Introduzione alla statistica bayesiana

Statistica bayesiana e probabilità soggettiva, distribuzione a priori, approcci attuali (parte 3)

Docente: Matilde Trevisani

DEAMS

A.A. 2025/2026
(aggiornato: 2025-10-02)

Agenda

Introduzione alla Statistica Bayesiana

- Probabilità soggettiva
- Distribuzioni a priori
- Approcci odierni
- Note finali

Probabilità soggettiva

XX century

La statistica Bayesiana fu riscoperta nel XX-esimo secolo

per **nuove domande**

- decifrare il codice *Enigma* durante la seconda guerra mondiale (see *Edward Simpson: Bayes at Bletchly Park*, http://mathcenter.oxford.emory.edu/site/math117/bayesTheorem/)

- combinare informazioni storiche e correnti nella determinazione delle tariffe assicurative (la tecnica attuariale nota come teoria della credibilità risulta essere basata sul ragionamento bayesiano)

- stimare la probabilità di eventi come la

    - probabilità di un incidente aereo che coinvolga due aerei (negli anni '50)
    - probabilità di un'esplosione accidentale di una bomba H

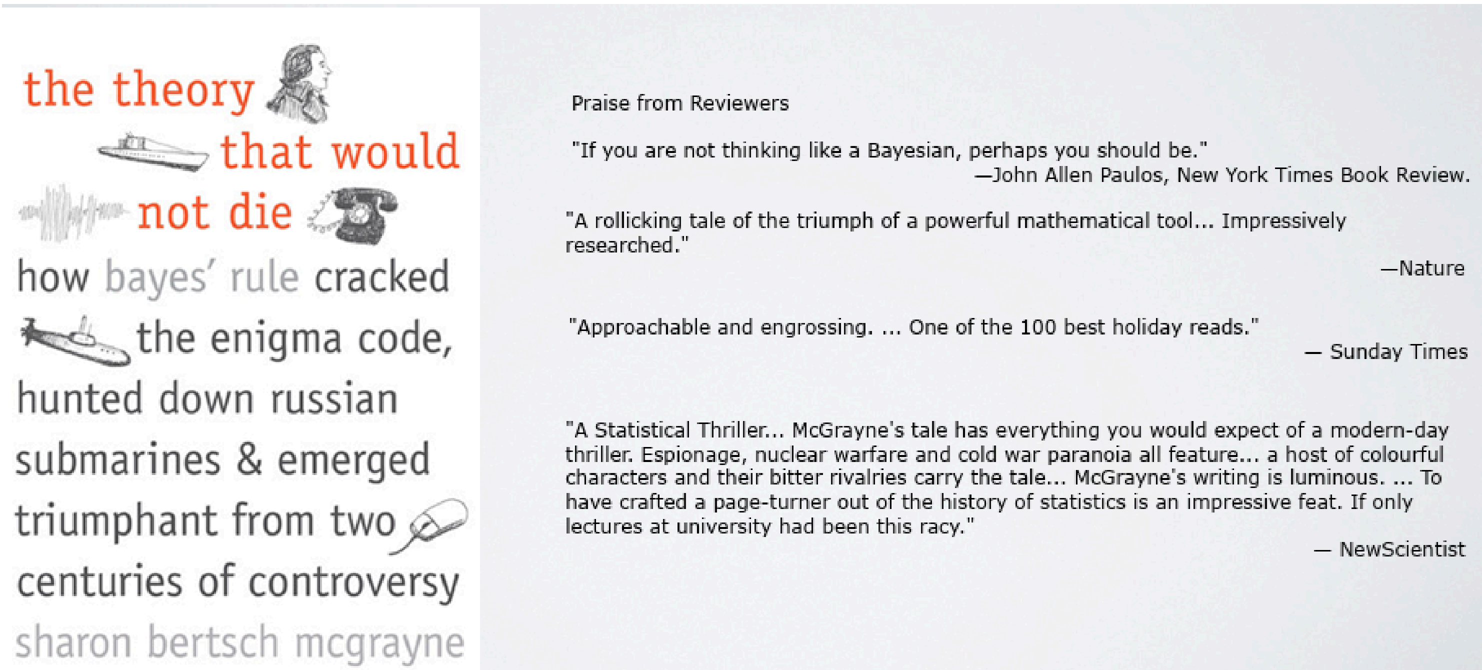
thanks to the **availability of computers**

- Bayesian analytical results are available only for simple problems
- Bayesian computation is rather intensive (Monte Carlo methods are fundamental)

Bayes' Rule is just conditional probability
So why is it such a big deal?

McGrayne (2011)

> A vivid account of the generations-long dispute over Bayes' rule, one of the greatest breakthroughs in the history of applied mathematics and statistics



Dipende dalla domanda e dalle informazioni

Behind the choice of the preferred statistical approach, frequentist or Bayesian, there might be the question which is asked and the information available.

E.g., Fisher (1890-1932), che lavorava nel campo della genetica, effettivamente eseguiva degli esperimenti dove

- non c'è bisogno di un'informazione a priori
- risulta facile inquadrare l'interpretazione nel paradigma del campionamento ripetuto

Molte applicazioni nel XX secolo riguardavano il bisogno/desiderio di

-> **combinare diverse fonti di informazione**
-> **valutare la probabilità di eventi che non sono mai stati osservati**

Problems with probability

Although employed in special contexts in XX century, Bayesian Statistics did not achieve general acceptance (far from it).

One of the reasons why Bayesian statistics was difficult to accept is related to the frequentist definition of probability.

It is conceptually difficult to frame a distribution on the model (parameter) as a frequentist probability.

We need another probability!

Let us take a step back and discuss about this.

Limiti dell'interpretazione frequentista della probabilità

(Richiamo)
La probabilità di un evento si definisce come limite della frequenza relativa del verificarsi dell'evento in un numero di ripetizioni.

