Metodi Matematici e Statistici per Giuristi

I parte - Lezione 03: probabilità

Dario Trevisan

25/09/2023

Section 1

Introduzione alla probabilità

Ragionamento logico e buon senso

Le regole di inferenza logica **non** permettono di dedurre, dalla validità di una implicazione materiale $A \to B$ e della conseguente B, la validità di A. Ad esempio,

$$A \to B = \text{"se piove, Luca prende l'ombrello"}$$

$$B = \text{"Luca prende l'ombrello"}$$

La ragione ci dice che Luca potrebbe prendere l'ombrello per altri motivi (ad esempio lo porta a riparare), quindi

$$A = \text{"Piove"}$$

non è necessariamente vera, ma il **buon senso** ci spinge a credere che forse, dopotutto, stia piovendo.

Probabilità (soggettiva) come grado di fiducia

Considerando tanti esempi simili si capisce che i valori di verità 0 (falso) e 1 (vero) sono troppo *rigidi* per le situazioni reali in cui non vi sia **certezza** che una affermazione A sia vera, ma si dispone solamente di una informazione (una proposizione I) che si ritiene vera, ma non è sufficiente a dedurre la validità o meno di A, è solamente *parziale* (un indizio).

Si introduce quindi la **probabilità** di una proposizione A, sulla base di informazione nota I, come misura del *grado di fiducia* di un soggetto razionale circa la validità di A (supponendo che I sia vera), e si scrive:

$$P(A|I)$$

- Ogni probabilità è sempre un numero compreso tra 0 (se la fiducia è nulla, ossia A è ritenuta falsa) ed 1 (se la fiducia è massima, ossia A è ritenuta vera).

Probabilità (soggettiva) come grado di fiducia

Considerando tanti esempi simili si capisce che i valori di verità 0 (falso) e 1 (vero) sono troppo *rigidi* per le situazioni reali in cui non vi sia **certezza** che una affermazione A sia vera, ma si dispone solamente di una informazione (una proposizione I) che si ritiene vera, ma non è sufficiente a dedurre la validità o meno di A, è solamente *parziale* (un indizio).

Si introduce quindi la **probabilità** di una proposizione A, sulla base di informazione nota I, come misura del *grado di fiducia* di un soggetto razionale circa la validità di A (supponendo che I sia vera), e si scrive:

$$P(A|I)$$

- Ogni probabilità è sempre un numero compreso tra 0 (se la fiducia è nulla, ossia A è ritenuta falsa) ed 1 (se la fiducia è massima, ossia A è ritenuta vera).
- Nel gergo probabilistico le proposizioni A, I, ecc. sono dette **eventi**.

- Quella descritta sopra è l'interpretazione **soggettiva** della probabilità, che dipende dal soggetto o meglio dall'informazione I di cui dispone.

- Quella descritta sopra è l'interpretazione **soggettiva** della probabilità, che dipende dal soggetto o meglio dall'informazione I di cui dispone.
- Questo non significa che ciascuno può attribuire la probabilità che preferisce: come la logica deduttiva, il soggetto è ideale, *completamente razionale*, e usa tutta e sola l'informazione I di cui dispone e di **regole di calcolo** precise.

- Quella descritta sopra è l'interpretazione **soggettiva** della probabilità, che dipende dal soggetto o meglio dall'informazione I di cui dispone.
- Questo non significa che ciascuno può attribuire la probabilità che preferisce: come la logica deduttiva, il soggetto è ideale, *completamente razionale*, e usa tutta e sola l'informazione I di cui dispone e di **regole di calcolo** precise.
- Altre intepretazioni della probabilità hanno lo svantaggio di essere applicabili solo a situazioni più controllate (ad esempio esperimenti ripetuti), e quindi *meno utili* nelle scienze umane.

- Quella descritta sopra è l'interpretazione **soggettiva** della probabilità, che dipende dal soggetto o meglio dall'informazione I di cui dispone.
- Questo non significa che ciascuno può attribuire la probabilità che preferisce: come la logica deduttiva, il soggetto è ideale, *completamente razionale*, e usa tutta e sola l'informazione I di cui dispone e di **regole di calcolo** precise.
- Altre intepretazioni della probabilità hanno lo svantaggio di essere applicabili solo a situazioni più controllate (ad esempio esperimenti ripetuti), e quindi *meno utili* nelle scienze umane.
- L'informazione nota I è **fondamentale** per stabilire la probabilità di A – un po' come le *ipotesi* in un ragionamento deduttivo – ed è buona pratica chiedersi sempre quale sia, se non è specificata. Spesso infatti si scrive solo $P(A)$, ma non bisogna pensare ad una probabilità "assoluta" di una affermazione A: è sempre "relativa" all'informazione di cui un soggetto dispone (che qui è solo sottointesa).

