

*Per comprendere il presente e prevedere
il futuro occorre conoscere il passato¹*

***Evoluzione della Statistica
da Thomas Bayes (1701-1761) a oggi (2021)***

(Firenze, ottobre 2021)

Bruno Chiandotto

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

¹ In questa nota si procede ad una sommaria sintesi del processo evolutivo della disciplina; una trattazione più ampia e articolata è riportata in **Bacci-Chiandotto** 2019 e **Chiandotto** 2020.

Introduzione

Per comprendere i fenomeni, qualunque sia la loro natura, 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 duplice:

- si vuole soddisfare una mera esigenza conoscitiva;
- la conoscenza è finalizzata alla risoluzione di uno specifico problema decisionale;

Qualunque problema decisionale, dal più banale al più complesso, richiede la chiara definizione del problema stesso e l'individuazione delle possibili relazioni che connettono i vari elementi e aspetti che lo caratterizzano.

La disciplina che si occupa della raccolta e del trattamento *scientifico* delle manifestazioni dei fenomeni è la **Statistica**, se poi la finalità che si intende perseguire è la risoluzione di un problema decisionale, cioè un problema che si risolve nella scelta 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 può essere intesa come generalizzazione ed estensione della Statistica che in questo modo risulta anche meglio caratterizzata nelle sue diverse articolazioni e meglio precisata nei contenuti.

La teoria statistica delle decisioni fissa principi razionali di comportamento che consentono la derivazione di regole di scelta ottimale, gli sviluppi più recenti di tale teoria consentono anche di valutare e correggere eventuali incoerenze e contraddizioni nel comportamento dei decisori.

Per la comprensione dei fenomeni l'elemento fondamentale di riferimento è l'insieme delle loro manifestazioni (disponibili e/o acquisibili), la statistica è la disciplina che tratta di metodi attraverso i quali le manifestazioni dei fenomeni di interesse, dovrebbero o potrebbero essere impiegati per ottenere, in funzione delle specificità del problema analizzato, una rappresentazione semplificata della realtà. Le **manifestazioni** vengono trasformate in **dati statistici**, dati che vengono successivamente trasformati in **informazione** e questa in **conoscenza** che fa emergere le caratteristiche che interessano in dipendenza degli obiettivi che s'intendono perseguire, se l'obiettivo è la risoluzione di un problema di scelta tra diverse alternative a disposizione la **conoscenza**, evidenziando anche i nessi di causalità eventualmente presenti tra le entità osservate, è utilizzabile a fini previsionali e di intervento.

Una convinzione abbastanza diffusa è che i dati siano in grado di “*parlare da soli*”, si tratta di una convinzione semplicistica e, spesso, priva di fondamento. In realtà i dati sono fatti ma “*un fatto è come un sacco, vuoto non si regge, perché si regga, bisogna farci entrare la ragione e i sentimenti (le cause) che lo hanno generato*”.

In tale ottica, assumono rilevanza fondamentale elementi quali l’esatta definizione della qualità e quantità d’informazione disponibile e/o che s’intende acquisire, la decisione sui tempi e sui modi di acquisizione e la valutazione dei costi connessi.

La statistica, che si occupa della raccolta e del trattamento finalizzato dei dati, entra, a pieno titolo, in ogni processo decisionale, finalizzato al conseguimento di uno specifico obiettivo, che prevede l’utilizzo di informazioni; un processo nel quale:

- l’**input** (*materia prima*) è costituito dai dati opportunamente trasformati in informazione riferita ad una specifica realtà, della quale rappresentano qualitativamente e/o quantitativamente uno o più aspetti;
- l’**output** (*prodotto*) è la conoscenza del fenomeno indagato la cui natura e interpretazione dipende dagli input utilizzati e dal meccanismo logico e metodologico di elaborazione e dalle finalità che s’intendono perseguire

Da quanto sopra sottolineato, occorre distinguere almeno tre insiemi di elementi fondamentali:

- le informazioni che vengono introdotte quali input nel processo di trasformazione (*dati e conoscenze pregresse*);
- le procedure di elaborazione adottate (*metodi e modelli statistici*);
- le *conseguenze*.

Se, ad esempio, si vuol procedere a una *compattazione* di dati osservati con l’obiettivo di mettere in risalto una caratteristica specifica del fenomeno, le conseguenze potranno essere descritte dalla perdita (di informazione) determinata dal processo, oppure, se il problema è finalizzato alla scelta di una modalità operativa, le conseguenze potranno essere identificate in perdite monetarie o di altra natura ed essere, eventualmente, collegate ad errori commessi nel processo di trasformazione.

Il criterio guida nell’operare la trasformazione dei dati in informazione consiste, per quanto possibile, nell’evitare conseguenze negative. Le conseguenze del processo di trasformazione assumono, pertanto, 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 precedentemente 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*.

Considerando che, con l’impostazione sopra delineata, sulla base di una funzione di perdita viene selezionato l’elemento ottimo (quello cioè cui corrisponde la perdita minima), risulta naturale attribuire agli specifici elementi che costituiscono l’insieme dei prodotti l’identità di *decisioni*, nel senso che essi rappresentano le alternative possibili del risultato di un problema del quale è stato specificato l’obiettivo. Si può, pertanto,

accantonare l'espressione *processo di trasformazione (compattazione) dei dati* e fare riferimento al *problema decisionale* quale schematizzazione che prevede la lista di tutti gli esiti possibili e la scelta di quello ottimo secondo il *criterio di minimizzazione della perdita attesa*.

Assumendo come struttura di riferimento quella sopra introdotta è possibile formulare, senza perdere in generalità e senza condizionamenti, ogni problema statistico in termini decisionali.

A proposito dell'ultima affermazione, si manifesta un ulteriore problema di estrema rilevanza riguardo ai vantaggi e alle difficoltà, se non addirittura alla ragionevolezza, propria di una impostazione quale è quella decisionale. Alcuni autori ritengono l'impostazione decisionale applicabile ai soli problemi con finalità operative, altri ritengono la logica decisionale semplicistica ed oltremodo riduttiva, alcuni autori considerano la logica decisionale applicabile, secondo modalità particolari, a tutte le problematiche descrittive e/o inferenziali della statistica anche quando queste sono caratterizzate da finalità esclusivamente conoscitive.

