

L'INFERENZA STATISTICA SECONDO BAYES, FISHER E NEYMAN-PEARSON

Nota del m.e. VITTORIO B. FROSINI (*)

(Adunanza del 3 ottobre 2019)

SUNTO. – Nella letteratura statistica esistono diverse proposte di procedura inferenziali, che sulla base di un campione casuale portano a inferire le caratteristiche di una popolazione, o variabile casuale; le procedure più comuni sono quelle esposte in questa nota. L'inferenza bayesiana è basata su una proposta di Thomas Bayes del 1763: a partire da una distribuzione di probabilità per un parametro della variabile casuale – da cui il campione è estratto – tale distribuzione è aggiornata sulla base delle informazioni campionarie. La proposta di Ronald Fisher utilizza solo le informazioni campionarie: una data ipotesi sulla popolazione è respinta solo se i dati disponibili non sono coerenti con tale ipotesi. La proposta di Jerzy Neyman e Egon Pearson richiede che venga precisata un'ipotesi alternativa, per cui l'inferenza eseguita sulla base del campione permette di scegliere fra le due ipotesi a confronto.

ABSTRACT. – In the literature about procedures of statistical inference, which – on the basis of a random sample – allow to accept a precise hypothesis concerning the probability distribution which yields the sample, there are many proposals. One of the oldest is the one by T. Bayes (1763): starting from a given distribution for the parameter of a random variable, this same distribution is revised conditionally on the sample results. Instead, R. Fisher's proposal is based only on the sample information: a given hypothesis is rejected only if the sample results are scarcely coherent with such hypothesis. The third proposal, made by J. Neyman and E. Pearson, takes into explicit consideration also an alternative hypothesis, and leads to the acceptance of the hypothesis more akin to the sample results.

(*) Università Cattolica del Sacro Cuore, Milano, Dipartimento di Scienze statistiche, Milano, Italy. E-mail: benito.frosini@unicatt.it

1. INTRODUZIONE

Come è noto, l'inferenza classica, o inferenza *tout court*, è sostanzialmente quella basata sui sillogismi aristotelici (di cui si farà un breve richiamo nel seguito al § 3); in un sistema deduttivo, è l'insieme di regole mediante le quali, da una o più affermazioni assunte come vere – gli *assiomi* – vengono derivate altre affermazioni – i *teoremi* – che sono altrettanto vere. L'*inferenza statistica* differisce sostanzialmente dall'inferenza classica per il fatto che una o più affermazioni sono assunte come vere con un certo grado di probabilità; pertanto le conclusioni di un ragionamento inferenziale non potranno più essere certe; in generale, tali conclusioni saranno accettate in via provvisoria, fintanto che nuove informazioni non producano un loro cambiamento.

Di regola l'inferenza statistica ha come oggetto la *stima* di una o più caratteristiche di una popolazione (effettiva o teorica), o la valutazione di una *ipotesi* concernente la distribuzione di probabilità di una popolazione o variabile casuale, eseguite sulla base di un campione casuale proveniente dalla stessa (popolazione o variabile casuale). L'inferenza statistica eseguita può anche essere errata, e questo può accadere più facilmente se la dimensione del campione è piuttosto piccola (ad es. ≤ 20); potremmo essere stati sfortunati, con l'estrazione di un campione poco rappresentativo della popolazione.

Nel trattamento dell'inferenza statistica si preferisce generalmente riferirsi a una popolazione come a una variabile casuale (spesso scritta con le sole iniziali v.c.), che può assumere tutti i valori (per un fenomeno quantitativo unidimensionale) in un certo intervallo di numeri reali, o tutti i valori in un insieme finito (ad es. tutti i numeri interi compresi fra 0 e n , dove n è la dimensione del campione), o anche tutti i valori in un insieme infinito discreto, con date probabilità associate agli stessi valori o a intervalli di valori assumibili. Nel caso di variabili casuali riferite a fenomeni continui (o che conviene trattare matematicamente come tali), ai singoli valori della relativa v.c. non può essere associata una probabilità; in tal caso si preferisce parlare di *densità* (di probabilità). Fra le ipotesi (statistiche) concernenti tali v.c. (che sono modelli di fenomeni reali), si usano distinguere: (a) ipotesi *semplici*, quando caratterizzano completamente la distribuzione di probabilità della v.c. in oggetto, e (b) ipotesi *composte*, quando sono compatibili (cioè contengono) con più ipotesi semplici (ad es. quando si suppone che la media di una v.c. continua sia compresa fra 0,4 e 0,6).