Questa idea fu sviluppata e meglio definita da Jakob Bernoulli (*Ars Conjectandi*, 1713) e Abraham De Moivre (1733) nella **legge dei grandi numeri** che collega teoricamente la probabilità di un evento alla frequenza relativa in ripetizioni infinite.

Teorema (Legge (forte) dei grandi numeri (*Strong Law of Large Numbers*))

Sia $E_1, \ldots, E_n, \ldots$ una sequenza di eventi indipendenti tali che $P(E_i) = p \,\forall\, i$. Sia $S_n = \sum_{i=1}^{n} |E_i|$ il numero di eventi che si verificano nei primi n. Allora `

$$P\left(lim_{n \to \infty} \frac{S_n}{n} = p\right) = 1$$

La definizione frequentista di probabilità si applica a un insieme ristretto di eventi, quelli che possono essere incorporati, almeno idealmente, in una sequenza di ripetizioni.

Principio frequentista non è applicabile

However we can easily think of "events" for which it does not work:
(each one for a different reason)

- vincere alla lotteria alla prossima tornata (probabilità 0?)
- Il pallavolo femminile italiano vince il prossimo campionato mondiale (non applicabile rigorosamente)
- WWIII tra USA and Russia scoppierà nel prossimo anno (evento *one-off*);
- la moneta mostri testa quando sappiamo che la moneta è a doppia testa o a doppia croce, ma non sappiamo quale (un'ipotesi o è vera o è falsa)
- il numero di cittadini extra-europei in Italia il 1/1/2025 (un numero fisso, non casuale)

The last two cases neither describe an event (a possible outcome of a random experiment).

The model (and any *gear* of it) is a random thing

In statistical applications, there are situations analogous to the two-headed/two-tailed coin example.

Suppose you want to estimate *the number N of non UE citizens in Italy on 1/1/2025*, one may object that

- N is not a random quantity (it is intrinsically a fixed number, albeit unknown);
- I cannot specify a probability distribution on a non random quantity!

The frequentist definition does not help in interpreting a probability distribution on N.

In the Bayesian approach any parameter is random: N is then a random number in the Bayesian setting

Definizione assiomatica di probabilità

Possiamo liberarci dall'interpretazione frequentista considerando la **definizione assiomatica** di probabilità.

Definizione (probabilità)

La probabilità è una misura su un insieme di eventi (risultati) tale che

- non è negativa
- è additiva rispetto a eventi che si escludono a vicenda
- somma a 1 su tutti i possibili risultati che si escludono a vicenda

This does tell us nothing about how it is measured and what can be used for.

We will all agree that we can use it to describe limiting relative frequencies of occurrence of events in repeated sequences, we may not agree on whether we can use it for something else.

Do we need another probability?

First, should we use it for something else?

We may take the stance that only events for which a sequence of ideal repetition is thinkable are permitted.

This is unsatisfying intuitively and practically since we have to deal with more general kinds of uncertainty (and they are relevant, think the H-bomb accidents) and we do routinely deal with them, that is we do take decisions based on some evaluation of such uncertain (non repeatable) events (think betting or weather forecasts).

Still, we might say that this kind of events is dealt with by **common sense** and is out of scope for a formal treatment by probability, but it might also be the case that probability could describe how common sense works.

Common sense: deductive logic $\rightarrow$ plausible logic

An example of common sense is an inference like

$$\left.\begin{array}{c}\text{if } A \text{ then } B\\ A \text{ is true}\end{array}\right\} \Rightarrow B \text{ is true}$$

which is described by **deductive logic**.

We also do inferences like the following

$$\left.\begin{array}{c}\text{if } A \text{ then } B\\ B \text{ is true}\end{array}\right\} \Rightarrow A \text{ more plausible}, \qquad \left.\begin{array}{c}\text{if } A \text{ then } B\\ A \text{ is false}\end{array}\right\} \Rightarrow B \text{ less plausible}$$

or even

$$\left.\begin{array}{c}\text{if } A \text{ then } B \text{ is more plausible}\\ B \text{ is true}\end{array}\right\} \Rightarrow A \text{ is more plausible}$$

This is a common type of reasoning (**plausible logic**), even in everyday life.

It is sensible to try to describe it, that is, to quantify less/more plausible.

Probability as extension of true-false logic

If the aim is to represent the state of uncertainty on a "fact", then conditional probability is the only system which satisfies the following axioms

I. States of uncertainty are represented by real numbers.

II. Qualitative correspondence with common sense.

- If the truth value of a proposition increases, its probability must also increase.
- In the limit, small changes in propositions must yield small changes in probabilities.

III. Consistency with true-false logic.

- Probabilities that depend on multiple propositions cannot depend on the order in which they are presented.
- All known propositions must be used in reasoning -- nothing can be arbitrarily ignored.
- If, in two settings, the propositions known to be true are identical, the probabilities must be as well.

Si può provare che la logica del "buon senso" (logica deduttiva e logica della plausibilità) segue gli assiomi della probabilità.

Un'altra prova che la probabilità definita dagli assiomi è l'unico modo ragionevole per descrivere l'incertezza è l'argomento del ***Dutch book***.

Coerenza delle scommesse (recall)

Definiamo la probabilità di un evento $P(E)$ come

- il prezzo che pagheresti in cambio di un ritorno di 1 se l'evento si verifica e 0 altrimenti,
- il prezzo che accetteresti in cambio di dover pagare 1 se l'evento si verifica e 0 in caso contrario.

In altre parole, una volta dichiarato $P(E)$, compreresti o venderesti l'importo casuale $|E|$ in cambio di $P(E)$.

La probabilità è quindi il **prezzo equo** di un evento incerto.

Il valore atteso della scommessa è 0.

