informazione di incertezza in modo tale che sia prontamente accolta e interpretata da chi deve prendere le decisioni.

Le varie informazioni comunicate all'elaboratore possono derivare, ovviamente, da diverse fonti e l'aggregazione di tali informazioni, com'è noto, va sotto il nome di *formazione di un consenso di esperti*, anche se, in realtà, alcuni dei cosiddetti esperti rappresentano previsioni da equazioni di regressione oppure da altri dati disponibili o da modelli teorici o, comunque, da fonti non soggettive.

L'incertezza può essere rappresentata formalmente in termini di probabilità e lo scopo ultimo di un procedimento del consenso è pervenire a distribuzioni di probabilità delle quantità incerte che interessano. Queste distribuzioni che, come abbiamo ricordato, sono chiamate *distribuzioni del consenso*, devono riflettere chiaramente le informazioni, nel loro insieme, fornite dagli esperti.

Per l'aggregazione di distribuzioni di probabilità soggettive di esperti sono stati suggeriti vari metodi che implicano retroazioni e, o, discussioni di gruppo, e che vanno da medie semplici o ponderate, delle singole distribuzioni, all'impiego di distribuzioni a priori coniugate per un formale procedimento Bayesiano.

Sono state fatte ricerche anche per mettere in luce l'influenza delle relazioni fra gli errori di stima dei diversi esperti sulle distribuzioni del consenso.

Queste distribuzioni poi, com'è noto, possono essere caratterizzate in vari modi,

ossia mediante tutti i loro parametri, oppure con uno o più momenti, per descriverne le caratteristiche essenziali — quali valori medi, indici di variabilità e di forma — od anche soltanto l'andamento crescente o decrescente in un certo intervallo, e così via.

Emerge dunque, con sufficiente chiarezza come le informazioni che abbiamo definite supplementari — indispensabili in tutte le ricerche scientifiche ed operative — abbraccino concetti tanto diversi che vanno dalle informazioni oggettive — ossia dati ottenuti da osservazioni dirette o da esperimenti od elaborazioni — alle informazioni soggettive rappresentate dall'assunzione di modelli teorici — in particolare, distribuzioni di parametri incogniti e di frequenze — e dalla formulazione di giudizi e di opinioni.

A chiusura di queste pagine giova aggiungere che problemi di impiego di informazioni supplementari si possono presentare anche in ricerche empiriche o semplicemente per finalità descrittive.

Basterà considerare, quali esempi, alcuni problemi chiamati dal Gini di *interpolazione per induzione*, ossia problemi di valutazione di modalità mancanti.

Così, ad esempio: «in una serie dei prezzi annuali del pane in una data città che presenti delle lacune, in corrispondenza a certi anni, i dati mancanti possono essere interpolati basandosi sui prezzi che il pane ha mostrato negli anni corrispondenti in una città vicina oppure sui prezzi che nella stessa città ha mostrato la farina».

È facile intuire, tuttavia, innumerevoli casi concreti nei quali l'interpolazione per induzione, alla luce di informazioni supplementari, può essere effettuata nell'ambito di fenomeni economici e sociali, con vari criteri, senza modificare i dati osservati purché sia possibile applicare statisticamente criteri di analogia che si risolvono, in gran parte, in termini formali, in successioni di proporzioni numeriche.

Ma l'odierna metodologia statistica affronta problemi di valori mancanti nei riguardi anche di piani sperimentali e di grandi indagini campionarie con procedimenti che si avvalgono di altri tipi di informazioni supplementari e che, non di rado, utilizzano, in qualche modo, anche valori osservati.

È dunque sconfinato il campo delle ricerche — empiriche, scientifiche ed operative — nel quale le indagini statistiche non possono prescindere da informazioni supplementari e fra queste informazioni emergono quasi sempre particolari assunzioni di ipotesi.

In generale, infatti, un'ipotesi è un concetto assunto quale punto di partenza per un'ulteriore elaborazione e che va poi comprovato dai risultati ottenuti attraverso il successivo svolgimento del problema, dell'esperimento o del ragionamento.

In ogni caso, un'ipotesi, pur nella sua costitutiva incertezza, non può — affinché il suo uso sia legittimo — sottrarsi a un determinato insieme di *informazioni iniziali*, ossia di rapporti e di relazioni riguardanti ciò che è già noto in un certo campo scientifico in cui essa viene usata, quale elemento essenziale per la sua verificabilità e per la sua effettiva portata scientifica.

Non è facile, pertanto, proporre una classificazione, che non sia arbitraria, ma comprensiva, dei principali campi della ricerca statistica nei quali l'analisi concreta e teorica, si avvale di informazioni supplementari, tuttavia riteniamo che, fra le tante classificazioni convenzionali possa apparire non inutile una distinzione fra alcune fondamentali categorie di ricerche pur non essendo possibile stabilire fra di esse una rigorosa separazione.

Invero, pur con le necessarie riserve, le principali categorie di ricerche statistiche che, in generale, si avvalgono di indispensabili informazioni supplementari possono essere così classificate:

- 1) ricerche empiriche nell'interpolazione per induzione e, in generale, per la stima di valori mancanti;
- 2) indagini campionarie probabilistiche e analisi della covarianza;
- 3) per svolgere complete inferenze Bayesiane;
- 4) per migliorare le stime dei parametri di modelli lineari indipendentemente da ogni approccio Bayesiano;
- 5) con l'impiego di variabili cosiddette strumentali;
- 6) per l'assunzione di modelli di distribuzioni di parametri incogniti;
- 7) per l'assunzione di modelli di distribuzioni di frequenza;
- 8) in generale, per la costruzione di modelli statistici;
- 9) per ottenere le cosiddette distribuzioni del consenso;
- 10) per la costruzione e la realizzazione di razionali sistemi di previsione socio-economica;
- 11) per una razionale interpretazione concreta dei risultati di qualsivoglia indagine statistica, scientifica od operativa.

SILVIO VIANELLI