

Università degli studi di Padova

FACOLTA' DI SCIENZE STATISTICHE  
CORSO DI LAUREA IN STATISTICA E TECNOLOGIE  
INFORMATICHE



TESI DI LAUREA TRIENNALE

**ALCUNE CONSIDERAZIONI SUL  
FATTORE DI BAYES NEI PROBLEMI  
DI VERIFICA D'IPOTESI**

Relatore: Ch.ma prof.essa Ventura Laura

Laureanda: Scaranaro Letizia  
Matricola: 484292

A.A. 2004/2005

***A Lucia e Natale, perché  
possano essere orgogliosi  
di me.***

# Indice

1	Nozioni di Statistica Bayesiana	5
1.1	Introduzione	5
1.2	Il Teorema di Bayes	7
1.3	Esempi	9
1.4	Conclusioni	13
2	Fattore di Bayes come strumento per la risoluzione di problemi di verifica d'ipotesi	14
2.1	Il Fattore di Bayes	14
2.2	La verifica d'ipotesi	16
2.3	Due ipotesi semplici	17
2.4	Ipotesi generali	21
2.5	Ipotesi alternativa composta	22
2.6	Problema della scelta tra modelli	24
3	Estensioni e aspetti computazionali legati al Fattore di Bayes	27
3.1	Estensioni del Fattore di Bayes	27
3.1.1	Fattore di Bayes Intrinseco	27
3.1.2	Fattore di Bayes Frazionario	28
3.2	Aspetti computazionali	29
3.2.1	Il metodo Monte Carlo	30
3.2.2	Il metodo Monte Carlo Importance Sampling	31
3.2.3	Metodo Monte Carlo basato sulle catene di Markov	32

4	La verosimiglianza integrata	33
4.1	Introduzione	33
4.2	Un'applicazione	35
4.3	Un'applicazione con dati reali	37
4.3.1	Descrizione del problema	37
4.3.2	Descrizione delle procedure impiegate	38
4.3.3	Risultati ottenuti	41
Appendice		
A1	I dati	43
A2	Le funzioni in R	44
Bibliografia		50

## Capitolo 1

# Nozioni di Statistica Bayesiana

### 1.1 Introduzione

Nel mondo statistico si distinguono due diverse scuole di pensiero: la scuola classica o frequentista e la scuola Bayesiana. La motivazione di fondo che divide i due approcci statistici riguarda il significato che essi danno alla nozione di probabilità.

Dal punto di vista frequentista la probabilità è concepita in modo oggettivo. Essa rappresenta una caratteristica intrinseca degli eventi per i quali viene calcolata e deve essere depurata di ogni elemento soggettivo che la riguarda e la caratterizza. Dal punto di vista Bayesiano, invece, tali elementi non vengono considerati fonte di disturbo. Al contrario, sono il punto di partenza della definizione soggettiva di probabilità, secondo cui essa esprime un'opinione personale dell'individuo nei confronti di un determinato evento. Ciò sta a significare che la probabilità si colloca tra l'individuo e il mondo esterno e non già all'interno dell'evento. "E' dall'interazione che scaturisce tra l'individuo e l'evento, dall'interesse che per l'individuo suscita l'evento che nasce la valutazione della probabilità" (Liseo, 2004). Se si chiede, ad esempio, ad un gruppo di persone di valutare la probabilità che la squadra di calcio A superi la squadra B nella prossima partita di campionato, è lecito aspettarsi tante differenti risposte: un tifoso della squadra A può reputare più probabile la vittoria di A rispetto ad un tifoso della squadra B.

Nella statistica classica, dunque, l'inferenza parametrica viene effettuata non considerando affatto l'eventualità della presenza di informazioni a priori sul fenomeno che si sta analizzando, assumendo cioè che le probabilità iniziali non esistano e che le valutazioni personali non debbano entrare nella trattazione dell'incertezza. L'identificazione del modello statistico, ovvero l'identificazione dei suoi parametri, è pertanto ottenuta esclusivamente basandosi sulle informazioni sperimentali, cioè sulle informazioni che scaturiscono da dati campionari ottenuti con la misura di variabili d'interesse. Il motivo di questo atteggiamento è che nell'impostazione classica della statistica la probabilità è intesa come una frequenza e le sue proprietà sono definite come proprietà asintotiche legate ad un numero infinito di dati e quindi solo per fenomeni replicabili.