Supponi di valutare le probabilità per un insieme di eventi, quindi se le tue probabilità non soddisfano gli assiomi è possibile escogitare una combinazione di scommesse che portano a una perdita (guadagno) sicura (sicuro).

Probabilità per descrivere l'incertezza

Per alcuni, queste considerazioni rendono l'uso della probabilità per rappresentare l'incertezza una scelta obbligata e quindi il ragionamento bayesiano (che è una conseguenza della probabilità) l'unico modo ragionevole per aggiornare le informazioni (probabilità).

> Bayesian Statistics offers a rationalist theory of personalistic beliefs in contexts of uncertainty, with the central aim of characterising how an individual should act in order to avoid certain kinds of undesirable behavioural inconsistencies.
>
> (Bernardo and Smith)

Questo porta abbastanza naturalmente alla definizione soggettiva di probabilità.

Probabilità soggettiva

> Probabilities are states of mind and not states of nature.
> -- *Leonard J. Savage, 1954*

Si accetta che la probabilità non sia una proprietà oggettiva di un fenomeno ma piuttosto l'opinione di una persona e si definisce

Definizione (probabilità soggettiva)

> La probabilità di un evento è, per un individuo, il suo grado di convinzione sull'evento.

Bruno de Finetti (1906-1985 circa), probabilista e attuario italiano (per Generali), in *Theory of probability* (1970) scrive

> Probability does not exist

e propone la definizione soggettiva di probabilità e l'impianto della coerenza, basato sull'interpretazione della scommessa.

Se la probabilità è un grado soggettivo di convinzione, essa dipende dalle informazioni che sono soggettivamente disponibili, ed è anche chiaro che per ***casuale*** si intende ***non noto per mancanza di informazioni***.

Quindi, un parametro è casuale perché è non noto.
(Nell'approccio frequentista esiste un parametro "vero" generalmente incognito che si cerca di "stimare".)

- Frequentists define probability as the long-run frequency of a certain measurement or observation. The frequentist says that there is a single truth and our measurement samples noisy instances of this truth. The more data we collect, the better we can pinpoint the truth. The archetypal example is the repeated tossing of a coin to see what the long-run frequency is of heads or tails; this long-run frequency then asymptotes to the truth.

- Bayesians define probability as the plausibility of a **hypothesis** given incomplete knowledge. To a Bayesian there is no Platonic truth out there which we want to access through data collection (or perhaps we should say there may be a Platonic truth, but it will always remain outside our experience). For a Bayesian, there is just data which we can use as evidence for particular hypotheses. A Bayesian coin-tosser just observes a series of coin tosses and then uses this information to make deductions about, for example, how likely it is that the coin is fair.

The real-world coin-toss experiment lives in the Bayesian domain: the experiment is done under limited knowledge of the initial conditions and precise set-up. Given that limited knowledge, how do we use the experimental outcomes to evaluate a statement such as “the coin is fair”? This is a Bayesian question. Bayesian statistics is the statistics of the real world, not of its Platonic ideal.

Statistica bayesiana e probabilità soggettiva

La definizione soggettiva di probabilità è la più compatibile con il paradigma bayesiano, che afferma:

- il parametro da stimare è una quantità ben specificata ma non è noto per mancanza di informazioni
- una distribuzione di probabilità è (soggettivamente) specificata per il parametro da stimare, ed è chiamata ***distribuzione a priori***
- dopo aver osservato i risultati sperimentali, la distribuzione di probabilità sul parametro viene aggiornata utilizzando il teorema di Bayes che combina i risultati sperimentali (verosimiglianza) e la distribuzione a priori per ottenere la distribuzione a posteriori

Probabilità soggettiva e regola di aggiornamento bayesiana (che è effettivamente una conseguenza delle regole della probabilità) stabiliscono un sistema per descrivere l'inferenza in cui gli input sono le convinzioni a priori e i dati e l'output è costituito dai convincimenti aggiornati (a posteriori).

Uses of probability

- To describe variation
  - using aleatory or phenomenological or descriptive probability
  - analogous to saying probability of rolling a 3 with an apparently fair die is 1/6.
- To quantify uncertain knowledge
  - by making epistemic or inferential statements
  - analogous to saying you are 90% sure that Greenland is part of the Kingdom of Denmark.

Bayesian inference

- Founded on an inferential principle of equivalence: there is only **one kind of probability** for both inferential and descriptive purposes, with inferential statements being obtained from descriptive ones merely by applying the laws of probability.

Bayes soggettivo

Questo approccio, talvolta chiamato **Bayes soggettivo**, ebbe molti seguaci dagli anni '60 (vedi Lindley (1970), Lindley (2013), Savage (1972a)).

In effetti, per molti è diventato l'unico fondamento coerente della statistica, mentre le alternative (Fisher, Neymann-Pearson e simili) sembravano una raccolta di strumenti ad hoc privi di un'adeguata giustificazione.

Ciò ha portato alla formazione di due fazioni, ciascuna che rifiuta i metodi dell'altra; da parte degli anti bayesiani, le critiche si sono concentrate sul fatto che ammettendo una natura soggettiva delle conclusioni ciò le rendeva inutili dal punto di vista scientifico.

Anche se accettiamo che Bayes soggettivo sia una buona descrizione del ragionamento in condizioni di incertezza in senso lato, è ancora rilevante discutere se questo sia accettabile in un contesto scientifico: semplificando un po', il ruolo delle distribuzioni a priori è per questo fondamentale.

Distribuzione a priori

Necessità della a priori

Un punto critico nell'inferenza bayesiana (e uno dei motivi per cui non è stato accettato all'inizio) è la necessità dell'informazione a priori.

Mentre la statistica classica si occupa solo delle informazioni provenienti dai dati, la statistica bayesiana è una regola con cui aggiornare le informazioni in base ai dati: dobbiamo iniziare da qualche parte.

