Statistica: una visione personale della disciplina[1]

Bruno Chiandotto

Dipartimento di Statistica, Informatica, Applicazioni “G. Parenti” (DiSIA)

Università di Firenze

Firenze, ottobre 2023

[1] Sintesi, la trattazione estesa è riportata in ***Bacci & Chiandotto*** 2019 e ***Chiandotto*** 2023.

It is easy to lie with statistics
It is hard to tell the truth without it

Andrejs Dunkels

1. Introduzione

Per comprendere i fenomeni reali occorre procedere all'analisi delle loro manifestazioni, se ci si domanda poi per quale ragione si è interessati ad una tale comprensione, la risposta è che si può voler soddisfare una esigenza conoscitiva o che la conoscenza è finalizzata alla risoluzione di uno specifico problema decisionale.

La disciplina che si occupa della raccolta e del trattamento *scientifico dei dati* (*manifestazioni dei fenomeni collettivi*[2]) e della loro trasformazione in *informazione* è la *Statistica*, se poi le informazioni stesse devono essere utilizzate per risolvere uno specifico problema decisionale, cioè un problema di scelta ottimale di una tra diverse alternative a disposizione, allora il contesto di riferimento è la *Teoria delle decisioni*. Nella fusione delle due discipline si sostanzia un'altra disciplina scientifica: "*La Teoria statistica delle decisioni*" o "*Teoria delle decisioni statistiche*" che deve essere intesa come generalizzazione ed estensione della Statistica che in questo modo risulta meglio caratterizzata nelle sue diverse articolazioni e meglio precisata nei contenuti.

La teoria delle decisioni fissa principi che consentono l'individuazione di regole di *scelta razionale*[3], gli sviluppi più recenti consentono anche di valutare e correggere eventuali incoerenze e contraddizioni nel comportamento dei decisori.

Nel contesto empirico l'elemento fondamentale di riferimento sono i dati (disponibili e/o acquisibili), la Statistica è la disciplina che tratta della raccolta delle manifestazioni dei fenomeni reali, dei *metodi e modelli* attraverso i quali le manifestazioni dei fenomeni, dovrebbero essere impiegati per ottenere una rappresentazione semplificata della realtà.

L'impiego dei modelli condiziona l'interpretazione dei risultati essendo il modello stesso caratterizzato da uno specifico schema concettuale di riferimento, si tratta di una interpretazione parziale da confrontare con altre possibili interpretazioni di risultati ottenuti impiegando *modelli* diversi; la logica decisionale fa emergere in modo inequivocabile tale necessità.

In tale ottica, assumono rilevanza fondamentale elementi quali l'esatta definizione della qualità e quantità d'*informazione disponibile*, la decisione sui tempi e sui modi di acquisizione e la valutazione dei costi connessi.

---

[2] Si definisce **collettivo** il **fenomeno** la cui misura e conoscenza richiede l'osservazione di una pluralità di sue manifestazioni.

[3] Si definisce *razionale* la *scelta* che soddisfa un serie di regole di comportamento (*assiomi di comportamento razionale*) che consentono al decisore, conformemente al proprio schema di preferenze, il conseguimento del massimo beneficio.

Il criterio guida nell'operare la trasformazione dei dati consiste, per quanto possibile, nell'evitare conseguenze negative. Ne deriva che l'informazione circa le conseguenze assume una rilevanza esclusiva ed un ruolo condizionante rispetto ad ogni altra tipologia disponibile per la quale si renderà, appunto, necessario un confronto o, meglio, un'integrazione con i dati di perdita già definiti.

La traccia originale viene, quindi, specificata e organizzata secondo lo schema seguente: avendo definito l'insieme dei risultati possibili (*le informazioni finali*) e avendo individuato le perdite corrispondenti, l'elaborazione deve essere effettuata avendo come obiettivo la *minimizzazione della perdita,* in tale struttura di riferimento è possibile formulare, senza perdere in generalità e senza condizionamenti, ogni problema statistico in termini decisionali.

Alcuni autori ritengono l'impostazione decisionale applicabile ai soli problemi con finalità operative, altri considerano la logica decisionale applicabile, secondo modalità particolari, a tutte le problematiche descrittive e/o inferenziali anche quando queste sono caratterizzate da finalità esclusivamente conoscitive, altri ancora ritengono la logica decisionale semplicistica ed oltremodo riduttiva.

Gli elementi a sostegno dell'impostazione decisionale sono innumerevoli e di varia natura, si può, innanzi tutto, osservare che la *duplice finalità, conoscitiva* e *operativa*, assegnata alla statistica quale disciplina scientifica, con conseguente attribuzione dei problemi decisionali alla seconda finalità, non è sostenibile, infatti, in entrambi i casi il problema si risolve sempre in una decisione, che poi questa sia orientata al cosa dire o al cosa fare è solo una questione di specificità della situazione in cui si opera. Se si parla poi di atti o decisioni in termini più generali, nel senso di scelte, la suddivisione diventa addirittura artificiosa se si pensa che ogni azione può essere considerata come l'effetto dell'affermazione: "*la decisione d è la migliore possibile poiché comporta la minore perdita di informazione"*.

Un altro rilevante aspetto della prospettiva decisionale, risiede nella logica interna propria *della teoria delle decisioni* che induce a formulare ed interpretare correttamente un problema statistico. Si tratta di una correttezza che può essere ricondotta a due fatti essenziali: gli *obiettivi che s'intendono perseguire* e il *patrimonio informativo a disposizione*. L'obiettivo della minimizzazione della perdita evidenzia, infatti, la parzialità e la particolarità del risultato che scaturisce dall'elaborazione: problemi analoghi affrontati con specificazioni diverse della funzione di perdita possono condurre, anzi generalmente conducono, a conclusioni diverse in quanto collegate ad elementi diversi dell'insieme delle decisioni possibili.