Proprietà di monotonia

Si consideri il seguente quesito:

Linda ha 31 anni, single, aperta e molto brillante. Si è laureata in filosofia. Da studente, era molto impegnata nei problemi di discriminazione e giustizia sociale, e ha anche partecipato a manifestazioni antinucleari.

- Dire quale delle due seguenti affermazioni è più probabile:

Proprietà di monotonia

Si consideri il seguente quesito:

Linda ha 31 anni, single, aperta e molto brillante. Si è laureata in filosofia. Da studente, era molto impegnata nei problemi di discriminazione e giustizia sociale, e ha anche partecipato a manifestazioni antinucleari.

- Dire quale delle due seguenti affermazioni è più probabile:
  a. Linda lavora in banca.

Proprietà di monotonia

Si consideri il seguente quesito:

Linda ha 31 anni, single, aperta e molto brillante. Si è laureata in filosofia. Da studente, era molto impegnata nei problemi di discriminazione e giustizia sociale, e ha anche partecipato a manifestazioni antinucleari.

- Dire quale delle due seguenti affermazioni è più probabile:
  a. Linda lavora in banca.
  b. Linda lavora in banca ed è un'attivista nel movimento femminista.

Prima di considerare le regole di calcolo precise (quantitative), c'è una proprietà "qualitativa" della probabilità molto utile che afferma: date due proposizioni, A, B e l'informazione nota I, se B è vera in qualsiasi situazione in cui A ed I siano entrambe vere, allora

$$P(A|I) \leq P(B|I).$$

- **Esempio**: Consideriamo $A =$ "Luca è toscano", $B =$ "Luca è italiano", $I =$ "Luca studia a Pisa". Allora, poiché se A è vera allora lo è anche B (in realtà qualsiasi sia I), segue che è più probabile B rispetto ad A.

Prima di considerare le regole di calcolo precise (quantitative), c'è una proprietà "qualitativa" della probabilità molto utile che afferma: date due proposizioni, A, B e l'informazione nota I, se B è vera in qualsiasi situazione in cui A ed I siano entrambe vere, allora

$$P(A|I) \leq P(B|I).$$

- **Esempio**: Consideriamo $A =$ "Luca è toscano", $B =$ "Luca è italiano", $I =$ "Luca studia a Pisa". Allora, poiché se A è vera allora lo è anche B (in realtà qualsiasi sia I), segue che è più probabile B rispetto ad A.
- La proprietà di monotonia per quanto evidente può non essere intuitivamente ovvia in certi casi.

Section 2

Regole di calcolo fondamentali

Regole fondamentali

Come nella logica, il calcolo delle probabilità consiste nel determinare la probabilità del risultato di operazioni tra proposizioni. In un certo senso, estendiamo le tavole di verità al caso della probabilità.

Si può riassumere il calcolo in due regole fondamentali, che si occupano di determinare le probabilità della disgiunzione inclusiva

$$P(A \text{ oppure } B|I)$$

e della congiunzione

$$P(A \text{ e } B|I).$$

Da queste seguono poi altre regole, ad esempio per la negazione $P(\text{non A}|I)$, o la formula di Bayes.

Regola della somma

Date due affermazioni A, B, e un'informazione nota I, si ha che

$$P(A \text{ oppure } B|I) = P(A|I) + P(B|I)$$

purché A **e** B **non siano mai vere entrambe** (supponendo vera I), ossia "A e B" è falsa. Tale **regola della somma** è detta anche *additività* della probabilità.

- Due affermazioni A, B che non siano mai vere entrambe (supponendo vera l'informazione I) sono dette **incompatibili** o *mutuamente esclusive*.

Regola della somma

Date due affermazioni A, B, e un'informazione nota I, si ha che

$$P(A \text{ oppure } B|I) = P(A|I) + P(B|I)$$

purché A **e** B **non siano mai vere entrambe** (supponendo vera I), ossia "A e B" è falsa. Tale **regola della somma** è detta anche *additività* della probabilità.