Gli elementi a sostegno dell'impostazione decisionale sono innumerevoli e di varia natura. Si può, innanzi tutto, osservare che la duplice finalità, conoscitiva ed operativa, assegnata alla statistica quale disciplina scientifica, con conseguente attribuzione dei problemi decisionali alla seconda finalità, si risolve, semplicemente, nella specificazione della duplice tipologia di prodotti che costituiscono l'insieme delle decisioni espresse come:

- *azioni* da intraprendere e da realizzare concretamente;
- *affermazioni* da formulare le quali, a loro volta, possono configurarsi come asserzioni che specificano la *conformità* dell'evidenza osservata nei confronti di una o più ipotesi assunte a priori;
- *asserzioni* che specificano la stima di una quantità incognita e/o l'adattamento di un modello teorico ad una specifica realtà di interesse, ecc..

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 lo statistico 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. Ne scaturisce, allora, una differenza solo verbale, terminologica e, quindi, non sostenibile dal punto di vista concettuale.

Un altro rilevante aspetto che può consigliare la scelta decisionale, risiede nella logica interna propria *della teoria statistica 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 i dati utilizzabili. L'obiettivo della minimizzazione della perdita attesa 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 tra due soluzioni possibili? 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 attesa minima.

L'ultima riflessione si ricollega alla necessità di attribuire ai dati una specifica forma per poter ottenere una rappresentazione degli stessi significativa. La realtà non è né descrivibile né rappresentabile senza ricorrere a schemi concettuali di riferimento e in corrispondenza di ognuno di essi si ottiene un risultato, evidentemente parziale e condizionato allo schema specificato. La logica decisionale, i cui risultati sono condizionati alla particolare funzione di perdita specificata, fa emergere in modo inequivocabile tale consapevolezza.

Non è infrequente imbattersi in problemi operativi nei quali la decisione determina mutamenti nella situazione reale, 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 valutando l'impatto sulle conseguenze relative alle singole azioni attivando, eventualmente, interventi in grado di modificare, a proprio vantaggio, l'evoluzione naturale dei fenomeni analizzati.

1. Statistica descrittiva e inferenza statistica

Come sottolineato, il primo tema di cui si occupa la Statistica è la raccolta delle manifestazioni dei fenomeni di interesse. La rilevazione può riguardare tutte le possibili manifestazioni del fenomeno oggetto di analisi (*rilevazioni censuarie*) o soltanto di un insieme parziale di sue manifestazioni (*rilevazioni campionarie*).

Da segnalare che non sono affatto infrequenti le situazioni, soprattutto nelle scienze osservative, nelle quali lo statistico non interviene nella fase di raccolta dei dati; molto più diffuso è, per contro, l'intervento dello statistico nelle scienze sperimentali essendo stato sviluppato in questo contesto un insieme di metodologie, usualmente denominate *disegno degli esperimenti* o *piano degli esperimenti*, che consentono il perseguimento ottimale delle finalità di interesse.

La trasformazione dei dati in informazioni può essere perseguita attraverso:

- a. l'esecuzione di rappresentazioni tabellari e grafiche che mettano in evidenza certi aspetti generali del fenomeno o dei fenomeni oggetto di analisi (*rappresentazioni tabellari e grafiche*);
- b. il calcolo di indici che mettano in evidenza quello che c'è di tipico nelle manifestazioni dei fenomeni (*rappresentazioni sintetiche: valori medi*);

- c. il calcolo di indici che mettano in evidenza quello che c'è di mutabile e/o variabile nelle manifestazioni di fenomeni (*rappresentazioni sintetiche: indici di variabilità e/o mutabilità*);
- d. la misura della relazione tra fenomeni diversi ma logicamente collegati mediante appropriati indici (*rappresentazioni sintetiche: misure di associazione*).

L'operazione di trasformazione dei dati in informazioni, che si risolve sostanzialmente in una compattazione dei dati, costituisce l'oggetto di quella che usualmente viene denominata *statistica descrittiva* (cfr. Fig. 1). Si tratta di operazioni che possono essere applicate ai dati disponibili sia nel caso in cui i dati costituiscono l'insieme di tutte le possibili manifestazioni del fenomeno oggetto di analisi (*popolazione*) sia che si tratti di un insieme parziale (*campione*) di manifestazioni, ovviamente le conclusioni cui si perviene hanno una valenza limitata all'insieme dei dati analizzati.

Statistica descrittiva

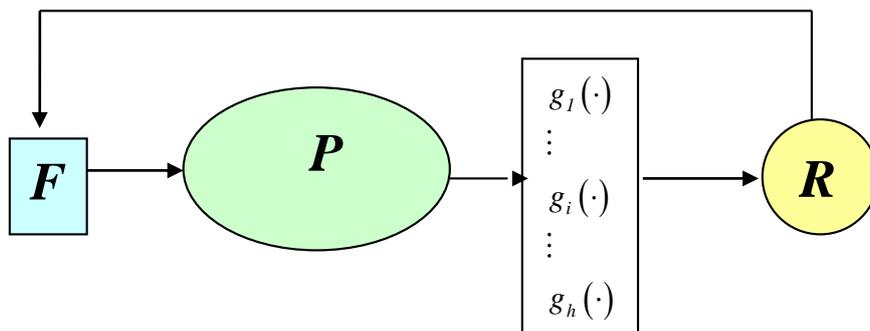


Fig. 1 - Rappresentazione schematica della struttura logica della statistica descrittiva

Legenda - *F* rappresenta il fenomeno analizzato, *P* l'insieme di tutte le possibili manifestazioni (popolazione) del fenomeno, $g_i(\cdot)$ per $i = 1, 2, \dots, h$ l'operazione di trasformazione (compattazione) dei dati in informazione e *R* l'insieme delle rappresentazioni sintetiche (valori medi, indici di variabilità, ecc.).

Nel caso di dati campionari, l'estensione delle conclusioni cui si perviene relativamente al campione all'intera popolazione costituisce per contro l'oggetto dell'*inferenza statistica classica (frequentista)* (cfr. Fig. 2).