Riassumendo, la scuola classica si basa su due principi fondamentali: quello della verosimiglianza e quello della ripetizione dell'esperimento. Il primo stabilisce che tutte le informazioni che un generico campione  $x$  può fornire sono contenute nella funzione di verosimiglianza. Tale funzione è definita come la funzione  $L : \Theta \rightarrow [0, \infty)$  che associa ad ogni valore di  $\theta \in \Theta$ , la probabilità  $p(X = x; \theta)$ , nel caso discreto, oppure la densità di probabilità  $f(x; \theta)$ , nel caso assolutamente continuo. Essa è pertanto lo strumento attraverso cui vengono soppesati i diversi valori dei parametri: i valori più plausibili avranno verosimiglianza maggiore. Il secondo principio considera il campione effettivamente ottenuto alla stregua di uno dei possibili campioni che si sarebbero potuti ricavare ripetendo un gran numero di volte, nelle stesse condizioni, l'operazione di campionamento (Cifarelli e Muliere, 1989).

Si vuol far notare inoltre che la funzione di verosimiglianza  $L(\theta)$  non rappresenta una distribuzione di probabilità su  $\Theta$ , in quanto nell'impostazione classica dell'inferenza sono i dati ad essere considerati come realizzazioni di variabili aleatorie, mentre il parametro ignoto  $\theta$  è considerato come deterministico.

L'approccio Bayesiano si basa principalmente sulla definizione soggettiva della probabilità. Essa cioè è interpretata come rappresentazione del "grado di

fiducia” che un individuo ripone nel verificarsi di un determinato evento e dipende dallo stato di conoscenza, o di ignoranza, di tale evento, che fa parte di ogni individuo in modo diverso. Tale definizione si ripercuote sull’inferenza Bayesiana, in quanto la determinazione dei parametri del modello statistico è ottenuta basandosi sulla prenoscenza, ovvero sulla disponibilità di informazioni a priori rispetto ai dati osservati nel campione, che dipende strettamente dall’esperienza precedentemente accumulata dall’individuo. La prenoscenza è quantificata da una distribuzione di probabilità detta appunto a priori. Per questo motivo i parametri ignoti sono considerati variabili casuali, mentre i dati sono trattati come costanti. Nelle procedure inferenziali Bayesiane le informazioni espresse dalla probabilità a priori dei parametri possono essere aggiornate alla luce dei dati osservati, grazie al teorema di Bayes. In questo modo l’inferenza assume la forma di distribuzione del parametro  $\theta$  condizionata ai dati, detta probabilità a posteriori del parametro (Cipollini e Stefanini, 1995).

Come detto poc’anzi il teorema di Bayes è alla base di tutta l’inferenza Bayesiana. Prima però di enunciare formalmente il teorema di Bayes è bene esplicitare alcuni concetti in modo da renderne più semplice la comprensione.

## 1.2 Il Teorema di Bayes

E’ noto dal calcolo delle probabilità che, presi due eventi qualsiasi  $H$  ed  $E$ , la probabilità dell’intersezione di tali eventi può essere scritta (assiomi di Kolmogoroff) come

$$P(H \cap E) = P(H | E)P(E) \quad ,$$

oppure

$$P(H \cap E) = P(E | H)P(H) \quad .$$

Eguagliando le due equazioni, ed esplicitando rispetto a  $P(H | E)$ , si ottiene:

$$P(H | E) = \frac{P(E | H)}{P(E)} P(H) \quad .$$

Questa equazione rappresenta la forma più semplice del teorema di Bayes e da essa si può vedere che la probabilità che si verifichi un evento  $H$  non è una caratteristica intrinseca dell'evento stesso, ma può essere calcolata tenendo in considerazione tutte le informazioni che si hanno a disposizione. Infatti, il verificarsi dell'evento  $E$ , nell'esempio, cambia la probabilità di  $H$  in  $P(H|E)$ , detta probabilità a posteriori, in quanto l'evento  $E$  si è verificato prima di  $H$ . La quantità  $P(H)$ , invece, rappresenta la probabilità a priori dell'evento  $H$ , la quale racchiude le conoscenze pregresse che si hanno del fenomeno, ovvero le informazioni a priori, da qui il nome di probabilità a priori.