- Due affermazioni A, B che non siano mai vere entrambe (supponendo vera l'informazione I) sono dette **incompatibili** o *mutuamente esclusive*.
- **Esempio**: $I =$ "Maria prende solamente un mezzo di trasporto per venire all'università"; $A =$ "Maria prende l'autobus", $B =$ "Maria usa la bicicletta". Se supponiamo che $P(A|I) = 15\%$, $P(B|I) = 12\%$, allora

$$P(\text{Maria prende l'autobus o la bici}|I) = 27\%.$$

Probabilità del complementare

Nel caso in cui $B =$ "non A", poiché "A oppure B" è sempre vera, segue che

$$1 = P(A|I) + P(\text{"non } A\text{"}|I),$$

quindi

$$P(\text{"non } A\text{"}|I) = 1 - P(A|I).$$

Estensione a più di due affermazioni

La regola della somma (e del complementare) si estende al caso di tre o più affermazioni A, B, C, ecc. purché **a due a due** incompatibili, ossia tali che, supponendo vera l'informazione I, *al più* una tra queste sia vera, ad esempio

$$P(A \text{ o } B \text{ o } C|I) = P(A|I) + P(B|I) + P(C|I)$$

- Se invece le affermazioni non sono incompatibili, la regola della somma diventa più complicata: ad esempio vale

$$P(A \text{ o } B|I) = P(A|I) + P(B|I) - P(A \text{ e } B|I).$$

Sistemi di alternative

Una collezione A_1, A_2, ..., A_n di proposizioni tale che, nota l'informazione I, *esattamente una e una sola* tra loro è sicuramente vera (senza sapere quale) è detta **sistema di alternative**. La regola della somma si estende:

$$1 = P(A_1|I) + P(A_2|I) + \ldots + P(A_n|I).$$

- La regola tuttavia non prescrive come attribuire le probabilità $P(A_1|I)$ ecc. Una possibilità è usare il **principio di Laplace**: *se l'informazione nota I non favorisce alcuna alternativa rispetto alle altre, si pongono tutte le probabilità uguali tra loro:*

$$P(A_1|I) = \ldots, = P(A_n|I) = \frac{1}{n}.$$

Sistemi di alternative

Una collezione A_1, A_2, ..., A_n di proposizioni tale che, nota l'informazione I, *esattamente una e una sola* tra loro è sicuramente vera (senza sapere quale) è detta **sistema di alternative**. La regola della somma si estende:

$$1 = P(A_1|I) + P(A_2|I) + \ldots + P(A_n|I).$$

- La regola tuttavia non prescrive come attribuire le probabilità $P(A_1|I)$ ecc. Una possibilità è usare il **principio di Laplace**: *se l'informazione nota I non favorisce alcuna alternativa rispetto alle altre, si pongono tutte le probabilità uguali tra loro:*

$$P(A_1|I) = \ldots, = P(A_n|I) = \frac{1}{n}.$$

- Se la probabilità è **uniforme** come sopra segue che, se una affermazione B è la disgiunzione di k alternative (i casi "favorevoli"):

$$P(B|I) = k/n = \text{"casi favorevoli"}/\text{"casi possibili"}.$$

Esempio: Si lancia un dado: per ciascuna faccia si introduce l'alternativa

$$A_i = \text{"esce la faccia } i\text{"}.$$

Non avendo alcuna informazione sul dado o su come viene lanciato, poniamo probabilità uniforme sulle 6 alternative.

Data una affermazione riguardante il lancio, ad esempio $B =$ "esce una faccia pari", basta riscriverla come disgiunzione delle alternattive *favorevoli* e sommare tali probabilità. Ne segue la classica formula di probabilità come *casi possibili*/*casi favorevoli*:

$$P(\text{"esce pari"}|\text{"si lancia un dado"}) = \frac{3}{6} = \frac{1}{2}.$$

- **Attenzione**: non tutti i sistemi di alternative hanno probabilità uniforme! Esempio: o "passo l'esame" o "non passo l'esame", ma sapendo di aver studiato la probabilità di passare è $> 1/2$

Regola del prodotto

La regola del prodotto indica come calcolare la probabilità di una congiunzione di due affermazioni A, B (sapendo l'informazione I):

$$P(A \text{ e } B|I) = P(A|I)P(B|A \text{ e } I).$$

- Le affermazioni A, B sono qualsiasi (non necessariamente incompatibili)