Inferenza statistica classica

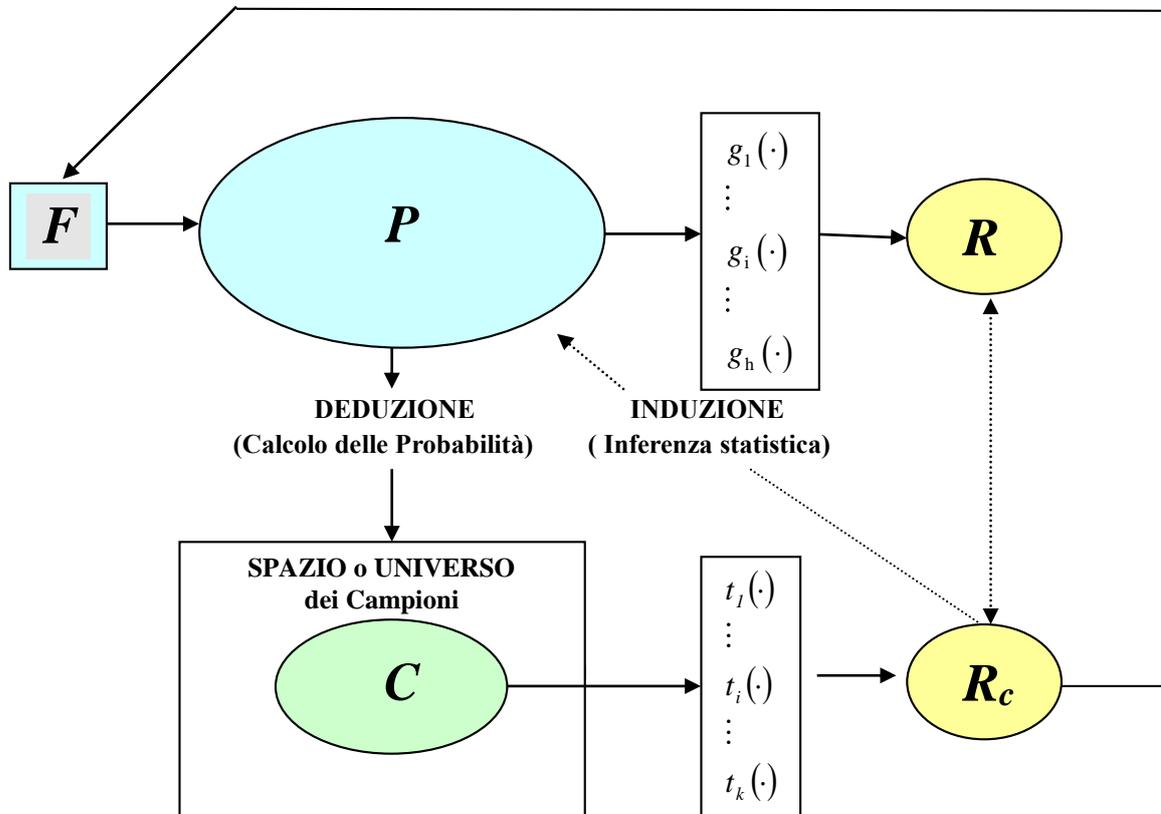


Fig. 2 - Rappresentazione grafica del processo di induzione statistica (inferenza classica)

Legenda - F , P , $g_i(\cdot)$ e R hanno il significato specificato nella Fig. 1, C rappresenta il campione delle manifestazioni del fenomeno, $t_i(\cdot)$ per $i = 1, 2, \dots, k$ l'operazione di trasformazione (compattazione) dei dati campionari in informazione e R_c l'insieme delle rappresentazioni sintetiche delle manifestazioni campionarie.

Un diverso modo per risolvere i problemi di induzione statistica è quello usualmente noto come **approccio bayesiano** all'inferenza statistica (cfr. Fig. 3). L'elemento caratterizzante di questo approccio è l'utilizzo esplicito e formalizzato del bagaglio conoscitivo pregresso (**informazioni a priori**) sul fenomeno d'interesse. Si deve, però, sottolineare che l'elemento distintivo del paradigma bayesiano non risiede tanto nell'utilizzo delle informazioni a priori (queste infatti vengono utilizzate anche nell'approccio classico all'inferenza statistica ad esempio specificando analiticamente il processo generatore dei dati), quanto nel diverso modo attraverso il quale le informazioni a priori vengono utilizzate.

Inferenza statistica bayesiana

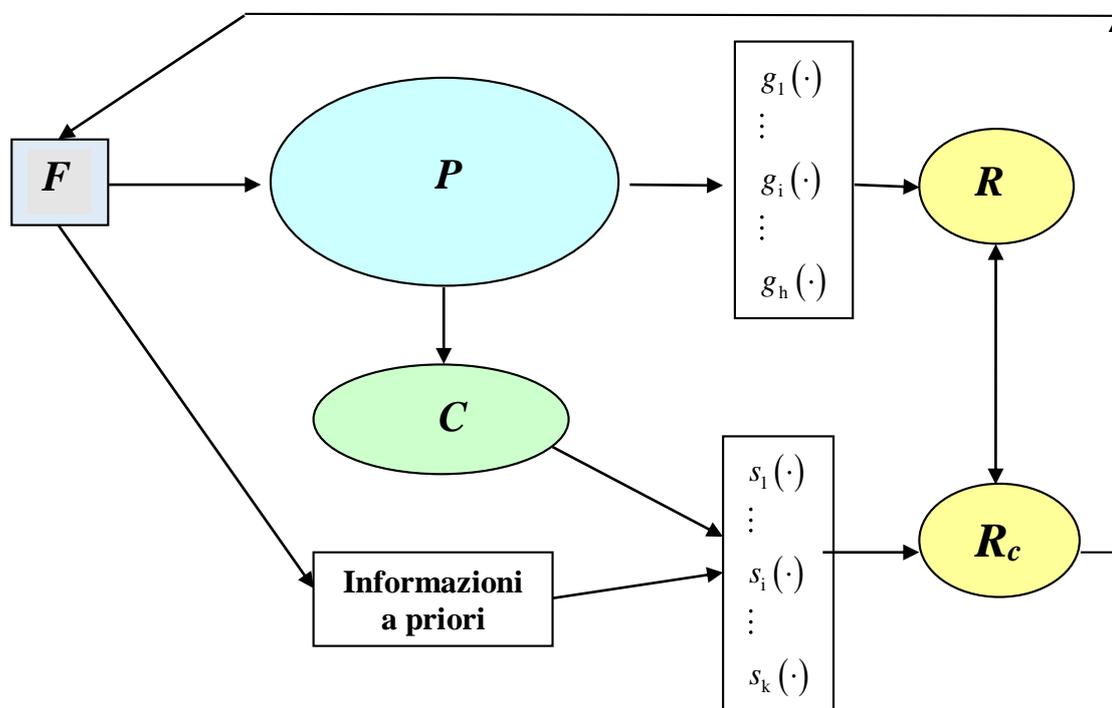


Fig. 3 - Rappresentazione grafica del processo di induzione statistica (inferenza bayesiana)

Legenda - F , P , $g_i(\cdot)$ e R hanno il significato specificato nella Fig. 1, C il significato specificato nella Fig. 2, $s_i(\cdot)$ per $i = 1, 2, \dots, k$ l'operazione di compattazione dei dati campionari e delle conoscenze pregresse (informazioni a priori), R_c l'insieme delle rappresentazioni sintetiche delle manifestazioni campionarie combinate alle informazioni a priori.