Cià è stato visto come un problema importante poiché introduce un elemento di soggettività nell'analisi.

Di questo si parlerà più avanti, facciamo ora due note preliminari in merito:

- da dove viene la priori;
- la soggettività (nel senso di arbitrarietà) dei risultati.

Source of prior information

Think of the female birth example again, but with the following sample:

> In 2010 in Muggia (small city near Triest) 38 males and 47 females were born.

According to likelihood inference (for θ, pr. of a female birth)

- the ML estimate is $\hat{\theta} = 0.553$
- the 95% c.i. is [0.441, 0.659]
- the p-value for $H0 : \theta \geq 0.5$ is 0.8

What do you think of this information?

You probably think something along the lines of "This sample tells me nothing".

Why is that? Well, because you have, in fact, prior information.

Di solito abbiamo informazioni a priori

E' raro modellare una situazione ove non **abbiamo informazioni a priori**.

Informazioni a priori possono provenire da

- conoscenza sostanziale del processo che genera i dati (potremmo non essere sicuri del meccanismo esatto, ma di solito sappiamo qualcosa),
- osservazioni fatte in passato.

Con questo in mente, la distribuzione a priori non dovrebbe apparire così strana.

Soggettività dei risultati: a priori che svaniscono

Le informazioni a priori possono non piacere: due persone con gli stessi dati possono giungere a conclusioni diverse perché iniziano da diverse informazioni a priori.

Anche se questo è vero, è anche vero che, se le informazioni a priori non sono irragionevoli, ***le conclusioni tendono*** (asintoticamente) ***a diventare uguali quanti più dati vengono raccolti***.

Discuteremo cosa significa "irragionevole", ma il requisito di base è che non escludiamo alcuna possibilità (assegnando una probabilità a priori nulla).

Inoltre, discuteremo come distinguere le distribuzioni a priori con rispetto a quanto pesano sulla conclusione (quanto sono informative): ***esistono metodi per garantire che le conclusioni siano meno influenzate dalle priori***.

Soggettività dei risultati: a priori standard

Quanto detto sopra presuppone che la priori possa (e debba) rappresentare le convinzioni prima delle osservazioni.

È anche possibile adottare un ***approccio diverso***, all'interno del paradigma bayesiano.

Nell'esempio della nascita femminile Laplace assunse una distribuzione a priori uniforme su θ: vide questo come un modo per esprimere indifferenza rispetto alle possibilità.

Questo è abbastanza ragionevole, sebbene problematico per alcuni aspetti, ma l'idea può essere resa più precisa.

Soggettività dei risultati: rendere irrilevante la a priori

L'idea è che non si richiede alla priori di esprimere informazioni, piuttosto essa è considerata ***come una componente tecnica del modello***.

Questa idea sta dietro alle cosiddette

- priori non informative (*non informative priors*)
- priori di riferimento (*reference priors*)

il cui nome dice tutto, anche se forse è troppo ottimistico:

- "informatività" non è un concetto ben definito (attenzione a dare un significato preciso all'idea intuitiva),
- la a posteriori dipende ancora dalla a priori

Con questi avvertimenti diciamo che si possono definire distribuzioni a priori particolari che evitano l'interpretazione soggettiva della distribuzione a priori.

Questo approccio è talvolta chiamato **Bayes oggettivo** (o Bayes automatico).

Riepilogo

- Calcolo delle probabilità
  problemi diretti <-> P(Dati|Modello)

- La prima inferenza statistica è "bayesiana" (Bayes, Laplace)
  probabilità delle cause <-> P(Modello|Dati)

- Approccio Bayesiano messo da parte (per ragioni filosofiche e tecniche)
  statistica classica <-> P(Dati|Modello) (Gosset, Fisher and many others)

  - likelihood
  - principio del campionamento ripetuto

- Interesse per statistica Bayesiana rinnovato (nuovi problemi, progressi computazionali)
  Ci sono problemi legati alla soggettività dei risultati

  - Bayes soggettivo viene proposto come paradigma convincente (de Finetti, Lindley, Savage)
  - Bayes oggettivo come compromesso tra stat classica e Bayesiana (Jeffreys)

Vedi un riepilogo sull'excursus storico più lungo in appendice. →

Approcci odierni

E oggi?

Ci manca ancora un'idea condivisa sul fondamento dell'inferenza statistica.

Questa non è solo una questione astratta, si sostiene che sia alla radice di problemi pratici nelle applicazioni della statistica: la questione dei test d'ipotesi nelle scienze applicate (vedi Nuzzo (2014); Goodman (2016), vedi anche Pauli (2018) per una panoramica della questione).

Di seguito vengono presentate alcune panoramiche moderne dello scenario sui fondamenti della statistica (Senn, Efron and Hastie, Royall).

Mappa degli approcci di Senn

Senn (2011) mappa i vari approcci a seconda che si concentrino su

- probabilità dirette o inverse da un lato;
- inferenza o decisione d'altra parte (qui decisione significa che siamo interessati alle conseguenze dell'utilizzo di un determinato criterio).

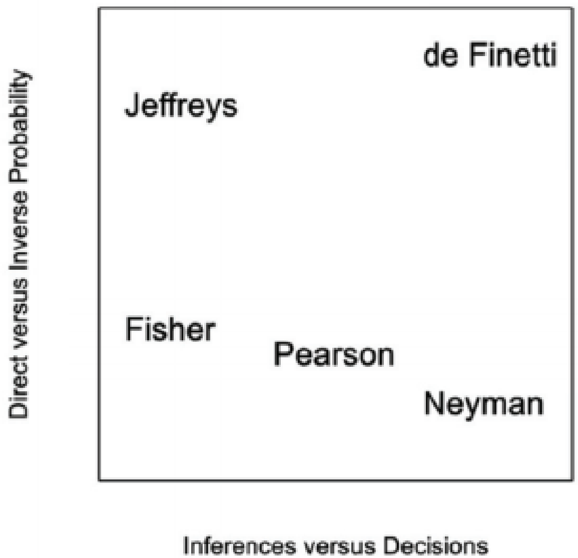
(Keep in mind that any scheme like this is bound to oversimplify.)