Regola del prodotto

La regola del prodotto indica come calcolare la probabilità di una congiunzione di due affermazioni A, B (sapendo l'informazione I):

$$P(A \text{ e } B|I) = P(A|I)P(B|A \text{ e } I).$$

- Le affermazioni A, B sono qualsiasi (non necessariamente incompatibili)
- In generale $P(B|A \text{ e } I)$ può essere maggiore, minore oppure uguale a $P(B|I)$. Se è uguale,

$$P(B|A \text{ e } I) = P(B|I),$$

B è detta **indipendente** da A (sapendo I). In tal caso vale

$$P(A \text{ e } B|I) = P(A|I)P(B|I).$$

Esempio: La probabilità che piova a Pisa oggi (A) è il 15%. Se piove, Luca porta l'ombrello (B) con probabilità 70%. La probabilità di "A e B", ossia che oggi piova a Pisa e Luca porti l'ombrello è quindi

$$15\% \cdot 70\% = 10,5\%$$

- Notiamo che è minore delle probabilità di entrambe le affermazioni (per la proprietà di monotonia).

Esempio: La probabilità che piova a Pisa oggi (A) è il 15%. Se piove, Luca porta l'ombrello (B) con probabilità 70%. La probabilità di "A e B", ossia che oggi piova a Pisa e Luca porti l'ombrello è quindi

$$15\% \cdot 70\% = 10,5\%$$

- Notiamo che è minore delle probabilità di entrambe le affermazioni (per la proprietà di monotonia).
- Intuitivamente B *non* è indipendente da A. Tuttavia potrebbe esserlo se $P(B) = 70\%$, ossia se Luca porta l'ombrello con probabilità 70% anche senza sapere che fuori piove. L'indipendenza di B da A è proprio questo: **sapere che A è vera non modifica il grado di fiducia circa la validità di B.**

Section 3

Inferenza bayesiana

Indipendenza e inferenza

L'indipendenza di due affermazioni A, B è utile perché permette di calcolare agevolmente la probabilità di una congiunzione "A e B",

$$P(\text{"}A \text{ e } B\text{"}|I) = P(A|I)P(B|I),$$

ma impedisce di **imparare** qualcosa su B se si osserva che A è vera: il grado di fiducia sulla validità di B non cambia.

- Per una efficace inferenza probabilistica è fondamentale essere in grado di **aggiornare** le probabilità tenendo conto di tutta l'informazione I disponibile.

Indipendenza e inferenza

L'indipendenza di due affermazioni A, B è utile perché permette di calcolare agevolmente la probabilità di una congiunzione "A e B",

$$P(\text{"}A \text{ e } B\text{"}|I) = P(A|I)P(B|I),$$

ma impedisce di **imparare** qualcosa su B se si osserva che A è vera: il grado di fiducia sulla validità di B non cambia.

- Per una efficace inferenza probabilistica è fondamentale essere in grado di **aggiornare** le probabilità tenendo conto di tutta l'informazione I disponibile.
- È anche importante notare che l'indipendenza di B da A (sapendo I) potrebbe non essere più vera se arriva nuova informazione J al posto di I (si pensi ad una scoperta che collega le due affermazioni, all'inizio apparentemente scollegate).

Formula di Bayes

La regola del prodotto è **simmetrica**: possiamo scambiare il ruolo di A e B (omettiamo I per chiarezza)

$$P(\text{"}A \text{ e } B\text{"}) = P(A)P(B|A) = P(B)P(A|B).$$

Dividendo per $P(B)$ (supponendolo positivo) si trova la **formula di Bayes**:

$$P(A|B) = P(A)\frac{P(B|A)}{P(B)}.$$

Nonostante la semplicità, ha un ruolo fondamentale nell'*inferenza* probabilistica.

La formula di Bayes esplicitando l'informazione nota I diventa

$$P(A|B \text{ e } I) = P(A|I)\frac{P(B|A \text{ e } I)}{P(B|I)}.$$

- La formula prescrive come **aggiornare** la probabilità di A se si aggiunge l'informazione che B sia vero (oltre ad I).