Quale soluzione è quella giusta è quale è quella sbagliata? Nessuna delle due, oppure entrambe se viste in ottiche diverse; il giudizio non deve essere formulato in termini di correttezza o errore, si può solo dire che, ritenendo valida (*accettabile, verosimile*) una struttura di perdita così come è rappresentata dalla funzione prescelta, la decisione migliore è quella che risulta dall'imposizione della condizione di perdita minima.

L'ultima riflessione si ricollega alla rappresentazione analitica (*modelli probabilistici*) delle manifestazioni dei fenomeni; i modelli oltre a facilitare la comprensione della realtà fattuale, consente l'effettuazione di analisi aventi finalità concettualmente diverse quali, ad esempio, l'esecuzione di calcoli previsionali.

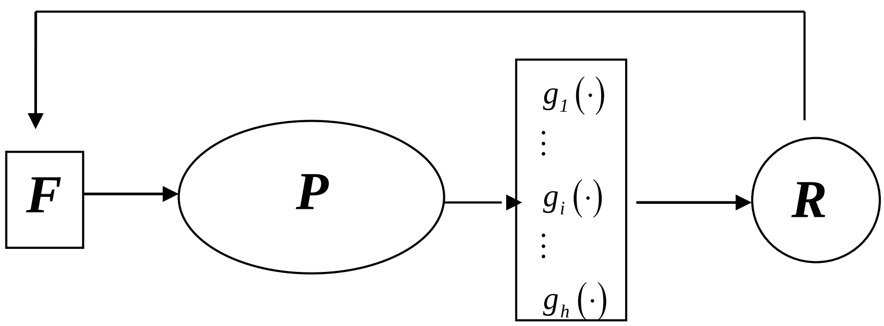
Non è infrequente imbattersi in situazioni operative nelle quali la decisione si riflette sulla situazione reale determinandone in qualche modo i mutamenti, diventa allora indispensabile procedere ad una ulteriore approfondimento dell'analisi avendo come obiettivo l'*individuazione dei nessi causali* presenti nel contesto di interesse; nessi causali che, una volta definiti nelle loro specificità, devono essere inseriti nella procedura di analisi valutandone l'impatto.

2. Articolazione della disciplina

L'articolazione tradizionale della disciplina distingue la *Statistica* in[4]:

- *Statistica descrittiva;*
- *Inferenza statistica classica;*
- *Inferenza statistica bayesiana.*

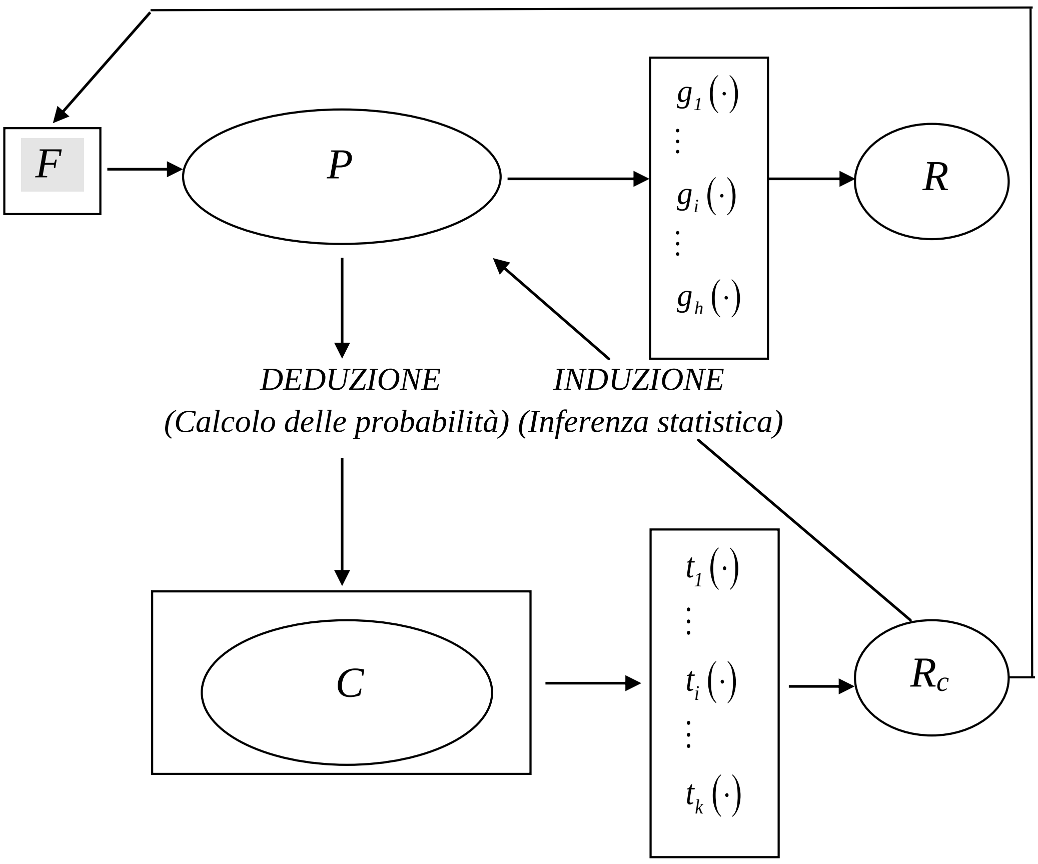
Nelle *Figg, 1, 2* e *3* è riportata una rappresentazione grafica dell'articolazione, nella *Fig. 4* , la rappresentazione grafica dell'inferenza statistica basata su modello.