Statistica classica

- Verosimiglianza: Fisher
- Test d'ipotesi: Neyman-Pearson

Statistica bayesiana

- Oggettiva: Jeffrey
- Soggettiva: de Finetti



Mappa di Senn

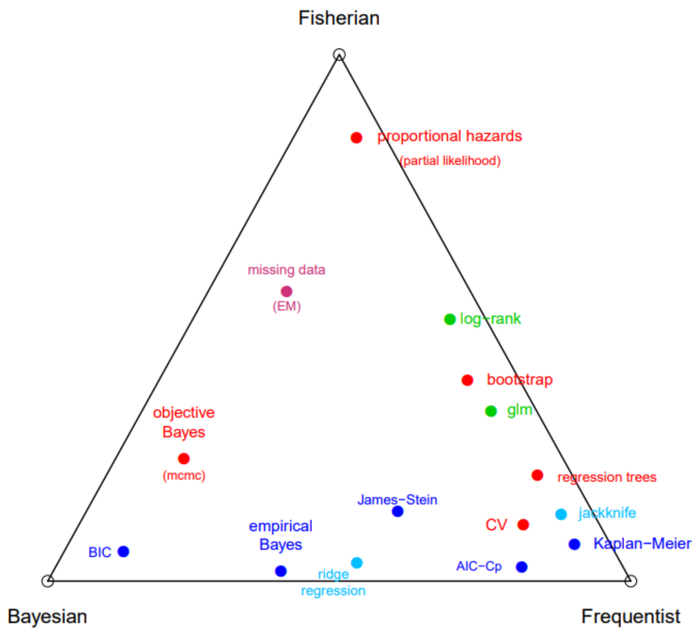
Map of techniques/approaches by Efron and Hastie

Efron and Hastie (2021) map some statistical techniques with respect to their guiding principles and the **relevance of the computational aspect**.

The triangle is not meant to give a complete picture: it shows a **selection of techniques**, as described in the text (*Early Computer-Age Methods*: 15 major topics, **1950s through 1990s**), and the Bayesian, frequentist, and Fisherian influences on them. It shows how these techniques are **based on a mixture of approaches**

Color indicates the **importance of electronic computation** in their development ->

It would not be easy to place some XXI century machine learning developments (a philosophically atheistic approach to statistical inference) here (of course they would all be red)



- red, crucial;
- violet, very important;
- green, important;
- light blue, less important;
- blue, negligible.

Some quotes from *Computer age statistical inference, student edition: algorithms, evidence, and data science*

> Statistical inference is an unusually wide-ranging discipline, located as it is at the triple-point of mathematics, empirical science, and philosophy. The discipline can be said to date from 1763, with the publication of Bayes' rule. The most recent quarter of this 250-year history—from the 1950s to the present—is the **"computer age"**, the time when computation, the traditional bottleneck of statistical applications, became faster …. The book is an examination of how statistics has evolved over the past sixty years.
>
> The role of electronic computation is central to our story. This doesn't mean that every advance was computer-related. [But] Almost all topics in twenty-first-century statistics are now computer-dependent.
>
> Very broadly speaking, algorithms are what statisticians do while inference says why they do them.
>
> A particularly energetic brand of the statistical enterprise has flourished in the new century, **data science**, emphasizing algorithmic thinking rather than its inferential justification.

Royall: Approcci diversi per domande diverse

Royall (2004) distingue i metodi in base alle domande a cui cercano di rispondere.

Tre domande possono essere poste ai dati

1. Cosa dovrei credere?
2. Come dovrei comportarmi?
3. Qual è l'evidenza?

La posizione di Royall è questa

La (3) è risposta unicamente dalla verosimiglianza,

La (1) è risposta dalla posteriori (necessita della verosimiglianza e della priori),

La (2) necessita della posteriori e dei costi degli errori.

Notare che la (2) è diversa dalla 'decisione' che Senn ha in mente.

Mix

La nota positiva è che le fazioni non ci sono più: almeno in una certa misura, gli statistici sono desiderosi di prendere ciò che è rilevante da ogni approccio.

> It seems quite clear that both Bayesian and frequentist methodology are here to stay, and that we should not expect either to disappear in the future. ... Philosophical unification of the Bayesian and frequentist positions is not likely, nor desirable, since each illuminates a different aspect of statistical inference.
>
> -- *Bayarri and Berger (2004)*

In pratica questo ha comportato che

- ora è ritenuto ragionevole da molti bayesiani valutare l'adeguatezza del modello (questo è incoerente con il considerare (posterior distribution on) models as beliefs, as beliefs non possono essere sbagliate per principio),
- vengono studiate le proprietà frequentiste delle procedure bayesiane. hybridization

> An assessment of strengths and weaknesses of the frequentist and Bayes systems of inference suggests that ... inferences under a particular mode! should be Bayesian, but model assessment can and should involve frequentist ideas (Little, 2006)

Today

- L'**atteggiamento** più comune tra gli statistici applicati è **pragmatico**: è ragionevole utilizzare l'approccio che è più adatto alla situazione in questione.