La formula di Bayes esplicitando l'informazione nota I diventa

$$P(A|B \text{ e } I) = P(A|I)\frac{P(B|A \text{ e } I)}{P(B|I)}.$$

- La formula prescrive come **aggiornare** la probabilità di A se si aggiunge l'informazione che B sia vero (oltre ad I).
- La probabilità $P(A|I)$ è detta *a priori* (perché **prima** di osservare che B sia vero)

La formula di Bayes esplicitando l'informazione nota I diventa

$$P(A|B \text{ e } I) = P(A|I)\frac{P(B|A \text{ e } I)}{P(B|I)}.$$

- La formula prescrive come **aggiornare** la probabilità di A se si aggiunge l'informazione che B sia vero (oltre ad I).
- La probabilità $P(A|I)$ è detta *a priori* (perché **prima** di osservare che B sia vero)
- La probabilità $P(A|B \text{ e } I)$ è detta *a posteriori* (**dopo** aver osservato che B è vero)

La formula di Bayes esplicitando l'informazione nota I diventa

$$P(A|B \text{ e } I) = P(A|I)\frac{P(B|A \text{ e } I)}{P(B|I)}.$$

- La formula prescrive come **aggiornare** la probabilità di A se si aggiunge l'informazione che B sia vero (oltre ad I).
- La probabilità $P(A|I)$ è detta *a priori* (perché **prima** di osservare che B sia vero)
- La probabilità $P(A|B \text{ e } I)$ è detta *a posteriori* (**dopo** aver osservato che B è vero)
- Il termine $P(B|A \text{ e } I)$ è detto anche **verosimiglianza** di A rispetto a B (sapendo I) e si indica anche con

$$L(A; B) = P(B|A \text{ e } I)$$

(dall'inglese **Likelihood**)

- Il denominatore $P(B|I)$ non è di solito rilevante, perché spesso si vuole solo valutare se A sia più o meno probabilie della sua negazione "non A", dopo aver osservato B. Basta allora considerare il rapporto

$$\frac{P(A|B \text{ e } I)}{P(\text{"non } A\text{"}|B \text{ e } I)} = \frac{P(A|I)}{P(\text{"non } A\text{"}|I)} \cdot \frac{P(B|A \text{ e } I)}{P(\mathrm{B}|\text{non } A \text{ e } I)}$$

e determinare se è maggiore o minore di 1.

- Il denominatore $P(B|I)$ non è di solito rilevante, perché spesso si vuole solo valutare se A sia più o meno probabilie della sua negazione "non A", dopo aver osservato B. Basta allora considerare il rapporto

$$\frac{P(A|B \text{ e } I)}{P(\text{"non } A\text{"}|B \text{ e } I)} = \frac{P(A|I)}{P(\text{"non } A\text{"}|I)} \cdot \frac{P(B|A \text{ e } I)}{P(\mathrm{B}|\text{non } A \text{ e } I)}$$

e determinare se è maggiore o minore di 1.

- Si può anche riscrivere il rapporto usando la verosimiglianza:

$$\frac{P(A|B \text{ e } I)}{P(\text{"non } A\text{"}|B \text{ e } I)} = \frac{P(A|I)}{P(\text{"non } A\text{"}|I)} \cdot \frac{L(A; B)}{L(\text{"non } A\text{"}; B)}.$$

  Il **rapporto di verosimiglianza** $L(A; B)/L(\text{"non } A\text{"}; B)$ determina quindi quanto il grado di fiducia a posteriori sarà sulla validità di A oppure di "non A". Se è maggiore di 1, allora la probabilità di A aumenta.

Esempio: Supponiamo di essere molto incerti circa se oggi pioverà (A) e quindi *a priori* $P(A) = 1/2$ (non scriviamo l'informazione nota I). Se piove, Luca porterà l'ombrello con probabilità 70% ossia

$$L(\text{piove; Luca ha l'ombrello}) = P(B|A) = 70\%,$$

mentre se non piove la probabilità scende al 10%, ossia

$$L((\text{non piove; Luca ha l'ombrello}) = P(B|\text{"non A"}) = 10\%.$$

Il rapporto di verosimiglianza vale

$$\frac{L(\text{piove; Luca ha l'ombrello})}{L(\text{non piove; Luca ha l'ombrello})} = 7,$$

quindi il rapporto tra le probabilità cresce di un fattore 7: è molto più probabile che piova avendo osservato che Luca ha portato l'ombrello.

- Se la probabilità a priori è $P(A) = 5\%$ un rapporto di verosimiglianza 7 non basta a "spostare" il grado di fiducia a favore di A, perché si trova

$$\frac{P(A|B)}{P(\text{"non } A\text{"}|B)} = \frac{5\%}{95\%} \cdot 7 \approx 37\%.$$

Ragionamento bayesiano

Il *metodo bayesiano* consiste nell'applicare la formula di Bayes per quantificare il grado di fiducia circa una o più affermazioni, sulla base di osservazioni (indizi, evidenze, ecc.).