F – Fenomeno reale, *P* - Popolazione (*Insieme di tutte le possibili manifestazioni del fenomeno, R* - Insieme delle rappresentazioni statistiche (*compattazione dei dati*) su *S.*

Fig. 1 - Statistica descrittiva.

[4] Per necessità espositive si considera il caso di un solo carattere X (*variabile casuale semplice*) e di un solo un solo parametro discreto θ (*stato di natura*): *costante incognita* nel contesto dell'inferenza statistica classica, *variabile casuale* nel contesto dell'inferenza statistica *bayesiana*; alle stesse conclusioni si perviene quando si considerano più caratteri (*variabile casuale multipla*) e più parametri che assumono valori nello spazio continuo e/o pluridimensionale.



C - Universo dei campioni, R_c - Insieme delle rappresentazioni statistiche su C.

Fig. 2 - Inferenza statistica classica.

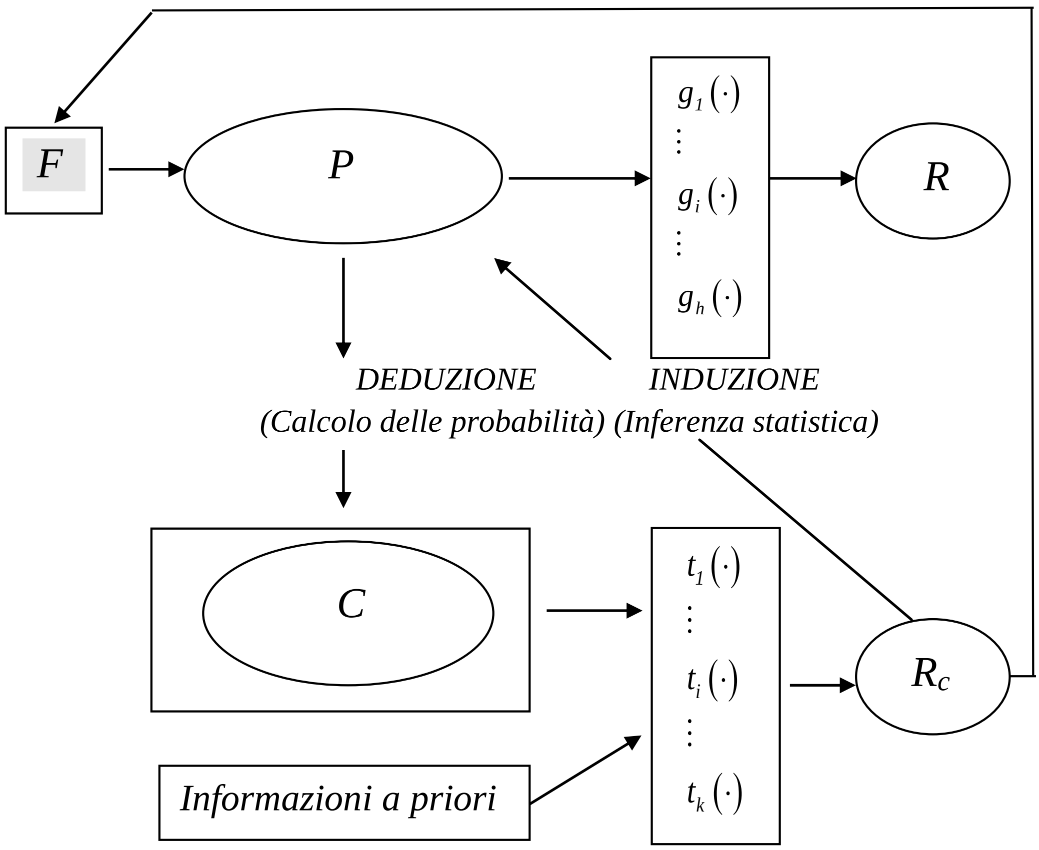


Fig. 3 - Inferenza statistica bayesiana.

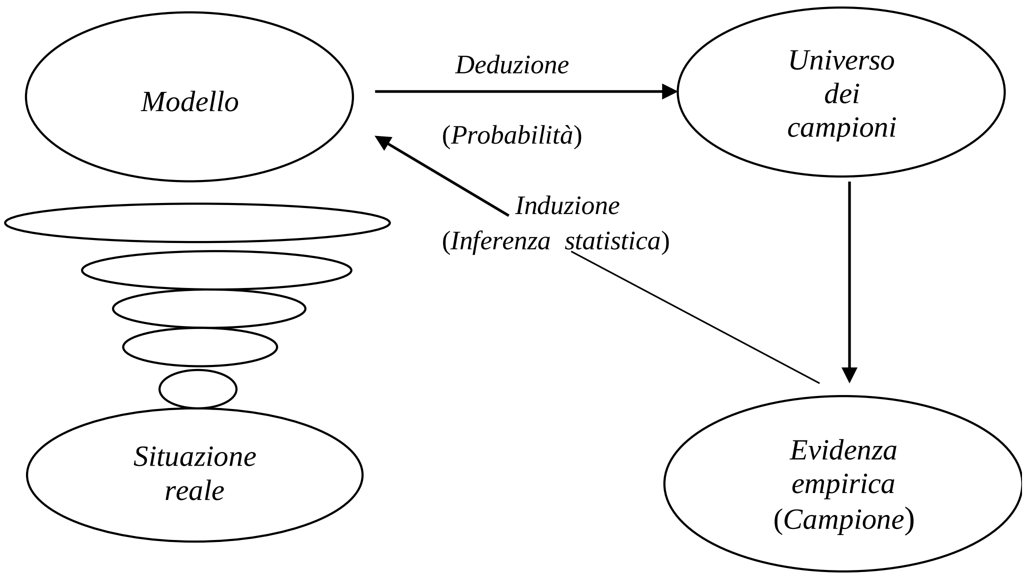


Fig. 4 - Inferenza statistica basata su modello.