  > "Pure" Bayes, "pure" frequentist, "pure" any statistical philosophy, pairs nicely with Port, but when you leave port for the high seas of applications, some degree of impurity is usually necessary. Consequently, statisticians who engage in important studies use their paradigm as an aid to navigation, not as a straightjacket. The goal is to do a good job, and one can't be (too) doctrinaire
  >
  > -- *Tom Louis, 2019*

- È anche ragionevole interpretare le tecniche bayesiane come tecniche di modellazione piuttosto che come una posizione filosofica (scollegandola così dall'interpretazione soggettiva), in questo senso il ruolo della a priori può essere minimizzato, da fonte di informazione a **dispositivo di regolarizzazione** (parte di un modello). See Objective Bayes

- Qualunque sia l'atteggiamento che adotterete, tenete presente, tuttavia, che l'approccio bayesiano è **l'unico corretto** (su una base di ragionamento probabilistico) e tutte le altre procedure sono giustificate solo come approssimazioni di quelle bayesiane.

Ragionamento Bayesian oltre la statistica

Natura convincente del ragionamento bayesiano

Si ricordi:

> Bayesian Statistics offers a rationalist theory of personalistic beliefs in contexts of uncertainty, with the central aim of characterising how an individual should act in order to avoid certain kinds of undesirable behavioural inconsistencies (Bernardo and Smith).

Abbiamo notato che questo ha portato alcuni a sostenere di assumere il ragionamento bayesiano come il fondamento dell'inferenza statistica.

Abbiamo infatti affermato che il ragionamento bayesiano (in quanto ragionamento probabilistico) potrebbe essere il paradigma per estendere la logica deduttiva alla logica plausibile.

Role of Bayesian reasoning

Ancor di più, alcuni hanno pensato che il ragionamento bayesiano potrebbe (dovrebbe) essere usato come paradigma della logica induttiva, cioè al di fuori dell'ambito statistico: una ricetta per il ragionamento umano in generale.

Alcuni esempi di contesti in cui le opinioni sono importanti: fino a che punto il ragionamento bayesiano è utile nella pratica?

- diagnostica: dove l'interesse risiede nel diagnosticare se una persona sottoposta a test sia malata
- legge: dove l'interesse risiede nella convinzione di colpevolezza o innocenza di un imputato.
- scienza (epistemologia): dove l'interesse risiede nella verità di una teoria

La questione è se (in che misura, a quali condizioni) il ragionamento bayesiano può descrivere (modellare) il processo di ragionamento di uno scienziato (giudice / giurato, clinico) che accetta / rifiuta le teorie (decide colpa / innocenza, diagnostica i pazienti).

Diagnostic example: fully appropriate

Law example: Here the reasoning works, the point is that most if not all the probabilities which are involved have to be elicited and are debatable.

Scienze : works when there are clear alternative theories.

How well does B reasoning describe reasoning?

Il principio è che il ragionamento bayesiano funzionerebbe se potessimo definire con precisione tutte le teorie alternative a priori, il che in generale non è realistico: *"because it is very hard to be sufficiently imaginative and because life is short."*.

Il ragionamento bayesiano non può descrivere tutto il ragionamento umano.

Il ragionamento bayesiano è un sistema convincente ma solo

- limitatamene alle ipotesi in esame (e limitatamente alla ragionevolezza di tali ipotesi),
- condizionato alla verosimiglianza data alle prove sotto le vari ipotesi.

In termini statistici, ciò si traduce in condizionatamente alla specificazione del modello (da cui l'importanza di valutare l'adattamento nel controllo dell'adeguatezza del modello)

Letture ulteriori

Per la **storia della statistica Bayesiana**, con esempi, si veda McGrayne (2011).

Per una moderna presentazione dell'approccio **subjective Bayes** si veda Jaynes (2003) and Lindley (2013), further readings include De Finetti (1974), Jeffreys (1998), Lindley (1970) and Savage (1972a).

L'**approccio classico all'inferenza** è descritto in Cox (2006), its principles are discussed in Mayo and Cox (2006) and Mayo (2011). The works which originated the approach are also readable although with some difficulty: Fisher (1922), Fisher (1925) and Neyman, Pearson, and Pearson (1933).

Un **approccio moderno all'inf Bayesiana** è in Gelman, Stern, Carlin, Dunson, Vehtari, and Rubin (2013), see also Gelman and others (2011) and Gelman and Shalizi (2013) per il ruolo dell'inferenza Bayesiana nella scienza.

References

Bayarri, M. J. and J. O. Berger (2004). "The interplay of Bayesian and frequentist analysis".

Cox, D. R. (2006). *Principles of statistical inference*. Cambridge university press.

De Finetti, B. (1974). *Theory of probability*. John Wiley & Sons.

Efron, B. and T. Hastie (2021). *Computer age statistical inference, student edition: algorithms, evidence, and data science*. Vol. 6. Cambridge University Press.

Fisher, R. (1922). "On the mathematical foundations of theoretical statistics". In: *Phil. Trans. R. Soc. Lond. A* 222.594-604, pp. 309-368.

Fisher, R. (1925). *Statistical methods for research workers.* Oliver and Boyd, Edinburgh.

Gelman, A. and others (2011). "Induction and deduction in Bayesian data analysis". In: *Rationality, Markets and Morals* 2.67-78, p. 1999.

Gelman, A. and C. R. Shalizi (2013). "Philosophy and the practice of Bayesian statistics". In: *British Journal of Mathematical and Statistical Psychology* 66.1, pp. 8-38.

Gelman, A., H. S. Stern, J. B. Carlin, et al. (2013). *Bayesian data analysis*. Chapman and Hall/CRC.

Goodman, S. N. (2016). "Aligning statistical and scientific reasoning". In: *Science* 352 (6290), pp. 1180-1181.

References (continue)

Jaynes, E. T. (2003). *Probability theory: The logic of science*. Cambridge university press.

Jeffreys, H. (1998). *The theory of probability*. OUP Oxford.

Lindley, D. V. (1970). *Introduction to probability and statistics from a Bayesian viewpoint*.

Lindley, D. V. (2013). *Understanding uncertainty*. John Wiley & Sons.

Mayo, D. G. (2011). "Statistical science and philosophy of science: Where do/should they meet in 2011 (and beyond)?"