Di regola le v.c. (variabili casuali) più utilizzate dipendono da un parametro, a una o più dimensioni. Ad esempio, se il campionamento fornisce la presenza o l'assenza su ogni individuo di una data caratteristica A , con $P(A) = p$ (che può assumere tutti i valori fra 0 e 1), la v.c. X considerata è spesso una v.c. *binomiale*, che descrive il numero x di A nel campione:

$$P(X = x ; p) = C_{n,x} p^x (1-p)^{n-x} \text{ per } x = 0, 1, \dots, n ; 0 < p < 1$$

dove $C_{n,x}$ è il c.d. coefficiente binomiale (numero di combinazioni di n oggetti presi x alla volta).

Altro esempio: se la v.c. X è continua, e può assumere (almeno nel modello) tutti i valori reali, spesso si usa la v.c. *normale*, con funzione di densità:

$$f(x ; \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp \left\{ -\frac{(x - \mu)^2}{2\sigma^2} \right\}$$

dove $-\infty < x < \infty ; -\infty < \mu < \infty ; \sigma > 0$.

La v.c. normale dipende da due parametri, μ (parametro di localizzazione, coincidente con la media, la moda e la mediana di X), e da σ (parametro che descrive la dispersione intorno alla media).

Il parametro (a una o più dimensioni) da cui dipende la distribuzione di probabilità della v.c. X è in genere indicato con θ . Ad esempio, $\theta = p$ nel caso della variabile binomiale, $\theta = (\mu, \sigma)$ nel caso della normale.

Nella letteratura statistica sono presenti diverse proposte di procedure inferenziali; per una rassegna critica generale si può consultare il pregevole volume di Vic Barnett (1982). Fra queste procedure, di gran lunga le più note e concretamente utilizzate sono quelle che verranno presentate succintamente nei tre paragrafi successivi, e che sono citate nel titolo di questa nota.

2. L'INFERENZA BAYESIANA

Si avverte che nella esposizione dell'inferenza bayesiana, come pure per i tipi di inferenza trattati nelle sezioni 3 e 4, non saranno considerati i casi più generali, ma solo le applicazioni particolari che verranno presentate; questo perché l'autore di questa nota ritiene essen-

le la comprensione della logica inferenziale esposta, che è possibile apprezzare anche (o forse soprattutto) nelle applicazioni più semplici.

Sembra conveniente introdurre la formula fondamentale dell'inferenza bayesiana, che è un modo di scrivere la formula della probabilità di un *evento condizionato*, mediante un semplice esempio. Si consideri allora l'estrazione di una carta da un mazzo normale di 40 carte, consistente di 10 carte per ciascuno dei semi tradizionali (cuori, quadri, fiori, picche); tale estrazione casuale può ricevere 40 determinazioni, che costituiscono lo *spazio campionario* di questo esperimento casuale.

Siano A e B due eventi casuali, aventi rispettiva probabilità $P(A)$ e $P(B)$. Ad esempio, per l'esperimento casuale in discorso sia A = estrazione di un asso con $P(A) = 4/40 = 1/10$, e B = estrazione di una carta di fiori con $P(B) = 10/40 = 1/4$. Si indichi con $A|B$, e si legga "A dato B" l'*evento condizionato* consistente nell'estrazione di un asso, ammesso che sia stata estratta una carta di fiori; la *probabilità condizionale* $P(A|B)$ potrebbe scriversi direttamente (data la semplicità dell'argomento) come $1/10$ (dato che vi è un solo asso fra le dieci carte di fiori); però conviene mostrare, per le sue implicazioni, che a tale probabilità si può arrivare mediante la seguente formula:

$$P(A|B) = \frac{P(A)P(B|A)}{P(B)} = \frac{\frac{1}{10} \cdot \frac{1}{4}}{\frac{1}{4}} = \frac{1}{10}$$

Una applicazione della probabilità condizionale è la formula di Bayes (1673), che scriviamo con riferimento a possibili *cause* A_1, A_2, \dots e a un *effetto* H :

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

dove $P(H) = \sum_r P(A_r) P(H|A_r)$. Questa formula permette di passare dalla conoscenza di probabilità condizionali del tipo $P(H|A_r)$ (le c.d. *verosimiglianze*, in genere abbastanza facili da calcolare), a probabilità condizionali del tipo $P(A_r|H)$, riferite alle possibili cause dell'evento H .