Se x rappresenta le manifestazione del fenomeno oggetto d'analisi, la rappresentazione formale del modello probabilistico è:

$$f(x;\theta)$$

dove $f(\cdot)$ rappresenta la forma analitica del modello e θ lo ***stato di natura*** che specifica il modello.

Nel contesto dell'inferenza statistica classica θ è una costante incognita

$$\left[f(x;\theta) \Rightarrow f(x/\theta) \right]$$

nel contesto dell'inferenza statistica bayesiana θ è una variabile casuale con una propria distribuzione di *probabilità a priori*

$$\pi\left(\theta/\delta\right)$$

dove δ (*iperparametro*) è una costante incognita che specifica il modello.

Se si dispone di un insieme di manifestazioni del fenomeno (*campione*) $\boldsymbol{x} = \left(x_1, x_2, ..., x_i, ..., x_n\right)$[5] si ha:

1. nel contesto dell'*inferenza statistica classica*

$$\left[f(\boldsymbol{x};\theta) \Rightarrow f(\boldsymbol{x}/\theta) \right];$$

2. nel contesto dell'inferenza statistica bayesiana

$$\left[f(\boldsymbol{x};\theta,\delta) \Rightarrow f(\boldsymbol{x},\theta/\delta) \right]$$

applicando la ***formula di Bayes*** si derivano le *probabilità a posteriori*

$$\pi(\theta;\boldsymbol{x}/\delta) = \frac{\pi(\theta/\delta)\, f(\boldsymbol{x}/\theta,\delta)}{E_\theta\left[\pi(\theta)\, f(\boldsymbol{x}/\theta)\right]} = \frac{\pi(\theta/\delta)\, f(\boldsymbol{x}/\theta)}{f(\boldsymbol{x})}.$$

La formula di Bayes è fondamentale nello sviluppo della *Teoria statistica delle decisioni*, la formula non solo consente l'*aggiornamento della conoscenza* del fenomeno oggetto di studio attraverso l'acquisizione di ulteriori informazioni, ma, interpretata come *formula della probabilità delle cause,* diventa anche *strumento di intervento* sulle cause stesse nelle situazioni nelle quali la situazione reale, cioè lo *stato di natura* θ, risulta dipendente dall'azione (*state dependent utility*).

3. Teoria delle decisioni

Indicando con $A = \left(a_1, a_2, ..., a_i, ..., a_m\right)$ l'insieme delle azioni alternative possibili, con $\boldsymbol{\theta} = \left(\theta_1, \theta_2, ..., \theta_j, ..., \theta_s\right)$ l'insieme dei possibili stati di natura, con $c_{ij} = h\ (a_i, \theta_j)$ l'insieme delle conseguenze, il processo decisionale può essere rappresentato da una *Tavola di decisione* -(*cfr. Tab. 1*) e/o da un *Albero di decisione* (*cfr. Fig. 5*).

[5] Il carattere in grassetto è utilizzato per rappresentare le variabili casuali multidimensionali (*vettori casuali*).

Azione	Stato di natura					
	θ_1	θ_2	.....	θ_j	.....	θ_s
a_1	c_{11}	c_{12}	.....	c_{1j}	.....	c_{1s}
a_2	c_{21}	c_{22}	.....	c_{2j}	.....	c_{2s}
.....	.....	.....	.....	.....	.....	.....
a_i	c_{i1}	c_{i2}	.....	c_{ij}	.....	c_{is}
.....	.....	.....	.....	.....	.....	.....
a_m	c_{m1}	c_{m2}	.....	c_{mj}	.....	c_{ms}

Tab. 1 - Tavola di decisione

In molte situazioni decisionali le conseguenze $c_{ij} = h\ (a_i, \theta_j)$ non sono espresse in termini numerici il che rende molto difficile, a volte impossibile, la risoluzione del problema di scelta tra le alternative a disposizione, comunque, se il comportamento del decisore è razionale, si dimostra l'*esistenza* e *l'unicità*, a meno di trasformazioni lineari positive, di una funzione a valori reali, definita *funzione di utilità* $c_{ij} = h\ (a_i, \theta_j) \Rightarrow u\ (a_i, \theta_j) = u_{ij}$ (cfr. *Tab. 2*).

Se il decisore conosce lo stato di natura (*situazione di certezza*) ad es. θ_j, il problema di scelta si riduce al confronto tra *m* conseguenze (nell'es. $u_{1j}, u_{2j}, \cdots, u_{ij}, \cdots, u_{mj}$), la *scelta razionale* è l'azione che comporta, conformemente al proprio schema di preferenze, il conseguimento del massimo beneficio per il decisore (*scelta giusta*[6])

$$a_i = arg\left\{\max_i u_{ij}\right\}.$$

In situazioni di certezza la *scelta razionale* equivale alla *scelta giusta*, sempre che siano note le conseguenze ed il decisore sia in grado di esprimere, in modo razionale, le proprie preferenze riguardo alle conseguenze stesse.

[6] Una decisione è giusta se si risolve in esiti ottimali per il soggetto che effettua la scelta.