Mayo, D. G. and D. R. Cox (2006). "Frequentist statistics as a theory of inductive inference". In: *Optimality*. Institute of Mathematical Statistics, pp. 77-97.

McGrayne, S. B. (2011). *The theory that would not die: how Bayes' rule cracked the enigma code, hunted down Russian submarines, & emerged triumphant from two centuries of controversy*. Yale University Press.

Neyman, J., E. S. Pearson, and K. Pearson (1933). "IX. On the problem of the most efficient tests of statistical hypotheses". In: *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character* 231.694-706, pp. 289-337. DOI: 10.1098/rsta.1933.0009.

Nuzzo, R. (2014). "Scientific method: Statistical errors". In: *Nature* 506.7487, pp. 150-152. ISSN: 0028-0836.

Pauli, F. (2018). "The p-value Case, a Review of the Debate: Issues and Plausible Remedies". In: *Studies in Theoretical and Applied Statistics*. Ed. by C. Perna, M. Pratesi and A. Ruiz-Gazen. Cham: Springer International Publishing, pp. 95-104. ISBN: 978-3-319-73906-9.

References (continue)

Royall, R. (2004). "The likelihood paradigm for statistical evidence". In: *The nature of scientific evidence: Statistical, philosophical, and empirical considerations*, pp. 119-152.

Savage, L. J. (1972a). *The foundations of statistics*. Dover Publications.

Senn, S. (2011). "You may believe you are a Bayesian but you are probably wrong". In: *Rationality, Markets and Morals* 2.48-66, p. 27.

Note finali

L'arte della modellazione probabilistica

L'arte della modellazione probabilistica consiste nel descrivere in una forma matematica (modello e distribuzioni a priori) cosa già sappiamo e cosa non sappiamo.

La parte "facile" consiste nell'usare la regola di Bayes per aggiornare le incertezze

- sfide computazionali

Altre parti dell'arte della modellazione probabilistica sono, per esempio,

- Controllo del modello: i dati sono in conflitto con la nostra conoscenza a priori?
- presentazione: presentazione del modello e dei risultati agli esperti dell'applicazione

I Vantaggi dell'approccio bayesiano

- Integro le incertezze per concentrarmi sulle parti d'interesse
- Utilizzo le informazioni a priori rilevanti
- I modelli gerarchici
- Controllo e valutazione dei modelli

Calcolo

Dobbiamo essere in grado di calcolare il valore atteso rispetto alla distribuzione a posteriori $p(\theta|y)$

$$E_{\theta|y}\left[g(\theta)\right] = \int p(\theta|y) g(\theta) d\theta$$

- Analitico
    - solo per modelli molto semplici
- Approssimazioni di calcolo con un numero finito di valutazioni di funzione
    - *grid*, *importance sampling*, Monte Carlo, Monte Carlo su catena Markov
    - generico
- Approssimazioni distribuzionali
    - ad es. Laplace, variazionale, Propagazione del valore atteso
    - meno generico, ma può essere molto più veloce con sufficiente accuratezza

Note a margine

Modello e verosimiglianza

Il termine $p(y|\theta, M)$ ha due nomi diversi a seconda del caso. A causa della notazione concisa utilizzata, si può generare confusione.

1. Il termine $p(y|\theta, M)$ è detto **modello** (a volte più specificamente *modello di osservazione* o *modello statistico*) quando è usato per descrivere l'incertezza su y dati θ e M. La notazione più lunga $p_y(y|\theta, M)$ mostra esplicitamente che è una funzione di y.
2. Nella regola di Bayes, il termine $p(y|\theta, M)$ è chiamato **funzione di verosimiglianza**. La distribuzione a posteriori descrive la probabilità (o densità di probabilità) per diversi valori di θ dato un y fissato, e quindi quando la posteriori è calcolata, i termini sul lato destro (nella regola di Bayes) sono anche valutati come funzione di θ dato un y fissato. La notazione più lunga $p_\theta(y|\theta, M)$ mostra esplicitamente che è una funzione di θ.

Il termine ha un proprio nome (verosimiglianza) per differenziarsi rispetto al modello. La funzione di verosimiglianza è distribuzione di probabilità non normalizzata che descrive l'incertezza relativa a θ (ed è per questo che la regola di Bayes ha il termine di normalizzazione per ottenere la distribuzione a posteriori). Vai a regola di Bayes

Notazione ambigua in statistica

In $p(y|\theta)$

- y può essere variabile o valore,
  - potremmo chiarire usando $p(Y|\theta)$ o $p(y|\theta)$
- θ può essere variabile o valore,
  - potremmo chiarire usando $p(y|\Theta)$ o $p(y|\theta)$
- p può essere una funzione discreta o continua di y o θ
  - potremmo chiarire usando P_Y, P_Θ, p_Y o p_Θ
- $P_Y(Y|\Theta=\theta)$ è una funzione di massa di probabilità, distribuzione campionaria, modello di osservazione
- $P(Y=y|\Theta=\theta)$ è una probabilità
- $P_\Theta(Y=y|\Theta)$ è una funzione di verosimiglianza (può essere discreta o continua)
- $p_Y(Y|\Theta=\theta)$ è una funzione di densità di probabilità, distribuzione campionaria, modello di osservazione
- $p(Y=y|\Theta=\theta)$ è una densità
- $p_\Theta(Y=y|\Theta)$ è una funzione di verosimiglianza (può essere discreta o continua)
- y e θ possono anche essere un misto di continuo e discreto

Valutazioni "frequentiste" nella statistica Bayesiana

La teoria bayesiana ha probabilità epistemiche e aleatorie

Le valutazioni "frequentiste" si concentrano sulle proprietà "frequentiste" dati il modello e la la ripetizione aleatoria di un'osservazione