Dopo questa semplice ed intuitiva spiegazione del teorema di Bayes, si dà la definizione di *classe completa*.

Un insieme di proposizioni  $H_j$  è detto formare una *classe completa* o *classe delle ipotesi*, se gode delle seguenti proprietà:

$$\bigcup_j H_j = \Omega \quad \text{e}$$

$$H_i \cap H_j = \emptyset \quad \text{per } i \neq j.$$

Questo concetto viene utilizzato anche nell'impostazione classica. Una classe completa gode della proprietà di normalizzazione se:

$$\sum_j P(H_j) = 1,$$

dove  $P(H_j)$  è la probabilità che si verifichi l'evento  $H_j$  prima dell'evento  $E$ , ovvero è la probabilità a priori di  $H_j$ .

### **Teorema di Bayes**

Sia dato un evento  $E$ . Allora, per ogni evento  $H_j$  vale la seguente formula:

$$P(H_j | E) = \frac{P(E | H_j)}{\sum_{i=1}^k P(H_i)P(E | H_i)} P(H_j), \quad (1)$$



dove  $k$  rappresenta il numero di eventi nella classe completa. Dalla (1) si osserva che  $P(H_j | E)$  è la probabilità che si verifichi l'evento  $H_j$ , o che sia vera l'ipotesi  $H_j$ , condizionata dalla realizzazione dell'evento  $E$ . Pertanto la (1) è la probabilità a posteriori alla luce dell'evento  $E$ .  $P(E | H_j)$  rappresenta invece la probabilità che si verifichi l'evento  $E$  sotto l'ipotesi  $H_j$ .

Si dà ora una diversa espressione della formula di Bayes:

$$\pi(\theta | x) = \frac{\pi(\theta)L(\theta; x)}{\int_{\Theta} \pi(\theta)L(\theta; x)d\theta} . \quad (2)$$

Nella (2)  $x$  rappresenta il campione, realizzazione della variabile casuale  $X$  con distribuzione di probabilità  $p(x; \theta)$ ,  $\theta$  è il parametro appartenente allo spazio dei parametri  $\Theta$ ,  $\pi(\theta)$  rappresenta la probabilità a priori del parametro, cioè quantifica tutte le informazioni che si posseggono su  $\theta$  in termini di una distribuzione di probabilità iniziale e  $L(\theta; x)$  è la funzione di verosimiglianza. Infine  $\pi(\theta | x)$  è la distribuzione finale o a posteriori, dopo aver osservato il risultato sperimentale  $x$ . La quantità

$$\int_{\Theta} \pi(\theta)L(\theta; x)d\theta$$

rappresenta la distribuzione marginale del vettore  $X$ .

### 1.3 Esempi

*Esempio 1 ([www.diee.unica.it](http://www.diee.unica.it))*

Due amici Stanlio e Ollio, dopo parecchio tempo, si incontrano in un bar. Stanlio propone ad Ollio un gioco a carte: chi perde paga da bere. Ollio accetta e perde sempre, anche nelle serate successive che si incontrano.

E' lecito che Ollio sospetti che l'amico sia un baro? Quali sono le probabilità che perda ancora nelle serate seguenti alla luce delle esperienze accumulate?

Ci sono due ipotesi per Stanlio: che sia onesto ( $O$ ) oppure sia un baro ( $B$ ). Sono due ipotesi che rappresentano una classe completa. Se Stanlio fosse un baro, la probabilità che vinca  $n$  volte di seguito nel gioco è ovviamente 1:

$$P(W_n | B) = 1 .$$

Con  $W_n$  si indica l'evento:  $n$  vittorie consecutive al gioco.