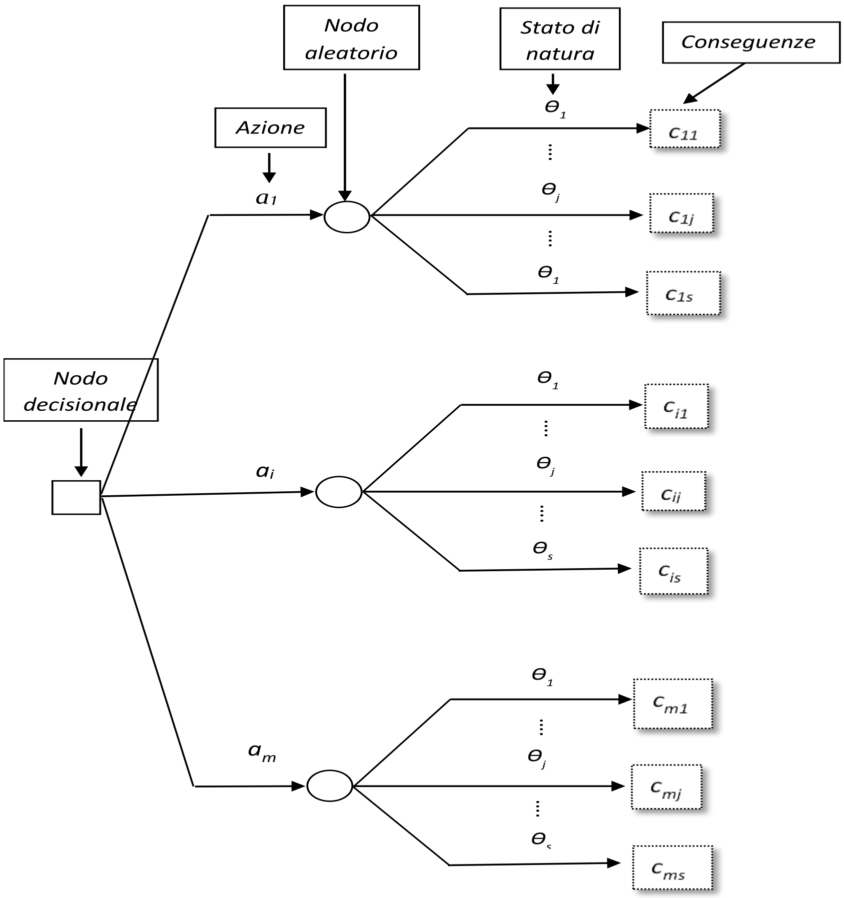


Fig. 5 - Albero di decisione.

Azione	***Stato di natura***					
	θ_1	θ_2	…..	θ_j	…..	θ_s
a_1	u_{11}	u_{12}	…..	u_{1j}	…..	u_{1s}
a_2	u_{21}	u_{22}	…..	u_{2j}	…..	u_{2s}
…..	…..	…..	…..	…..	…..	…..
a_i	u_{i1}	u_{i2}	…..	u_{ij}	…..	u_{is}
…..	…..	…..	…..	…..	…..	…..
a_m	u_{m1}	u_{m2}	…..	u_{mj}	…..	u_{ms}

Tab. 2 - Tavola di decisione con conseguenze espresse in termini di utilità.

Se il decisore non conosce lo stato di natura (*situazione di incertezza*) risulta impossibile effettuare la scelta giusta si può, però, effettuare *una scelta razionale*[7].

Nella generalità dei casi il decisore, dopo aver definito l'insieme delle azioni, l'insieme degli stati di natura, l'insieme delle conseguenze $c_{ij} \Rightarrow u\ (a_i, \theta_j)$ e il modello probabilistico $f\left(x;\theta\right)$ che intende impiegare per rappresentare il fenomeno analizzato, si trova ad operare in situazioni nelle quali lo stato di natura non è noto (*situazioni di incertezza*), risulta impossibile prendere una decisione giusta poiché gli esiti della scelta saranno noti solo a posteriori, è possibile, però, decidere razionalmente in funzione del patrimonio informativo disponibile

Una possibile classificazione delle teorie sviluppate per trattare scientificamente i diversi contesti operativi in funzione del *patrimonio informativo disponibile* è è riportata nella *Tab. 3*.

L'intervento indicato ai punti 5 e 6 trasforma il contesto di analisi da *osservazionale* a *quasi-sperimentale* soddisfacendo il principio basilare della causalità

"***No causation without manipulation***".

Nella *Tab. 3* è riportata la specifica del livello informativo sullo stato di natura.

Patrimonio informativo a disposizione	***Teoria***	***Probabilità associate allo stato di natura***
1. Assenza assoluta di informazioni	*Classica delle decisioni*	*Nessuna*
2. Solo informazioni a priori	*Bayesiana delle decisioni*	$\pi\left(\theta_1 / \delta\right)$
3. Solo informazioni campionarie	*Statistica classica delle decisioni*	*Nessuna*
4. Informazioni a priori e informazioni campionarie	*Statistica bayesiana delle decisioni*	$\pi\left(\theta_j / \delta, \mathbf{x}\right)$
5. Informazioni a priori e informazioni campionarie con intervento	*Probatoria delle decisioni (Evidential decision teory)*	$\pi\left(\theta_j / \delta, a_i,\ \text{osservo}\ \mathbf{x}\right)$
6. Informazioni a priori e informazioni campionarie con intervento	*Causale delle decisioni (Causal decidion theory)*	$\pi\left(\theta_j / \delta, a_i,\ \mathit{fisso}\ \mathbf{x}\right)$

Tab. 3 - Classificazione dei contesti operativi in funzione delle informazioni disponibili.