Nelle applicazioni statistiche della procedura inferenziale *bayesiana*, di regola il riferimento *formale* è a un campionamento a due stadi: nel primo stadio si sceglie un valore casuale per il parametro θ ; nel secondo stadio si sceglie un valore $X = x$, secondo una distribuzione di probabilità indicizzata dal valore prescelto θ . Si suppone pertanto che la v.c. X si distribuisca secondo una legge di probabilità $g(x | \theta)$. La pre-

cedente formula, già riferita alla probabilità di una causa A_r per un effetto H , può quindi risciversi come segue:

$$P(\theta | x) = \frac{P(\theta) P(x|\theta)}{P(x)}$$

La distribuzione $P(\theta)$ del parametro θ è chiamata *distribuzione a priori* (o distribuzione iniziale); la distribuzione $P(\theta | x)$ è la distribuzione di θ sotto la condizione che è stato osservato $X = x$, e si chiama *distribuzione a posteriori* (o distribuzione finale). Possiamo pensare a un numero (molto grande) di urne, indicizzate da una delle possibili determinazioni di θ , e al campionamento a due stadi, già accennato: prima si sceglie a caso un'urna (cioè una determinazione del parametro θ), poi si sceglie a caso un valore della variabile X secondo la distribuzione di probabilità dipendente dalla determinazione di θ scelta nel primo stadio.

Il più grande statistico italiano, Corrado Gini, e il più grande statistico mondiale, Ronald A. Fisher, hanno pienamente sottoscritto e condiviso l'impostazione bayesiana dell'inferenza *quando sussistono tutti i presupposti*, e in particolare quando tutte le probabilità in discorso, a cominciare dalle probabilità a priori, sono *oggettive*. Scrive Gini: "Se non possediamo tutti questi elementi di informazione – e, in pratica, ad eccezione di particolari esempi, non li possediamo – è assurdo pretendere di essere in grado di calcolare π_a [una formula equivalente a quella di $P(\theta | x)$]"

Gini espone il suo punto di vista con riferimento a un problema assai discusso nei primi decenni dello scorso secolo. Si ammetta di avere osservato un valore $X = x$ di una v.c. approssimativamente normale; si dirà che il valore osservato è un *valore accidentale* se appartiene a una v.c. X_0 di riferimento pre-assegnata; si dirà invece che x è un *errore sistematico* se proviene da una v.c. X_1 , che differisce in modo non trascurabile da X_0 . La formula che fornisce la probabilità π_a che un valore di $X \geq x$ sia da giudicarsi come *accidentale* deriva da una applicazione della "formula di Bayes":

$$\pi_a = \frac{p_0 P_0}{p_0 P_0 + p_1 P_1}$$

dove p_0 e p_1 sono rispettivamente la probabilità (a priori) che x sia accidentale oppure sistematico, mentre $P_0 = P(X_0 \geq x)$ e $P_1 = P(X_1 \geq x)$.

Applichiamo questa formula a un tipo di inferenza su cui si era scritto molto nel secolo XVIII, concernente il *rapporto di mascolinità delle nascite* (rapporto al totale delle nascite dei nati maschi). Dopo il

1662, anno di pubblicazione a Londra di un'operetta di John Graunt, basata su alcuni dati demografici delle parrocchie di Londra, era noto – almeno fra le persone colte – che sono prevalenti le nascite maschili, nella proporzione di circa 515 nati maschi ogni mille nati. La precedente formula può applicarsi anche a *valori puntuali*; si ammetta quindi che su mille nascite sia stato rilevato il valore $x = 0,515$ per il rapporto di nascite maschili sul totale, e che si sia scelta l'ipotesi di equiprobabilità di nascite maschili e femminili come “ipotesi accidentale”. Siano quindi: X_0 = nascite governate da identiche probabilità 0,5 per le nascite M e F, e X_1 = nascite maschili governate da una probabilità 0,515. Rispetto al valore osservato $x = 0,515$ su mille nascite si può calcolare:

$$P_0 = P(X_0 \geq 0,515) = 0,1718 ; P_1 = P(X_1 \geq 0,515) = 0,5.$$

Si ammetta in un primo tempo di inserire uguali probabilità a priori (cioè $p_0 = p_1 = 0,5$). Sostituendo nella formula di π_a si ottiene $\pi_a = 0,2557$ (e per l'evento contrario si ottiene la probabilità $1 - 0,2557 = 0,7443$); abbiamo quindi ottenuto probabilità favorevoli a una distribuzione *sistematica* delle nascite maschili. Se però si vuole tenere conto, nella assegnazione delle probabilità a priori, che il rapporto 0,515 è oggettivamente molto più plausibile di 0,5, possiamo assegnare i seguenti valori: $p_0 = 0,05$ e $p_1 = 0,95$. Sostituendo questi valori nella formula di π_a si ottiene $\pi_a = 0,01776$, e per la probabilità dell'evento contrario 0,9822. Se poi, lasciando questi valori per le probabilità a priori, ma riferendo le osservazioni al numero attuale di nascite in Italia in un intero anno, la probabilità π_a si riduce a un valore vicinissimo a 0, mentre diventa vicinissima a 1 la probabilità dell'evento contrario.

In questo caso vi è una *conoscenza oggettiva* (cioè indipendente da qualunque osservatore) che ci permette di calcolare le *probabilità finali* (ovvero *a posteriori*) in modo sicuro. *La situazione è ideale*: si parte da una distribuzione di probabilità, e si arriva – con un ragionamento *deduttivo* – a un'altra distribuzione di probabilità, che tiene conto delle osservazioni disponibili. *Ma questo accade molto raramente*. In effetti, però, in questo discorso il riferimento implicito è alla *ricerca scientifica*, o in ogni caso a conclusioni opponibili a chiunque. Altro discorso è quello della *ricerca operativa*, ad esempio per decisioni di tipo aziendale.

Negli anni venti del secolo XX Bruno de Finetti propose una nuova impostazione del Calcolo delle probabilità, e quindi sulle applicazioni statistiche basate sullo stesso Calcolo, coerente con l'imposta-

zione bayesiana: la distribuzione a priori poteva (o doveva) essere meramente soggettiva (ma non arbitraria, cioè doveva essere fondata su tutte le informazioni disponibili). Anche se questo non era il suo scopo, nella pratica delle applicazioni statistiche vennero a distinguersi due grandi aree: quella della ricerca scientifica (piuttosto ristretta), e quella della ricerca operativa (molto più ampia). In quest'ultimo ambito la probabilità soggettiva (o *epistemica*) poteva trovare un amplissimo campo applicativo. Esistono numerosi volumi, oltre a moltissimi articoli scientifici, dedicati all'inferenza bayesiana; per una esposizione già abbastanza estesa si possono raccomandare i volumi di Bernardo & Smith 1994 e di Lee 1997; un testo che tratta in generale le impostazioni inferenziali, assai raccomandabile, è quello di Cox & Hinkley 1974; una esposizione approfondita delle probabilità soggettive, con gli opportuni rimandi bibliografici, è quella di Regazzini 1988.

La debolezza dell'impostazione bayesiana con *probabilità soggettive* emerge in particolare nei processi civili e penali. Nonostante l'impegno di de Finetti in Italia, di Phil Dawid in Inghilterra e di J.B. Kadane negli Stati Uniti (al quale si deve un pregevole studio del celebre processo di Sacco e Vanzetti, che però non potrebbe mai essere accettato in un processo, dato l'impiego ineliminabile delle probabilità soggettive associate ad *eventi unici*), l'impostazione bayesiana è quasi assente nei processi civili e penali celebrati in questi paesi. Anche il successo di Dawid - con una argomentazione strettamente bayesiana - nel celebre processo di Sally Clark nel regno Unito, è stato basato su probabilità pienamente oggettive.

3. FISHER E I TEST DI SIGNIFICATIVITÀ

Sempre negli anni venti dello scorso secolo intervenne la proposta di Ronald Fisher, cioè di rinunciare alla assegnazione formale di probabilità a priori (quasi sempre soggettive), e impostare l'inferenza in modo *induttivo*, e *non deduttivo*, senza arrivare a una distribuzione di probabilità sulle ipotesi. In sostanza, Fisher ha formalizzato la procedura inferenziale più seguita nel passato; anche la prima applicazione a stampa di John Arbuthnot del 1710 era di questo tipo. La proposta di Fisher è sostanzialmente un adattamento ai fenomeni casuali del sillogismo classico denominato *Modus Tollens*, che ha la seguente struttura, dove H e K sono affermazioni che possono essere vere o false:

Premessa maggiore:	$H \text{ implica } K$
Premessa minore:	$K \text{ è falsa}$
Conclusione:	$H \text{ è falsa.}$

L'applicazione di questo argomento si riferisce a *leggi deterministiche*, e non può riferirsi a *leggi probabilistiche*, che possono ammettere fatti eccezionali, lontani dal comportamento "normale" (cfr. Frosini, 2009, p. 405). Era perciò indispensabile adattare il "modus tollens" ai fenomeni casuali, associando alla premessa maggiore (H implica K) una probabilità molto elevata, riferita agli eventi che possiamo aspettarci nella grande maggioranza dei casi, mentre la *non* accettazione di tale premessa (K è falsa) deve essere associata a uno o più eventi giudicati eccezionali. Fisher ha utilizzato, come introduzione ai test di significatività e ai piani sperimentali, il seguente esempio (che sembra sia stato effettivamente proposto a Fisher da una signora sua amica).

Questa signora afferma che assaggiando una tazza di tè con latte è in grado di distinguere se nella tazza è stato versato prima il tè o prima il latte. Allo scopo di sottoporre a verifica questa affermazione Fisher ha programmato un esperimento, consistente nel sottoporre alla signora otto tazze di tè con latte, quattro preparate in un modo e quattro nell'altro, disposte in ordine casuale (la c.d. *randomizzazione* è essenziale). Le permutazioni di 8 oggetti (le tazze), di cui quattro di tipo A e quattro di tipo B , sono 70. Se la signora *non ha* le pretese capacità discriminanti – è la nostra ipotesi H_0 (o *ipotesi nulla*) – queste 70 permutazioni si presentano simmetricamente; la signora avrebbe quindi probabilità $1/70$ di riconoscere (o meglio, di indovinare) la permutazione corretta, nota allo sperimentatore. Se si respinge l'ipotesi nulla H_0 si accetta implicitamente l'*ipotesi alternativa* (che in questo caso è facilmente individuabile, cioè si tratta dell'ipotesi che accetta la capacità discriminante del soggetto), che solitamente viene indicata con H_1 .

Se la signora riconosce la permutazione corretta, $1/70 = 0,0143$ può essere ritenuta da molti una probabilità *abbastanza piccola* da far giudicare plausibile la capacità discriminante del soggetto: si dice in tal caso che il risultato dell'esperimento casuale è *significativo*, ovvero che l'ipotesi alternativa viene *accettata*. Si chiama *livello di significatività* del test la probabilità totale, sotto l'ipotesi H_0 , dell'insieme di campioni che portano a respingere H_0 ; nell'esempio il livello di significatività è $1/70$.

Può essere interessante calcolare, secondo la formula di Bayes-

Gini, la probabilità associata a quattro “successi” nel suddetto esempio, sotto varie ipotesi di probabilità a priori:

con $p_0 = 0,99$ e $p_1 = 0,01$ (ipotesi a priori molto favorevole ad $H_0 =$ nessuna capacità discriminante), si ottiene $\pi_a = P(H_0 \mid X = 4) = 0,6111$, cioè una probabilità ancora favorevole ad H_0 , però meno favorevole della probabilità a priori;

con $p_0 = p_1 = 0,5$ si ottiene $\pi_a = 0,0156$, valore molto simile a $1/70 = 0,0143$, cioè con una probabilità a posteriori molto favorevole ad H_1 (ammissione di capacità discriminante);

con $p_0 = 0,05$ e $p_1 = 0,95$ (si ammette a priori che la signora abbia una effettiva capacità discriminante, anche se nel 5% dei casi può sbagliare), si ottiene $\pi_a = 0,000835$, cioè una probabilità moltissimo favorevole ad H_1 .

I termini *accettare un'ipotesi*, e *respingere un'ipotesi* sono entrati nell'uso statistico; ma Fisher e i suoi seguaci (che sono la grande maggioranza degli statistici) scrivono e ribadiscono più volte che si tratta di semplici etichette: accettare un'ipotesi significa semplicemente che i dati disponibili (fino ad ora) sono favorevoli a tale ipotesi, o almeno non solo tali da portarci a respingerla. Radhakrishna Rao, uno dei più grandi statistici dello scorso secolo, parla esplicitamente di *inferenza di prima approssimazione*, che richiede ulteriori verifiche e ulteriori campionamenti. In sostanza, un risultato “significativo” funziona solo come “campanello di allarme”. Per una analisi critica dell'impostazione di Fisher si possono consultare i suoi volumi del 1959 e del 1966, oltre all'impostazione generale contenuta nel volume di Cox & Hinkley 1974, già citato.