2. Modelli probabilistici e modelli statistici

La complessa natura dei fenomeni biologici, fisici, ingegneristici, economici, ecc. oltre a giustificare l'esistenza di metodi statistici sempre più raffinati, esige spesso l'introduzione di modelli analitici che, semplificando e schematizzando la realtà, ne rendono più facile la comprensione.

Poiché una serie innumerevole di fenomeni nelle varie scienze è governata da leggi aventi natura aleatoria, ne consegue immediatamente che i **modelli probabilistici** risultano essere i più appropriati per descrivere le loro manifestazioni (cfr. Fig. 4).

Una diversa categoria di modelli è rappresentata **modelli statistici**. Mentre i modelli probabilistici, quali rappresentazioni, nella generalità dei casi, approssimate della realtà, consentono una più immediata comprensione degli aspetti più rilevanti relativi ai fenomeni di interesse, i modelli statistici, oltre ad essere rappresentazioni semplificata della realtà, facilitano anche la **formulazione di previsioni** e consentono, come già

sottolineato, attraverso la ricerca di eventuali *nessi causali*, l'acquisizione di informazioni utilizzabili per l'attivazione di interventi in grado di modificare, a proprio vantaggio, l'evoluzione naturale dei fenomeni.

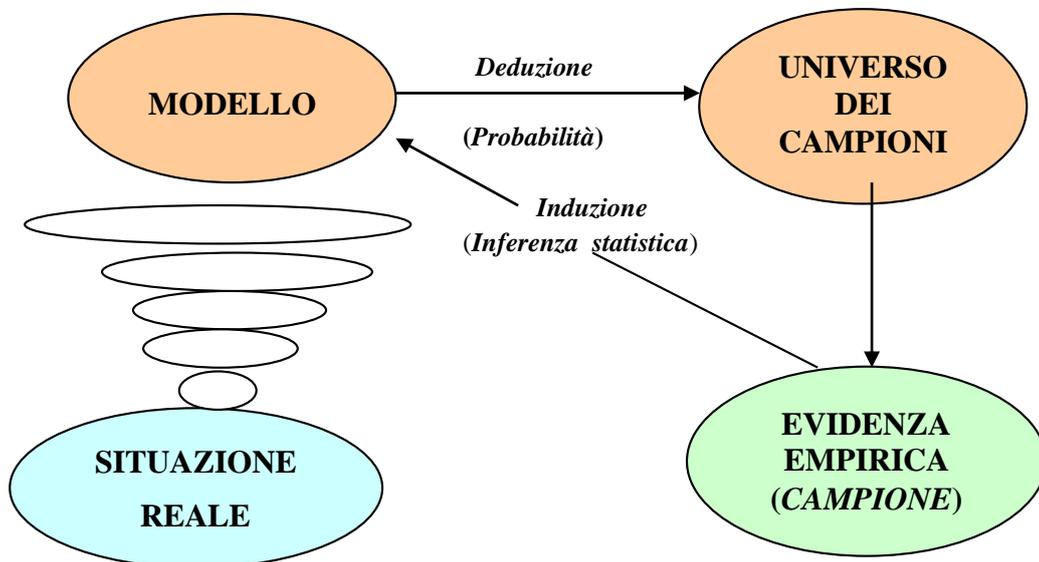


Fig. 4 - Relazioni tra modello probabilistico, statistica induttiva e evidenza empirica

Anche se fra i criteri logici od empirici che possono guidare nella individuazione del modello probabilistico e/o statistico ritenuti adeguati a rappresentare la realtà in esame, alcuni hanno validità generale, nella generalità dei casi i criteri sono strettamente connessi alla natura del fenomeno analizzato e agli obiettivi che s'intendono perseguire. Al riguardo si deve, inoltre, sottolineare che quando si procede all'introduzione di un modello probabilistico per esprimere le manifestazioni di un fenomeno, non ha più senso parlare di popolazione in quanto la popolazione è rappresentata dal modello stesso quale espressione del *processo di generazione dei dati*, e le manifestazioni del fenomeno o dei fenomeni esaminati hanno, necessariamente, sempre natura di manifestazioni campionarie essendo la popolazione rappresentata analiticamente attraverso il modello stesso (*super-popolazione*).

Riguardo l'impiego dei modelli (probabilistici e/o statistici) non ha senso chiedersi se un modello è vero o falso poiché, a stretto rigore, ogni modello è falso in quanto non coincide con la realtà. Ha senso, invece chiedersi se un modello è utile tenendo presente che le conclusioni cui si perviene sono strettamente condizionate dalla sua capacità di rappresentazione semplificata della realtà che si vuole interpretare.

In tale prospettiva assume rilevanza fondamentale la formula di Bayes sia come formula in grado di rappresentare in modo più che soddisfacente il *processo di apprendimento dell'esperienza* sia, e soprattutto, quando la si interpreta in termini causali.

Sulle manifestazioni dei fenomeni possono agire cause diverse che dipendono dalla situazione operativa usualmente denominata *stato di natura*, sullo stato di natura si posseggono informazioni a priori di natura aleatoria: *probabilità a priori delle cause*. L'acquisizione di informazioni (manifestazioni del fenomeno analizzato) consentono l'aggiornamento della probabilità a priori pervenendo alla misura delle *probabilità a posteriori delle cause*².

Lo schema teorico di riferimento afferma che possono agire più cause e che sulla scorta di risultati sperimentali (*informazione campionaria*) si procede all'aggiornamento della conoscenza relativamente al modello, considerato preesistente, posto a base dell'analisi.

3. Formalizzazione della rappresentazione tramite modelli

La rappresentazione analitica della realtà tramite modelli probabilistici e statistici è logicamente giustificata soltanto se si ritiene che la realtà d'interesse sia rappresentabile attraverso i modelli stessi, se si ritiene, cioè, che il fenomeno o i fenomeni, e le loro relazioni, siano governati da leggi esprimibili analiticamente. Al riguardo si deve, comunque osservare che alcuni modelli (probabilistici) non costituiscono una rappresentazione semplificata della realtà ma derivano dalla traduzione in termini analitici del processo generatore dei dati noto a priori.