***

4. Assenza di informazioni (*situazioni di estrema incertezza*)

Se non si possiede alcuna informazione i criteri di decisione più diffusi sono:

1. ***Criterio del max-min*** *(criterio di Wald)* - Il criterio consiste nello scegliere l'azione a_i che corrisponde al massimo dell'utilità minima

$$a_i = arg\left\{\max_i\left(\min_j u_{ij}\right)\right\}$$

2. ***Criterio del max-max*** - Il criterio consiste nello scegliere l'azione. a_i che corrisponde al massimo dell'utilità massima

$$a_i = arg\left\{\max_i\left(\max_j u_{ij}\right)\right\}$$

3. ***Criterio intermedio*** *(criterio di Hurwicz)* – Il criterio è definito dalla relazione

$$a_i = arg\left\{\max_i\left[(1-\alpha)\min_j\ u_{ij} + \alpha\ \max_j\ u_{ij}\right]\right\}$$

il coefficiente $0 \le \alpha \le 1$ della combinazione lineare convessa è definito *indice di ottimismo*: per $\alpha = 1$ si ha il criterio del *max-max* per $\alpha = 0$ si ha il criterio del *max-min*.

4. ***Criterio del min-max rimpianto*** *(criterio di Savage)* - Per applicare questo criterio occorre sostituire agli elementi di ciascuna colonna la differenza tra l'elemento che ha valore massimo e l'elemento che occupa quella posizione

$$r_{ij} = \max_i u_{ij} - u_{ij}$$

scegliendo poi l'azione a_i per la quale il massimo rimpianto assume valore minimo

$$a_i = arg\ \left\{\min_i\left(\max_j\ u_{ij}\right)\right\}$$

Attraverso il criterio del *min-max rimpianto* l'operatore cerca di minimizzare i danni di una decisione errata.

5. ***Criterio della ragione insufficiente*** *(criterio di Laplace)* - La scelta a_i risulta dalla relazione

$$a_i = arg\left\{\max_i\left(\sum_{j=1}^{s} u_{ij}\right)\right\}.$$

I primi quattro criteri di decisione risultano del tutto ragionevoli, naturalmente gli argomenti che possono essere addotti a sostegno dell'uno o dell'altro criterio, sono di natura diversa, il che in effetti non presenta gravi inconvenienti. L'aspetto più problematico riguarda invece l'applicazione dei quattro criteri: infatti se essi vengono adottati in uno stesso problema di decisione ne può risultare una scelta di azioni differenti; da alcuni autori tale fatto è stato preso a pretesto per affermare che uno o più criteri devono essere necessariamente errati. Non sembra comunque che esistano ragioni sufficienti a sostegno di tale punto di vista; più ragionevole è l'affermazione che tutti i criteri proposti hanno una portata applicativa limitata e che, a seconda delle circostanze, dovrà essere adottato il criterio che il decisore ritiene più appropriato.

Relativamente al quinto criterio (*criterio della ragione insufficiente*) si attribuisce implicitamente a tutti gli stati di natura θ_j uguale probabilità; ciò viene fatto in quanto non si hanno motivi sufficienti per ritenere che la distribuzione delle probabilità sia diversa da quella uniforme. Evidentemente una tale giustificazione risulta del tutto insoddisfacente perché non sono chiari i motivi per i quali lo stato d'ignoranza completa debba implicare necessariamente un'uguale probabilità degli stati di natura. Pertanto, o si prende atto che si sta operando in situazioni di estrema incertezza, cioè di ignoranza completa riguardo alla plausibilità (*probabilità*) dei vari stati di natura, e si agisce di conseguenza, oppure si dovrà procedere esplicitamente alla *valutazione* delle probabilità associate ai diversi valori di θ.

5. Disponibilità delle sole informazioni a priori $\pi\ (\theta_j / \delta)$

Se il decisore dispone delle sole informazioni a priori $\pi\ (\theta_j / \delta)$, il criterio di decisione di più largo impiego è il *criterio dell'utilità attesa* definito dalla relazione

$$a_i = arg \left\{ \max_i \left(\sum_{j=1}^{s} u_{ij}\ \pi(\theta_j / \delta) \right) \right\}.$$

6. Disponibilità delle sole informazioni campionarie $\mathbf{x} = (x_1, x_2, \dots, x_h, \dots, x_n)$

Se il decisore dispone delle sole informazioni campionarie $\mathbf{x} = (x_1, x_2, \dots, x_h, \dots, x_n)$ lo sviluppo della teoria fa riferimento alle perdite (*utilità negative*) $l_{ij} = -u_{ij} = l\ (a_i, \theta_j)$ e alle funzioni di decisione $d(\mathbf{x}) \Rightarrow a_i$. Si tratta dell'utilizzo tradizionale nell'ambito dell'inferenza statistica (es. errore quadratico medio quando si procede alla stima di θ, *probabilità dell'errore di I e II tipo* quando si procede alla verifica di ipotesi formulate su θ).