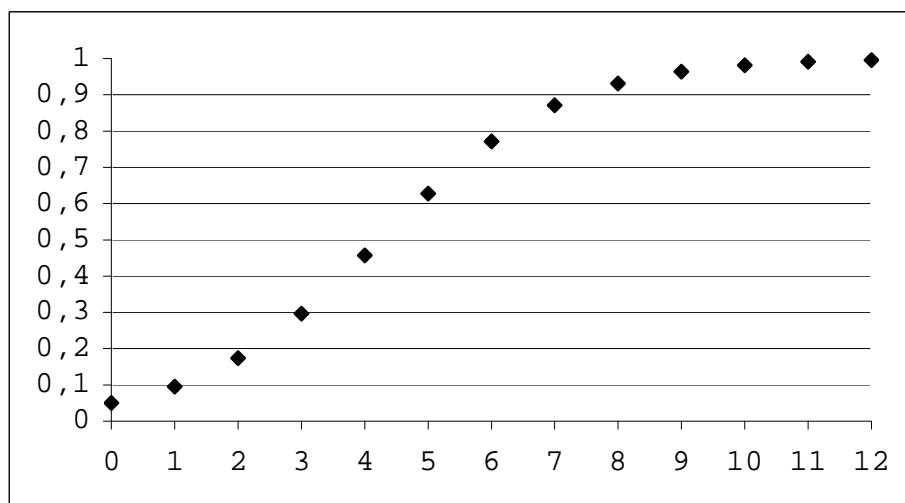
Mentre la probabilità di  $n$  vittorie consecutive, nell'ipotesi di gioco onesto, è invece pari a:

$$P(W_n | O) = \left(\frac{1}{2}\right)^n .$$

La probabilità che Stanlio sia un baro alla luce di  $n$  vincite successive al gioco è data dal teorema di Bayes:

$$P(B | W_n) = \frac{P(W_n | B)P_0(B)}{P(W_n | B)P_0(B) + P(W_n | O)P_0(O)} = \frac{P_0(B)}{P_0(B) + 2^{-n} P_0(O)} .$$

Restano da stabilire le probabilità a priori  $P_0(O)$  e  $P_0(B)$ . Se Ollio ha fiducia nell'amico gli può attribuire una probabilità  $P_0(B)$  molto bassa, ad esempio 5%. Ma all'aumentare delle vittorie dell'amico tale probabilità decresce sempre di più come si può vedere dal grafico:



All'aumentare delle prove tale probabilità ha sempre meno peso e domina l'evidenza sperimentale.

### *Esempio 2 (Liseo, 2004)*

Sondaggio pre-elettorale.

Per un determinato collegio elettorale, si vuole stimare la percentuale di elettori che, nelle prossime elezioni politiche, voterà per le coalizioni di centro-destra e centro-sinistra. Per semplificare un po' il problema si immagina di essere interessati soltanto alla percentuale di voti che otterrà lo schieramento di centro-sinistra.

Dalla popolazione di votanti in quel seggio elettorale si estrae un campione casuale di  $n$  elettori, con  $n$  fissato in anticipo, e a ciascuno di loro si chiede per chi voteranno alle prossime elezioni. A ciascun componente del campione si associa una variabile aleatoria  $Y_j$  che vale 1 se l'elettore vota per il centro-sinistra, 0 altrimenti. Le distribuzioni di probabilità possibili da associare alle variabili aleatorie  $Y_1, \dots, Y_n$  costituiscono, nel loro insieme, il modello statistico.

In questo caso il modello si costruisce partendo dalla considerazione che il valore incognito da stimare, ovvero la percentuale di votanti per la lista di centro-sinistra, è un valore  $\theta \in [0,1]$ . Condizionatamente al valore assunto da  $\theta$ , le  $n$  osservazioni  $Y_1, \dots, Y_n$  vengono supposte indipendenti e tutte con la stessa distribuzione di probabilità, ovvero per ogni  $j$  da 1 a  $n$  si ha:

$$P(Y_j = 1 | \theta) = 1 - P(Y_j = 0 | \theta) = \theta .$$

Si è in presenza di un modello statistico di Bernoulli, in cui

$$Y_i \sim Be(\theta) , \quad i = 1, \dots, n .$$

La funzione di verosimiglianza associata a tale esperimento è:

$$L(\theta; y) = \prod_{i=1}^n P(Y_i = y_i | \theta) ,$$

dove si indicano con le lettere maiuscole le variabili aleatorie e con le lettere minuscole le realizzazioni di tali variabili.