La rappresentazione analitica dei fenomeni attraverso modelli probabilistici si risolve, nel caso in cui si considera un solo carattere quantitativo³, nella introduzione di un simbolo x al quale è associata una funzione di massa di probabilità, nel caso discreto, di densità di probabilità, nel caso continuo:

$$f(x; \theta) \quad \text{per } x \in S \text{ e } \theta \in \Theta^4$$

dove:

- x rappresenta le manifestazioni del fenomeno che si sta analizzando;
- S rappresenta lo *spazio campionario* di definizione di x , cioè lo spazio (*supporto*) di tutti i valori assumibili dall'entità variabile x ;
- θ rappresenta il parametro o i *parametri caratteristici* del modello. Parametri che nell'inferenza statistica classica vengono supposte costanti incognite mentre nell'inferenza statistica bayesiana vengono assunte come entità aleatorie (variabili casuali);
- Θ rappresenta lo *spazio parametrico*, cioè lo spazio di tutti i valori assumibili dal parametro o parametri che caratterizzano il modello.

Quando si ritiene opportuna l'introduzione di un modello analitico per rappresentare le manifestazioni di un fenomeno di interesse, si possono distinguere almeno due situazioni di mancanza di conoscenza: la prima situazione è quella caratterizzata da una

² L'interpretazione della formula di Bayes in termini causali verrà discussa successivamente.

³ Considerazioni analoghe valgono nel caso in cui si considerano più caratteri.

⁴ Il carattere in grassetto sta ad indicare che il simbolo utilizzato può rappresentare sia un singolo valore sia un vettore sia una matrice di valori.

conoscenza parziale della funzione $f(x; \theta)$ nel senso che si conosce (conoscenza a priori) la sua forma analitica ma non si conosce il valore di tutti o di alcuni dei parametri che caratterizzano la funzione stessa, in questa circostanza si parla di **inferenza statistica parametrica (classica o bayesiana)**. La seconda situazione è quella di mancata conoscenza della forma analitica del modello, in questa circostanza si parla di **inferenza statistica non parametrica (classica o bayesiana)**. Una terza situazione, intermedia rispetto alle due precedenti, è quella in cui si specificano solo certe componenti del modello. Se si opera in tale contesto si parla di **inferenza statistica semi-parametrica (classica o bayesiana)**, nel senso che la forma analitica del modello è specificata solo parzialmente.

Nel **contesto classico**, dato il modello

$$f(x; \theta) \quad \text{per } x \in S \text{ e } \theta \in \Theta$$

- nel caso parametrico l'unica incognita è rappresentata dal parametro o parametri θ che lo caratterizzano e i dati campionari vengono utilizzati per trarre conclusioni su tale entità;
- nel caso non parametrico o semi-parametrico sono incogniti, totalmente o parzialmente, sia il parametro o parametri θ che caratterizzano il modello, sia la sua forma analitica $f(\cdot, \cdot)$ e i dati campionari vengono utilizzati per trarre conclusioni sia sulla forma analitiche che sui parametri.

In altri termini, se a titolo esemplificativo si considera il caso parametrico, specificato il modello probabilistico $f(x; \theta)$ rappresentativo del fenomeno di interesse, le sue manifestazioni x (campione di osservazioni) vengono utilizzate per.

- la misura del grado di rappresentatività del modello;
- la stima puntuale e/o di intervallo dei parametri che lo caratterizzano;
- la verifica di ipotesi statistiche sulla forma e/o sul valore dei parametri.

I metodi sviluppati nell'ambito dell'inferenza statistica classica, a ragione di una supposta pseudo-oggettività, trascurano numerose informazioni, di natura sia oggettiva che soggettiva, disponibili a priori. Si tratta di informazioni che, oltre a risultare fondamentali per la comprensione del fenomeno oggetto di analisi, consentono, se impiegate in modo adeguato, di trasformare le informazioni stesse in conoscenza utilizzabile a fini di intervento; obiettivo questo conseguibile attraverso l'impiego della formula di Bayes soprattutto quando la formula stessa viene interpretata in termini causali.

Nel **contesto bayesiano**, la prospettiva dell'analisi si modifica sostanzialmente; infatti, in tale contesto, i parametri θ non vengono più considerati costanti incognite ma entità aleatorie con una propria distribuzione di **probabilità**, nota **a priori**, $\pi(\theta)$ e i dati campionari vengono utilizzati per procedere all'aggiornamento della conoscenza disponibile a priori che si trasforma, dopo l'acquisizione delle informazioni campionarie x , nella **probabilità a posteriori** $\pi(\theta/x)$.

Nonostante la diversità del fondamento logico dell'analisi bayesiana dei dati rispetto all'analisi classica (frequentista) è ancora possibile fornire una risposta ai problemi di stima puntuale, stima di intervallo e di verisimiglianza di ipotesi relativi ai i parametri θ ⁵.

4. Bayes come formula della probabilità delle cause

Se si ritiene che un evento E possa essere generato da una delle possibili cause, necessarie e incompatibili, E_i ($i=1, 2, \dots, j, \dots, k$) si ha:

$$\pi(E_j / E) = \frac{\pi(E_j) f(E / E_j)}{P(E)} = \frac{\pi(E_j) f(E / E_j)}{\sum_{i=1}^k \pi(E_i) f(E / E_i)}$$

dove; $\pi(E_j)$ rappresenta la **probabilità a priori** della causa E_j , $f(E / E_j)$ la probabilità probativa (verosimiglianza nell'ambito dell'inferenza statistica) dell'evento E mentre $\pi(E_j / E)$ rappresenta **probabilità a posteriori** della stessa causa E_j

Nella Fig. 5 è proposta una rappresentazione grafica di un processo dove l'evento E può essere generato da una delle possibili 5 cause (incompatibili) E_j ($j=1, 2, 3, 4, 5$)

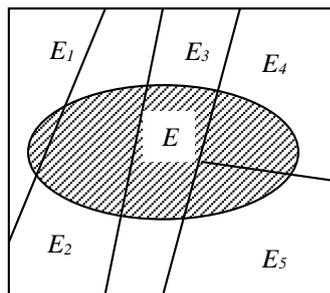


Fig. 5 – Rappresentazione grafica della formula di Bayes

La formula di Bayes, come più volte sottolineato, esprime in maniera molto semplice il processo di apprendimento dall'esperienza in contesti non deterministici.