Decisione	***Stato di natura***					
	$\theta_1 / \mathbf{x}$	$\theta_2 / \mathbf{x}$	…..	$\theta_j / \mathbf{x}$	…..	$\theta_s / \mathbf{x}$
d_1	$l(d_1, \theta_1 / \mathbf{x})$	$l(d_1, \theta_2 / \mathbf{x})$	…..	$l(d_1, \theta_j / \mathbf{x})$	…..	$l(d_1, \theta_s / \mathbf{x})$
d_2	$l(d_2, \theta_1 / \mathbf{x})$	$l(d_2, \theta_2 / \mathbf{x})$	…..	$l(d_2, \theta_j / \mathbf{x})$	…..	$l(d_2, \theta_s / \mathbf{x})$
…..	…..	…..	…..	…..	…..	…..
d_i	$l(d_i, \theta_1 / \mathbf{x})$	$l(d_i, \theta_2 / \mathbf{x})$	…..	$l(d_i, \theta_j / \mathbf{x})$	…..	$l(d_i, \theta_s / \mathbf{x})$
…..	…..	…..	…..	…..	…..	…..
d_r	$l(d_r, \theta_1 / \mathbf{x})$	$l(d_r, \theta_2 / \mathbf{x})$	…..	$l(d_r, \theta_j / \mathbf{x})$	…..	$l(d_r, \theta_s / \mathbf{x})$

Tab. 4 - Tavola con funzioni di decisione, informazioni campionarie e conseguenze espresse in termini di perdite monetarie

In questo contesto si procede al computo della *perdita attesa* introducendo il concetto di *rischio*

$$R_{ij} = R\left(d_i, \theta_j\right) = E_x\left[\, l\left(d_i, \theta_j / \boldsymbol{x}\right)\right] \quad \text{per } i = 1,2,\ldots,r;\ j = 1,2,\ldots,s\ .$$

la corrispondente tavola di decisione è riportata nella *Tab. 4*.

Azione	***Stato di natura***					
	θ_1	θ_2	.....	θ_j	.....	θ_s
d_1	$R(d_1,\theta_1)$	$R(d_1,\theta_2)$	.....	$R(d_1,\theta_j)$	.....	$R(d_1,\theta_s)$
d_2	$R(d_2,\theta_1)$	$R(d_2,\theta_2)$	.....	$R(d_2,\theta_j)$	.....	$R(d_2,\theta_s)$
.....	.....	.....	.....	.....	.....	.....
d_i	$R(d_i,\theta_1)$	$R(d_i,\theta_2)$	.....	$R(d_i,\theta_j)$	.....	$R(d_i,\theta_s)$
.....	.....	.....	.....	.....	.....	.....
d_r	$R(d_r,\theta_1)$	$R(d_r,\theta_2)$	.....	$R(d_r,\theta_j)$	.....	$R(d_r,\theta_s)$

Tab. 5 - Tavola di decisione e conseguenze espresse in termini di rischi.

La considerazione dei rischi, e non delle perdite associate a ciascuna azione non modifica il livello informativo, la situazione decisionale è sempre caratterizzata da estrema incertezza, pertanto, i criteri di decisioni adottabili sono quelli elencati per tale contesto.

.

7. Disponibilità di informazioni a priori e informazioni campionarie

Se il decisore dispone di informazioni a priori $\pi\left(\theta_j / \delta\right)$ e informazioni campionarie $\boldsymbol{x} = (x_1,\ x_2,\ \ldots\ ,x_h,\ \ldots, x_n)$, si deriva la *probabilità a posteriori* $\pi\left(\theta_j / \delta, \boldsymbol{x}\right)$, pertanto, il *criterio di decisione dell'utilità attesa* è definito dalla relazione

$$a_i = arg\left\{ \max_i \,(\sum_{j=1}^{s} u_{ij}\ \pi(\theta_j / \delta, x) \right\}.$$

8. Disponibilità di informazioni a priori e informazioni campionarie con intervento

Se le probabilità a posteriori, oltre a dipendere dallo stato di natura θ e dall'evidenza campionaria $\boldsymbol{x}$, dipendono anche dall'azione prescelta a_i (*intervento*) due sono le alternative possibili: *osservo* ***x*** o *fisso* ***x*** con probabilità rispettive

$$\pi\left(\theta_j / \delta, a_i, osservo\ \boldsymbol{x}\ \right)$$

$$\pi\left(\theta_j \,/\, \delta, a_i,\ fisso\ \mathbf{x}\right).$$

Il criterio di decisione dell'utilità attesa definisce le azioni ottimali nelle relazioni

$$a_i = arg\left\{\max_i \left(\sum_{j=1}^{s} u_{ij}\ \pi\left(\theta_j \,/\, \delta, a_i, osservo\ \mathbf{x}\ \right)\right\}$$

$$a_i = arg\left\{\max_i \left(\sum_{j=1}^{s} u_{ij}\ \pi\left(\theta_j \,/\, \delta, a_i,\ fisso\ \mathbf{x}\right)\right\}.$$

In letteratura è trattata la teoria definita *Teoria statistica delle decisioni in forma normale* che prevede l'introduzione di una probabilità a priori $\pi\left(\theta_j\right)$ e il successivo computo del *rischio atteso* $\bar{R}_{ij} = \sum_{i=1}^{s}\left[R_{ij} \cdot \pi\left(\theta_i \,/\, \delta\right)\right]$.

Applicando il criterio di decisione della minimizzazione del rischio atteso, si perviene alle stesse conclusioni cui si perviene applicando il criterio della massimizzazione dell'utilità attesa disponendo di informazioni a priori e informazioni campionarie.

Nonostante l'equivalenza delle conclusioni si deve, comunque, osservare che la *Teoria statistica in forma normale* è concettualmente inaccettabile; infatti, vale il giudizio espresso da ***L. J. Savage*** nei confronti dell'*Inferenza statistica fiduciale* proposta da ***R. A. Fisher***: "*fare una frittata bayesiana senza utilizzare le uova bayesiane"*, si ipotizza, infatti, una variabilità (*virtuale*) di un valore costante.

9. Conclusioni

Gli individui, gli enti, le istituzioni, … regolano il proprio comportamento basandosi su un processo, spesso inconscio, di ragionamento causale non supportato, nella generalità dei casi, da dati e analisi statistiche adeguate.