Di regola l'ipotesi nulla, come nell'esempio, comprende solo un'ipotesi *semplice*, cioè che descrive completamente una distribuzione di probabilità. Però le ipotesi in discorso possono anche essere *composte*, cioè comprendere più ipotesi semplici.

Molta letteratura, anche recente (Regina Nuzzo 2014) si è soffermata sul significato da attribuire al c.d. *valore-p* (in inglese *p-value*). Scelta una statistica X funzione dei dati campionari, per decidere fra le ipotesi H_0 e H_1 , si tratta della probabilità di ottenere un valore x di X altrettanto favorevole, o anche più favorevole, alla accettazione di H_1 , del valore osservato di X ; nel caso della *lady tasting tea*, si tratta semplicemente della probabilità di ottenere $X = 4$ (non ci sono valori più

estremi); in questo caso il *p-value* è $1/70$. Il *p-value* è anche chiamato “livello di significatività osservato”. Dato che il *p-value* è calcolato sulla base dell’ipotesi H_0 , si tratta della probabilità di sbagliare (respingendo H_0) in funzione dei dati campionari: è chiaro che è desiderabile ottenere un *p-value* molto basso (ma, osserva giustamente la Nuzzo, il *p-value* dipende dalla specifica ipotesi nulla che viene accettata).

Vi sono tre questioni principali riguardo alla interpretazione e all’uso del *p-value*:

- (1) non si tratta della probabilità che sia vera l’ipotesi H_0 ; tale probabilità è obiettivamente 1 o 0, secondo che H_0 sia vera o falsa, mentre può differire da questi valori nell’impostazione bayesiana, in cui si associano probabilità alle ipotesi;
- (2) se il *p-value* riguarda una statistica riferita alla differenza fra due popolazioni, un valore molto piccolo (ad es. $< 0,01$) non deve essere interpretato nel senso che le due popolazioni differiscono in modo notevole. Deve sempre essere ricordato che il *p-value* dipende anche dalla dimensione del campione, e che una piccola differenza fra le caratteristiche di due popolazioni può sempre risultare da un campione sufficientemente grande;
- (3) un requisito essenziale della ricerca scientifica è quello della *riproducibilità* su nuovi dati (v. ad es. Carl Hempel 1966); è perciò essenziale che un basso *p-value* ottenuto su un primo campione (magari di dimensione relativamente limitata) trovi conferma su un campione di nuovi dati. Ovviamente questo criterio della riproducibilità con nuovi dati deve valere con qualunque criterio inferenziale, non solo nel caso della inferenza eseguita con test di significatività.

4. L’INFERENZA STATISTICA NELL’IMPOSTAZIONE DI NEYMAN-PEARSON

Il polacco Jerzy Neyman e l’inglese Egon Pearson hanno proposto, nel periodo tra la fine degli anni venti e gli anni trenta dello scorso secolo, una nuova impostazione dell’inferenza statistica, che molti autori di libri di testo interpretano come un avanzamento rispetto all’impostazione di Fisher, mentre fu oggetto di un aspro dibattito tra Fisher e questi due autori.

Il punto di partenza di questi due autori fu quello di fare interve-

nire, esplicitamente e in modo operativo, l'ipotesi alternativa H_1 (che spesso è un'ipotesi composta, lasciata indeterminata). L'intero *spazio campionario* (insieme dei possibili campioni) viene suddiviso in due partizioni; se un campione appartiene alla prima partizione (*regione critica* C) si accetta l'ipotesi alternativa H_1 ; se appartiene alla partizione complementare A (regione di accettazione di H_0) si accetta H_0 . Si indicano solitamente con α e β le probabilità dei due tipi di errore che si possono commettere (dove \mathbf{X} indica il campione, che è una v.c. n -dimensionale):

$$\alpha = P(\mathbf{X} \in C ; H_0) : \text{si respinge } H_0 \text{ sotto la validità di } H_0;$$

$$\beta = P(\mathbf{X} \in A ; H_1) : \text{si respinge } H_1 \text{ sotto la validità di } H_1.$$

La probabilità α viene chiamata "probabilità dell'errore di prima specie"; la probabilità β viene chiamata "probabilità dell'errore di seconda specie".