Della realtà si possiede una conoscenza incerta, che viene espressa in termini di probabilità a priori, $\pi(E_j)$ queste probabilità si trasformano, al verificarsi dell'evento E (acquisizione di ulteriori informazioni), nelle probabilità a posteriori $\pi(E_j / E)$. Le probabilità condizionate si usano, quindi, per riassegnare le probabilità agli eventi una volta che sono state acquisite ulteriori informazioni (manifestazioni del fenomeno di interesse).

Inoltre, la formula di Bayes interpretata come formula della probabilità delle cause costituisce l'elemento fondamentale di riferimento della Teoria statistica delle decisioni;

⁵ Una trattazione molto approfondita dell'argomento è riportata in Chiandotto 2017, paragrafo 6.4.

in tale contesto θ rappresenta lo stato di natura, causa delle manifestazioni x del fenomeno analizzato:

$$\pi(\theta / x) = \frac{\pi(\theta) f(x / \theta)}{f(x)} = \frac{\pi(\theta) f(x / \theta)}{E_{\theta}[\pi(\theta) f(x / \theta)]}$$

dove

$$f(x) = E_{\theta}[\pi(\theta) f(x / \theta)] = \left\{ \begin{array}{ll} \sum_{j=1}^s f(x / \theta_j) \pi(\theta_j) & \text{nel discreto} \\ \int_{\theta} f(x / \theta) \pi(\theta) & \text{nel continuo} \end{array} \right\}$$

Se l'evoluzione naturale del fenomeno analizzato risulta sfavorevole si procede all'impiego di strumenti metodologici in grado di modificarne il percorso; si tratta, sostanzialmente, di operare la scelta di una specifica azione a capace di modificare la distribuzione delle probabilità $\pi(\theta)$ o $\pi(\theta / x)$ sugli stati di natura, cioè, di intervenire sulle cause che determinano gli esiti dei processi generatori dei dati misurandone le conseguenze.

Il contesto di riferimento è la *Teoria statistica causale delle decisioni* o *Teoria causale delle decisioni statistiche*; in tale contesto, se gli interventi si rivelano non efficaci o controproducenti; si procede all'attivazione di interventi correttivi⁶.

1. Rappresentazione grafica formale del processo evolutivo della Statistica

Nella Fig. 6 è riportata una rappresentazione grafica del processo evolutivo della Statistica sopra delineato, nelle figure successive si procede ad una chiarificazione del processo stesso.

⁶ Il lettore interessato all'utilizzo della metodologia statistica, nelle sue varie articolazioni, all'analisi di un fenomeno reale relativamente complesso, può utilmente consultare la monografia: "Il canone Rai TV 1954-2016", *Chiandotto*, 2020

Il Rapporto si caratterizza per l'impiego di svariati strumenti statistici di analisi dei dati, dai più semplici propri della statistica descrittiva (tabelle, grafici e indici sintetici), a quelli relativamente più avanzati (modello logistico di previsione e modelli a equazioni strutturali per variabili osservabili e variabili latenti in prospettiva decisionale).