Inferenze causali si possono trarre sia da dati sperimentali che non sperimentali; tuttavia, regole meccaniche non risolvono il problema, inoltre, l'analisi causale oltre a trovare piena giustificazione nella prospettiva decisionale, in tale contesto rivela tutte le sue potenzialità

Premesso che l'analisi statistica dei dati occupa una posizione di rilievo nel panorama delle analisi scientifiche formalizzate che comportano, nella generalità dei casi, l'introduzione di modelli probabilistici in grado di sintetizzare e rappresentare la complessa realtà fenomenica avendo come obiettivo la sua comprensione ed il suo controllo, si richiamano gli elementi fondamentali da tenere in considerazione per il perseguimento di esiti significativi sia dal punto di vista teorico sia, e soprattutto, dal punto di vista applicativo.

1. Approfondita conoscenza del fenomeno che s'intende analizzare.
2. Ipotesi alternative di possibili modelli probabilistici rappresentativi della situazione reale devono essere ipotizzate e adeguatamente verificate.

3. Tutti i modelli, per definizione, sono falsi l'importante è che siano utili.
4. Inferenze causali si possono applicare sia a dati sperimentali che osservazionali, tuttavia, regole meccaniche non risolvono il problema.
5. Ipotesi alternative di possibili percorsi causali devono essere formulate e adeguatamente verificate.

L'inferenza causale, nella prospettiva decisionale, sempre possibile se si fa riferimento alla definizione soggettiva della probabilità, oltre a soddisfare i requisiti elencati, consente il dimensionamento ottimale delle rilevazioni campionarie.

L'integrazione del patrimonio informativo a disposizione è una necessità che deve essere sempre attentamente valutata, la prospettiva decisionale consente di fornire una risposta soddisfacente a tale necessità.

Infine occorre tener presente che (***L. Pirandello***: "*6 personaggi in cerca di autore*")

" ***Un fatto è come un sacco: vuoto, non si regge. Perché si regga, bisogna prima farci entrar dentro la ragione e i sentimenti che lo han determinato***".

La Statistica, interpretata in prospettiva decisionale, oltre ad evidenziare i limiti dell'inferenza sia classica che bayesiana, rappresenta un ritorno alle origini della disciplina come risulta evidente dalla prima rilevazione censuaria effettuata nel 3800 aC nell'Impero Babilonese che ha riguardato la produzione di beni alimentari, il numero di uomini, donne, bambini, bestiame, schiavi presenti nel regno, censimento, effettuato per il perseguimento di finalità operative di intervento, soprattutto di ordine militare e di tassazione; finalità operative analoghe caratterizzano anche i primi censimenti effettuati in Cina (3000 aC), in Egitto (2500 aC) e nell'Impero Romano (600 aC).

Dal 3800 aC ad oggi la Statistica ha registrato un notevole sviluppo (*cfr. Fig. 6*), soprattutto a partire dal 1763, anno di pubblicazione dell'articolo di ***Thomas Bayes***: *Essay Towards Solving a Problem in the Doctrine of Chances* pubblicato postumo da ***R***. ***Price*** in *Philosophical Transactions of the Royal Society of London*, 53.

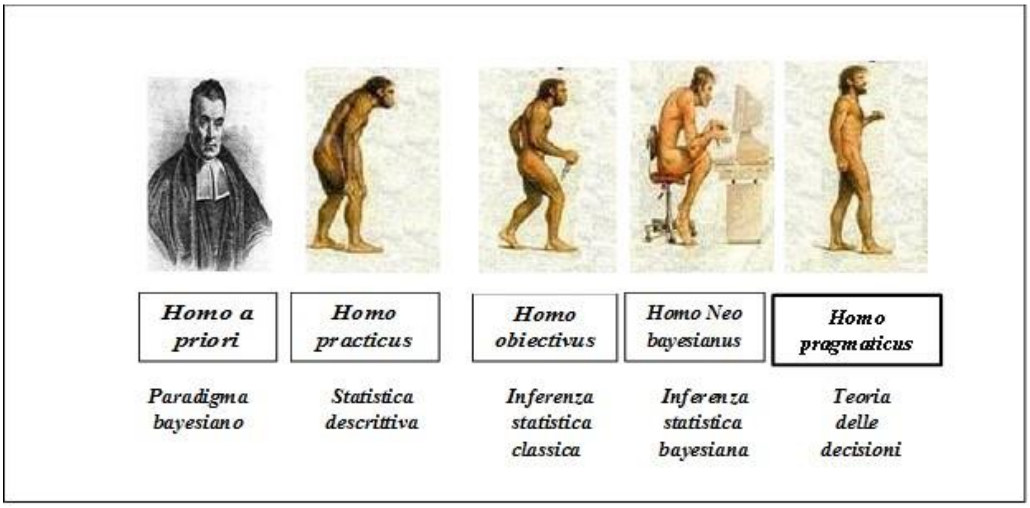


Fig. 6 - Evoluzione della Statistica

I momenti di svolta più significativi del processo evolutivo della Statistica sono:

- l'***introduzione di modelli probabilistici*** (*homo obiectivus*) rappresentativi della realtà fenomenica;
- lo ***sviluppo dell'informatica*** (*homo neo-bayesianus*);
- l'***intervento sulle cause*** (*homo pragmaticus*).

Riferimenti bibliografici

S*. *Bacci *e* ***B*. *Chiandotto*** (2019). *Introduction to Statistical Decision Theory, Utility Theory and Causal Analysis",* Chapman & Hall CRC, New York • London .

B*. *Chiandotto (2023). *Statistica e Decisioni*, in fase di completamento.