In particolare, nel caso di un sistema di alternative, $A_1, \ldots, A_n$, si può calcolarne le probabilità a posteriori applicando a ciascuna la formula di Bayes (sottointendiamo I per chiarezza)

$$P(A_i|B) = P(A_i)P(B|A_i) \cdot \frac{1}{P(B)} = P(A_i)L(A_i; B) \cdot \frac{1}{P(B)}.$$

- Il denominatore $P(B)$ è lo stesso per tutte le alternative, quindi è **irrilevante** se si devono solo confrontare per determinare l'alternativa più probabile.

Ricordando che la somma delle probabilità di un sistema di alternative vale 1, si può ricavare una formula per il denominatore

$$P(B) = \sum_{i=1}^{n} P(A_i)P(B|A_i) = \sum_{i=1}^{n} P(A_i)L(A_i; B).$$

- Se una alternativa è molto improbabile *a priori*, ossia $P(A_i)$ è molto vicina a 0, affinché diventi molto probabile *a posteriori* la verosimiglianza $L(A_i; B)$ deve essere molto grande. In altre parole, **teorie implausibili** a priori necessitano di **evidenze molto forti** per essere considerate alla pari o di più di alternative che invece non lo erano.

Ricordando che la somma delle probabilità di un sistema di alternative vale 1, si può ricavare una formula per il denominatore

$$P(B) = \sum_{i=1}^{n} P(A_i)P(B|A_i) = \sum_{i=1}^{n} P(A_i)L(A_i; B).$$

- Se una alternativa è molto improbabile *a priori*, ossia $P(A_i)$ è molto vicina a 0, affinché diventi molto probabile *a posteriori* la verosimiglianza $L(A_i; B)$ deve essere molto grande. In altre parole, **teorie implausibili** a priori necessitano di **evidenze molto forti** per essere considerate alla pari o di più di alternative che invece non lo erano.
- **Esempio:** un oggetto volante non identificato (UFO) potrebbe essere un'astronave aliena (A_1) oppure un aereo spia (A_2). Anche se osservando le foto dell'avvistamento la verosimiglianza potrebbe puntare verso A_1, la probabilità a priori è **o dovrebbe** essere sbilanciata a favore di A_2 (ad esempio per il semplice fatto che nel passato tutti gli UFO si sono rivelati tali o comunque non extraterrestri).

La formula di Bayes, con opportune probabilità a priori, può anche rappresentare il "pregiudizio" in un soggetto e quindi la necessità di dover portare prove "schiaccianti" per fargli cambiare opinione.

- **Esempio:** Per un caso di omicidio vi sono n sospettati (l'alternativa A_i significa che il sospettato i è colpevole). Se si trova una prova B che è a favore sia del primo che del secondo, in equal misura,

  $$L(A_1; B) = L(A_2; B),$$

  questa potrebbe in realtà scagionare solo uno dei due, ad esempio se $P(A_1)$ è molto più grande di $P(A_2)$.

La formula di Bayes, con opportune probabilità a priori, può anche rappresentare il "pregiudizio" in un soggetto e quindi la necessità di dover portare prove "schiaccianti" per fargli cambiare opinione.

- **Esempio:** Per un caso di omicidio vi sono n sospettati (l'alternativa A_i significa che il sospettato i è colpevole). Se si trova una prova B che è a favore sia del primo che del secondo, in equal misura,

$$L(A_1; B) = L(A_2; B),$$

  questa potrebbe in realtà scagionare solo uno dei due, ad esempio se $P(A_1)$ è molto più grande di $P(A_2)$.

- Va notato che nei soggetti reali, il metodo bayesiano, se è applicato, lo è di solito in forma *qualitativa* (senza calcolare probabilità). Vista questa dipendenza forte dalle probabilità a priori, sarebbe però auspicabile in molte occasioni che queste venissero dichiarate.

Stima di massima verosimiglianza

Se le probabilità a priori di un sistema di n alternative $P(A_i)$ sono uniformi, allora le probabilità a posteriori dipendono unicamente dalle verosimiglianze $L(A_i; B)$:

$$P(A_i|B) = L(A_i; B) \cdot \frac{1}{nP(B)},$$

dove abbiamo evidenziato i termini comuni che *non* dipendono da i.

Pertanto, se vogliamo determinare l'alternativa più probabile (*a posteriori*), ossia i_{max} in modo che

$$A_{i_{\mathrm{max}}} = \max_{i=1,\ldots,n} L(A_i; B).$$

- Tale i_{max} è detta stima di **massima verosimiglianza** (o **moda** a posteriori).