- Consistenza asintotica
- Correttezza
  - non così importante nell'inferenza bayesiana, un errore piccolo è più importante
- Efficienza
  - errore quadratico piccolo
  - altre funzioni di utilità/costo possibili
- Calibrazione
  - L'intervallo a posteriori al α% contiene il vero valore nel α% dei casi
  - l'intervallo predittivo al α% contiene i veri valori futuri nel α% dei casi
  - calibrazione *approssimata* con intervalli più brevi per probabili veri valori, più importante della calibrazione esatta con intervalli lunghi per tutti i valori possibili.

hybridization

Statistica frequentista

La statistica frequentista accetta solo probabilità aleatorie

- Le stime si basano sui dati
- L'incertezza delle stime si basa su tutti i possibili datasets che potrebbe essere stati generati dal meccanismo di generazione dei dati

Le stime vengono derivate per soddisfare le proprietà frequentiste

- La massima verosimiglianza soddisfa solo le proprietà frequentiste asintotiche
- Proprietà desiderate più comuni sono 1) correttezza, 2) minima varianza, 3) calibrazione dell'intervallo di confidenza
- Il requisito di correttezza può portare a una maggiore varianza o a stime senza senso (e.g., una stima corretta per un parametro strettamente positivo può essere negativa)
- L'intervallo di confidenza si dice contenere il vero valore nel α% dei casi di generazione ripetuta di dati dal meccanismo di generazione dei dati
    - non dice quanto è probabile che il valore vero sia all'interno dell'intervallo costruito sui dati osservati
    - non è necessario che sia utile per avere una calibrazione perfetta

Frequentisti vs Bayes vs altri

C'è una grande quantità di tecniche della statistica frequentista molto utili

- per modelli semplici e avendo molti dati non c'è molta differenza

Inferenza bayesiana

- più semplice per modelli complessi, ad es. gerarchici
- più semplice quando il modello cambia
- un modo coerente per aggiungere informazioni a priori

Gran parte delle tecniche di apprendimento automatico (*machine learning*) non sono nè puramente frequentiste nè puramente bayesiane

Appendice

Riepilogo: excursus storico

- Giochi d'azzardo
- **Probabilità classica**: probabilità elementari e calcolo combinatorio
- Girolamo Cardano (1501-1576), scrisse il primo trattato sistematico di probabilità nel 1576. Fu anche un giocatore d'azzardo
- **Probabilità frequentista**
- Jakob Bernoulli (1654-1705; Ars conjectandi, 1713) e Abraham de Moivre (1667-1754; Laws of Chancee, 1718) svilupparono l'interpretazione frequentista della probabilità con la formulazione della ***Legge dei grandi numeri***

Theorem ((Strong) Law of large numbers)

Let $E_1, \ldots, E_n, \ldots$ be a sequence of independent events such that $P(E_i) = p$ for all i. Let $S_n = \sum_{i=1}^{n} |E_i|$ be the number of events occurring among the first n. Then

$$P\left(\lim_{n\to\infty} \frac{S_n}{n} = p\right) = 1.$$

Teorema di Bayes e probabilità delle cause

- Thomas Bayes (1702-1761) fu un ministro presbiteriano. In *Essay Towards Solving a Problem* nella *Doctrine of Chances* (1763) egli considera il problema della probabilità inversa per cui formalizza una soluzione.
- Pierre-Simon Laplace (1749-1827) in Essai philosophique sur les probabilités (1814) fa un trattazione sistematica dell'approccio che oggi chiamiamo Bayesiano

Nuove domande, nuove risposte

- sir Francis Galton (1822-1911) studiò la eugenetica. Sviluppa la **regressione lineare**
- Karl Pearson (1857-1936) introduce il concetto di **correlazione** e di ***goodness of fit***
- William Gosset (1876-1937) lavorando alla Guiness sviluppa la distribuzione t ***di Student*** per valutare la qualità dell'orzo

Verosimiglianza e principio del campionamento ripetuto

- sir Ronald Fisher (1890-1932) introduce, tra l'altro, i concetti di **verosimiglianza**, **analisi della varianza**, **design sperimentale**. Inoltre, dà origine alle idee di **sufficienza**, **ancillarità** e **informazione**. Le sue opere principali: Statistical Methods for Research Workers (1925), The design of experiments (1935), Contributions to mathematical statistics (1950), Statistical methods and statistical inference (1956)
- Egon Pearson (1895-1980) con Jerzy Neyman sviluppa la teoria della **verifica d'ipotesi**.

Probabilità soggettiva

- Bruno de Finetti (1906-1985 ca.), probabilista e attuario italiano (per Generali) propone la **definizione soggettiva di probabilità** e il principio di coerenza, basato sull'interpretazione della **scommessa** (vedi Teoria di probabilità (1970), dove ha scritto *"La probabilità non esiste"*)

- Nel XX secolo l'interesse per la **statistica bayesiana** si è rinnovato per lo studio di nuovi problemi (valutare la probabilità di eventi che non sono mai stati osservati, combinare diverse fonti di informazione) e perché i progressi tecnici hanno permesso la risoluzione di problemi che richiedono un'intensità di calcolo elevata (i metodi Monte Carlo sono fondamentali).
    - R.A. Fisher usò nel 1950 per la prima volta il termine "bayesiano" per enfatizzare la differenza con il termine generale "teoria della probabilità". (Il termine è diventato rapidamente popolare, perché le descrizioni alternative erano più lunghe.)
    - L'**approccio soggettivo** di Bayes è proposto da Bruno de Finetti (1906-1985), Dannis Lindley, Leonard Savage.
- I contrasti tra l'approccio classico e quello bayesiano esistevano ancora in questo periodo. *Objective Bayes*, di cui Harold Jeffreys è uno dei principali fautori, è un tentativo di compromesso tra i due, cercando di evitare o limitando il ruolo della soggettività nell'analisi.

Torna a riepilogo ←

Grazie!