E' chiaramente desiderabile che α e β (entrambe *probabilità di errore*) siano piuttosto piccole. Purtroppo fra le due probabilità esiste una relazione inversa: se riduco α aumento β , e viceversa; si tenga presente, a questo proposito, che queste due probabilità di errore sono calcolate facendo uso di due diverse distribuzioni di probabilità, valide rispettivamente sotto H_0 e H_1 . Concretamente, di solito si fissa la regione critica C in modo da ottenere una probabilità α abbastanza bassa; a questo punto la probabilità β risulta dal calcolo (se è troppo alta e la riduciamo, l'effetto è quello di aumentare α). Due testi fondamentali per la comprensione e l'applicazione in generale della impostazione inferenziale di Neyman-Pearson sono quelli di Lehmann 1986, e di Lehmann-Casella del 1998.

Se ad esempio l'ipotesi alternativa H_1 è composta, e si conosce come la probabilità β dipende dalla specificazione del parametro θ , è possibile in molti casi considerare esplicitamente la funzione $\beta(\theta)$ (cioè come varia β al variare delle ipotesi alternative comprese nell'insieme C).

In pratica, anche in questi casi è spesso opportuno – e a volte è richiesto – associare al valore di α uno specifico valore di β (nel senso di un valore di β associato a una data ipotesi alternativa semplice). Ad esempio, nei processi civili e penali celebrati negli Stati Uniti, per malattie o morti dipendenti dall'esposizione a un ambiente presumibilmente cancerogeno, è richiesto di calcolare la probabilità – sotto l'*ipotesi alternativa* che l'ambiente sia effettivamente cancerogeno - che si abbia un numero di morti (per la coorte considerata nel processo) dop-

pio di quello che ci si può aspettare per una identica dimensione campionaria, ma *senza* un effetto ambientale cancerogeno.

Ad esempio, nella esperienza dell'autore relativa al processo penale per i morti nello stabilimento chimico di Porto Marghera, concernente una coorte di 1.658 dipendenti, il numero di morti è stato modellato con una distribuzione di Poisson. Tale distribuzione, se riferita a una coorte della stessa dimensione e per le stesse classi di età, ma utilizzando i dati della popolazione veneta nel suo complesso (da cui provengono quasi tutti gli individui compresi nella coorte di dipendenti dello stabilimento), e per una regione di accettazione che arriva a comprendere fino a 12 morti per malattie cancerogene di un certo tipo, ha fornito un valore di $\alpha = 0,0427$; il corrispondente valore di β , calcolato rispetto a una distribuzione dello stesso tipo ma con media doppia (15 invece di 7,5), è stato di 0,2676.

RIFERIMENTI BIBLIOGRAFICI

- V. Barnett 1982, *Comparative statistical inference* (2nd Edition), Chichester, Wiley, 1982.
 J.M. Bernardo & A.F.M. Smith 1994, *Bayesian Theory*, Chichester, Wiley, 1994.
 D.R. Cox & D.V. Hinkley 1974, *Theoretical Statistics*, London, Chapman and Hall, 1974.
 R.A. Fisher 1959, *Statistical methods and scientific inference* (2nd Edition), Edinburgh, Oliver and Boyd, 1959.
 R.A. Fisher 1966, *The Design of experiments* (8th Edition), Edinburgh, Oliver and Boyd, 1966.
 B.V. Frosini 2009, *Metodi statistici*, Roma, Carocci, 2009.
 C. Hempel 1966, *Philosophy of Natural Science*, Englewood Cliffs, Prentice-Hall, 1966.
 P.M. Lee 1997, *Bayesian Statistics. An Introduction* (2nd Edition), London, Arnold, 1997.
 E.L. Lehmann 1986, *Testing Statistical Hypotheses* (2nd Edition), New York, Springer-Verlag, 1986.
 E.L. Lehmann & G. Casella 1998, *Theory of Point Estimation* (2nd Edition), New York, Springer-Verlag, 1998.
 R. Nuzzo 2014, *Statistical Errors*, Nature (vol. 506, 13 February 2014), 150-152.
 E. Regazzini 1988, *Subjective Probabilities*, Encyclopedia of Statistical Sciences, vol. 9, Wiley, New York, 1988, 55-64.