Si suppone che l'esperimento si concluda con  $k$  successi e  $n - k$  insuccessi. Allora si ha

$$L(\theta) = \theta^k (1 - \theta)^{n-k} .$$

Con  $k = 7$  e  $n = 10$ , vale

$$L(\theta) = \theta^7 (1 - \theta)^3 .$$

Il valore  $\hat{\theta} = 0.7$  appare come il più verosimile alla luce dei dati raccolti. L'incertezza relativa a tale stima è espressa da:

$$Var(\hat{\theta}) = \frac{\hat{\theta}(1 - \hat{\theta})}{n} = \frac{0.21}{10} = 0.021 .$$

Ciò che non è stato considerato fin'ora sono le informazioni a priori che si hanno su  $\theta$ . Sulla base di elezioni precedenti però si può già sapere che tipo di percentuali aspettarsi per tale lista. Questa informazione si può inserire nell'analisi statistica attraverso il teorema di Bayes. Si suppone che l'informazione sia raccolta nella distribuzione di probabilità iniziale  $\pi_1(\theta)$ . Combinando la distribuzione a priori con la verosimiglianza si ottiene la distribuzione finale:

$$\pi_1(\theta; y) = \frac{\theta^7 (1 - \theta)^3}{\int_{\Theta} \theta^7 (1 - \theta)^3 \pi_1(\theta) d\theta} \pi_1(\theta) .$$

La scelta di  $\pi_1(\theta)$  non è univoca, potrebbe essere una distribuzione uniforme su  $\Theta$ , che nel caso specifico è l'intervallo  $[0,1]$ , oppure una distribuzione unimodale intorno a  $\theta = 0.5$ . E' comunque comune scegliere la distribuzione all'interno della famiglia di distribuzioni Beta.

## 1.4 Conclusioni

In conclusione si può affermare che la differenza tra l'approccio classico e l'approccio Bayesiano ai problemi di inferenza si ha sostanzialmente nell'uso della formula di Bayes.

La statistica classica rifiuta l'uso di tale formula in quanto non è sempre possibile dare alla distribuzione iniziale un carattere oggettivo, legato cioè alla nozione di frequenza.

La statistica Bayesiana, invece, affronta i problemi di inferenza basandosi sul legame tra la funzione di verosimiglianza, la probabilità a priori e la distribuzione finale, tre quantità fulcro del teorema di Bayes. L'approccio Bayesiano permette di aggiornare le informazioni pregresse che si hanno su un determinato fenomeno alla luce dei dati campionari, ottenendo così una nuova distribuzione di probabilità, ossia la probabilità finale.

Nel capitolo successivo verrà trattato il problema inferenziale della verifica di ipotesi semplici e composte in ambito Bayesiano introducendo il concetto di *Fattore di Bayes*.

## Capitolo 2

# Fattore di Bayes come strumento per la risoluzione di problemi di verifica d'ipotesi

### 2.1 Il fattore di Bayes

Dati due eventi denominati  $H$  e  $D$ , si ricava dal teorema di Bayes la formula della probabilità a posteriori:

$$P(H | D) = \frac{P(H)P(D | H)}{P(H)P(D | H) + P(\overline{H})P(D | \overline{H})} ,$$

dove  $\overline{H}$  indica la negazione dell'evento  $H$ , mentre  $P(H)$  rappresenta la probabilità a priori assegnata all'evento  $H$  e  $P(D | H)$  rappresenta la probabilità condizionata di  $D$  dato  $H$ . Volendo poi confrontare le probabilità a posteriori di due eventi,  $H$  e  $\overline{H}$ , si ottiene:

$$\frac{P(H | D)}{P(\overline{H} | D)} = \frac{P(H)}{P(\overline{H})} \frac{P(D | H)}{P(D | \overline{H})} .$$

Tale confronto corrisponde a trasformare la formula di Bayes in termini di *odds*, dove con *odds* s'intende il rapporto  $probability/(1-probability)$ , ad esempio  $P(H)/(1 - P(H))$ . Con diversa simbologia si ottiene:

$$\begin{aligned}
O(H | D) &= \frac{P(H | D)}{P(\overline{H} | D)} \\
&= \frac{P(H) P(D | H)}{P(\overline{H}) P(D | \overline{H})} \\
&= O(H) B_{H\overline{H}} ,
\end{aligned}$$

dove  $O(H | D)$  rappresenta il rapporto delle probabilità a posteriori, ovvero gli *odds* finali,  $O(H)$  rappresenta il rapporto a priori, ovvero gli *odds* iniziali, e

$$B_{H\overline{H}} = \frac{P(D | H)}{P(D | \overline{H})}$$

rappresenta il rapporto tra le verosimiglianze per  $D$ , che prende il nome di *Fattore di Bayes* a favore dell'evento  $H$ . In altre parole, il fattore di Bayes altro non è che il coefficiente per cui bisogna moltiplicare gli *odds* iniziali dell'evento  $H$  per ottenere gli *odds* finali, considerando il verificarsi dell'evento  $D$ . Esso ha pertanto il ruolo di indicare la veridicità dell'evento  $H$  rispetto all'evento  $\overline{H}$ , basandosi esclusivamente sull'evento osservato  $D$  e non su valutazioni soggettive.