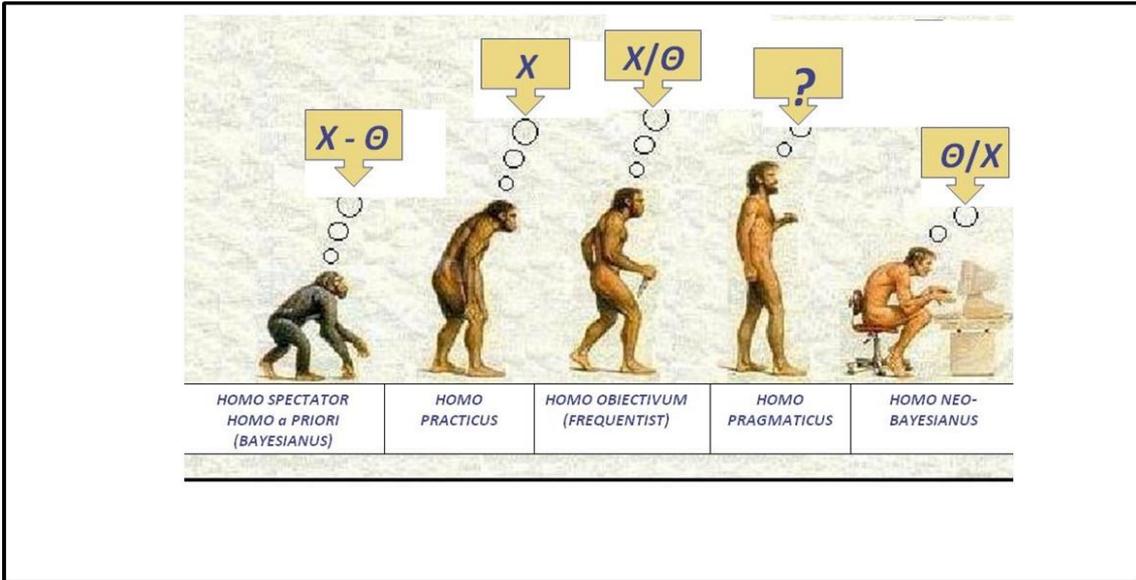


Fig. 6 – Processo evolutivo della Statistica

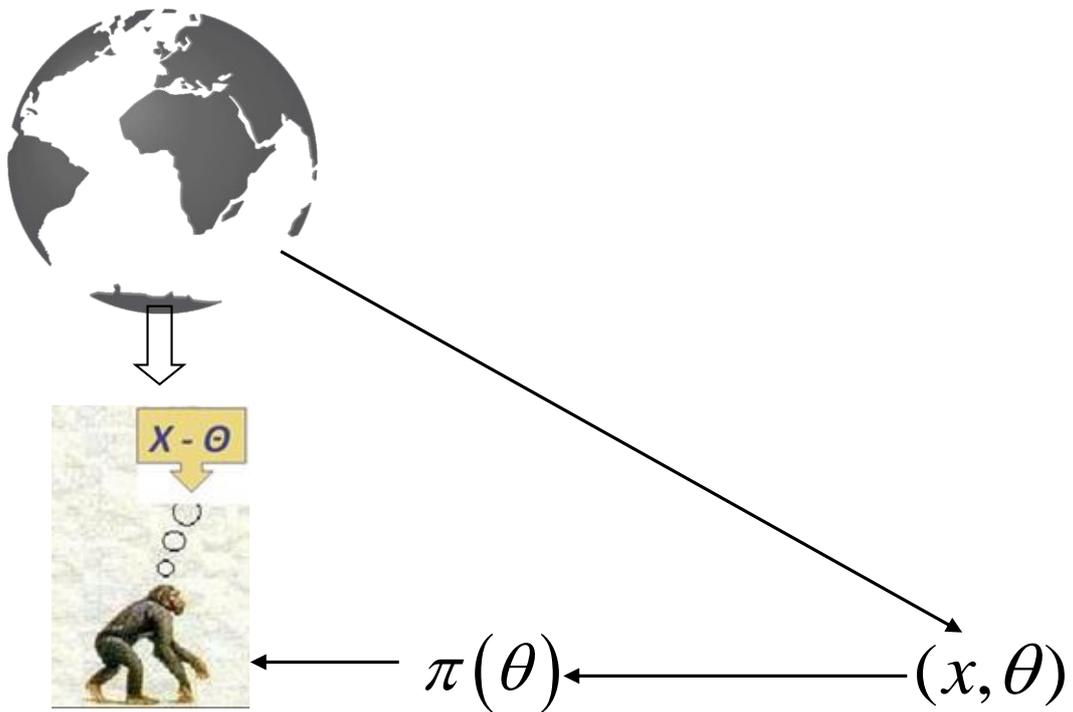


Fig. 7 – HOMO a PRIORI: Paradigma bayesiano

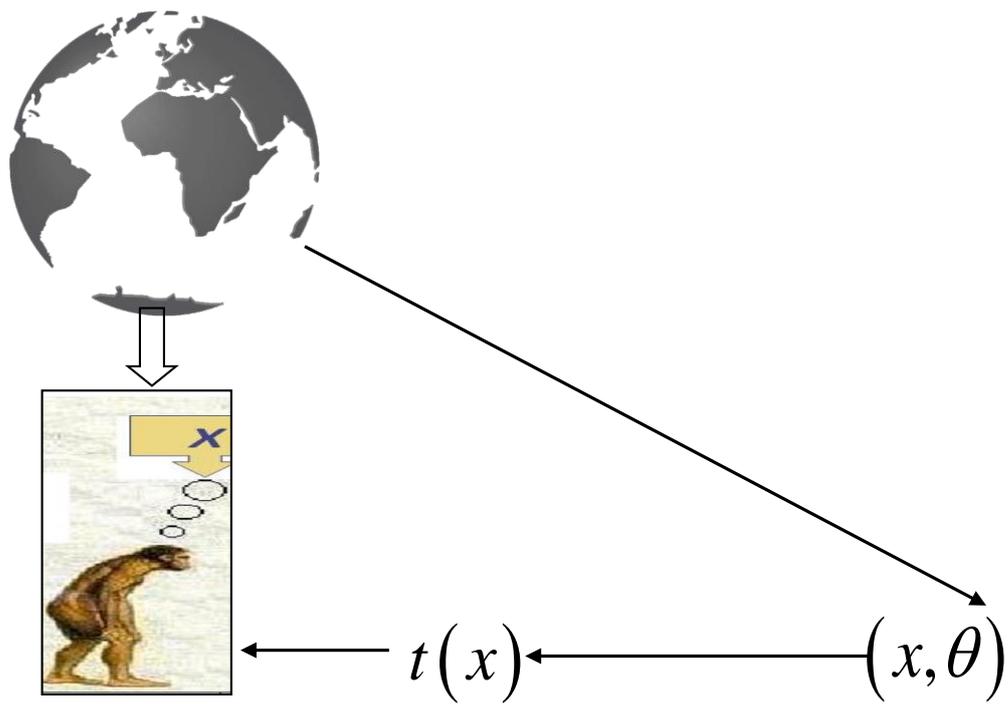


Fig. 8 – HOMO PRACTICUS: Statistica descrittiva

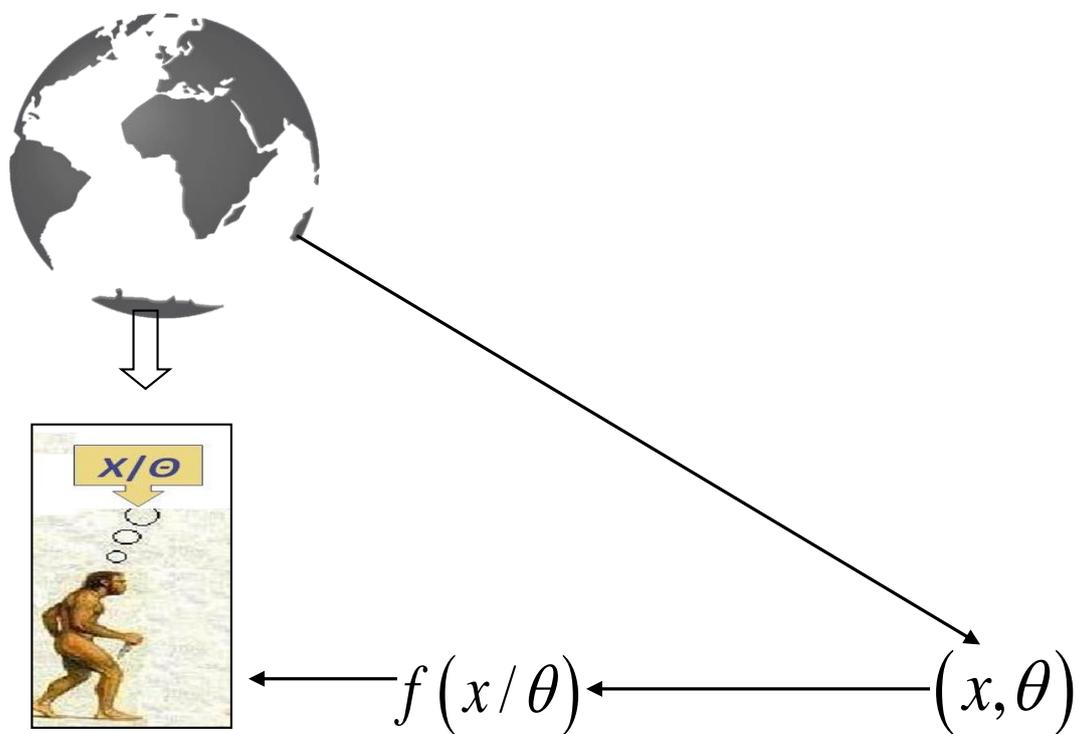


Fig. 9 – HOMO OBIECTIVUM: Inferenza statistica classica

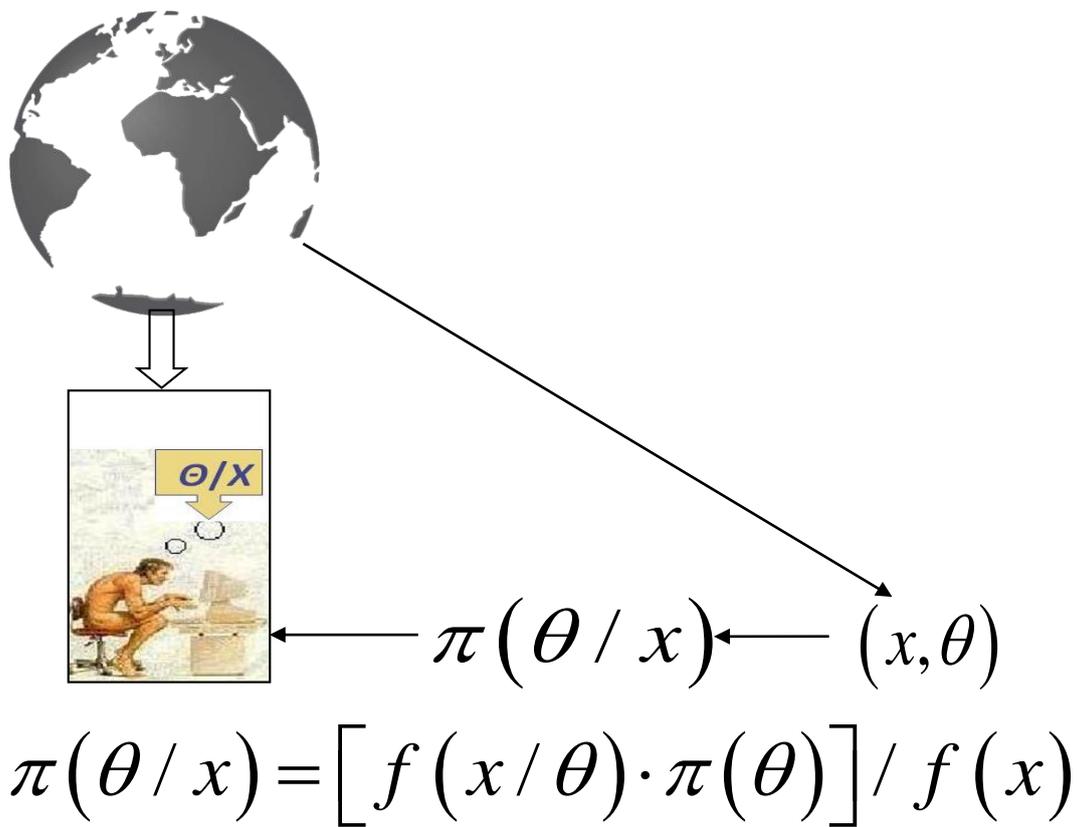


Fig. 10 – HOMO NEOBAYESIANUS: Inferenza statistica bayesiana

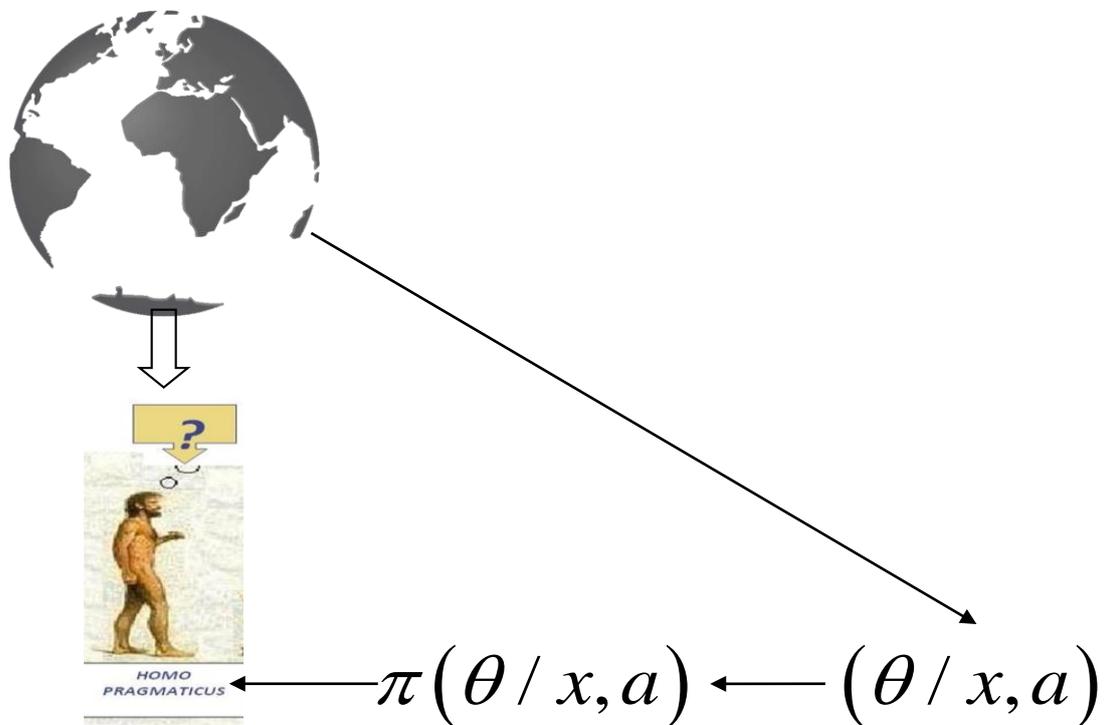


Fig.11 - HOMO PRAGMATICUS: Teoria statistica causale delle decisioni

Bibliografia

***B. Chiandotto (2017). Inferenza Statistica, Unifi-DISIA.
<https://local.disia.unifi.it/chiandotto>***

S. Bacci e B. Chiandotto (2019). Introduction to Statistical Decision Theory, Utility Theory and Causal Analysis, Chapman and Hall/CRC.

***B. Chiandotto (2020). Teoria Statistica delle Decisioni, Unifi-DISIA.
<https://local.disia.unifi.it/chiandotto>***

***B. Chiandotto (2020). Il canone Rai TV 1954-2016..
<https://local.disia.unifi.it/chiandotto>***