Si intuisce come il fattore di Bayes sia strettamente legato ai problemi di verifica d'ipotesi, i quali consistono nel confronto tra un'ipotesi nulla  $H$  e l'ipotesi complementare  $\overline{H}$ , o più semplicemente tra due ipotesi  $H_0$  e  $H_1$ , disgiunte, ma non esaustive. Il fattore di Bayes diventa:

$$B_{01} = \frac{P(D | H_0)}{P(D | H_1)}$$

e continua a valere l'equazione:

$$O(H_0 | D) = B_{01} O(H_0) ,$$

dove  $O(H_0 | D) = P(H_0 | D) / P(H_1 | D)$  rappresenta il rapporto degli *odds* finali e  $O(H_0) = P(H_0) / P(H_1)$  rappresenta il rapporto degli *odds* iniziali (si rimanda al § 2.2).

Inoltre il fattore di Bayes è visto da molti studiosi bayesiani come uno strumento utile per il problema della scelta tra modelli, argomento trattato nel § 2.6 (Piccinato, 1997).

## 2.2 La verifica d'ipotesi

Nell'impostazione classica della statistica i problemi di verifica d'ipotesi consistono nel confrontare due alternative: l'ipotesi nulla  $H_0 : \theta \in \Theta_0$  e l'ipotesi alternativa  $H_1 : \theta \in \Theta_1$ , con  $\Theta_0 \cup \Theta_1 = \Theta$  e  $\Theta_0 \cap \Theta_1 = \emptyset$ .

Fisher e Neyman-Pearson proposero due approcci differenti per affrontare tali problemi, dovendo rispondere a situazioni diverse.

Fisher affronta i problemi di verifica d'ipotesi come test di significatività del parametro in questione. In questo caso viene presa in considerazione la sola ipotesi nulla  $H_0$ , assegnando così all'ipotesi alternativa  $H_1$  tutte le altre possibili spiegazioni riferite all'esperimento in esame, non previste da  $H_0$ . Questo si traduce in una verifica di compatibilità dei dati osservati con l'ipotesi nulla. Il confronto avviene grazie al calcolo del p-value, il quale rappresenta la probabilità di osservare un campione con caratteristiche più diverse dall'ipotesi nulla, rispetto al campione osservato.

Neyman e Pearson affrontano invece i problemi di verifica d'ipotesi con un'impostazione decisionale. Al contrario dell'approccio proposto da Fisher, essi assegnano ad entrambe le ipotesi la stessa importanza, ovvero vengono ambedue utilizzate in modo esplicito. Infatti, nell'impostazione decisionale la bontà di una procedura per condurre un test di verifica d'ipotesi è valutata misurando la probabilità di commettere due tipi di errore: l'errore di primo tipo  $\alpha$ , definito come la probabilità di rifiutare  $H_0$  quando  $H_0$  è vera, e l'errore di secondo tipo  $\beta$ , definito come la probabilità di accettare  $H_0$  quando è vera  $H_1$ . L'avvicinamento alla scelta dell'ipotesi nulla avviene soltanto grazie alla ricerca del test ottimo, la quale prevede di fissare la probabilità di errore di primo tipo e di minimizzare la probabilità di commettere un errore di secondo tipo. Per avere una visione più precisa dei due paradigmi e per avere maggiore



chiarezza a riguardo delle differenze tra i due approcci, si rimanda ad esempio a Pace e Salvan (1996, § 3.2, Appendice B).

Nel contesto bayesiano, l'ipotesi nulla e l'ipotesi alternativa devono essere definite in modo esplicito e devono essere completamente probabilizzate. Per valutare poi il peso delle due ipotesi,  $H_0$  e  $H_1$ , si calcola la probabilità a posteriori nel modo seguente:

$$P(H_i | x) = \int_{\Theta_i} \pi(\theta | x) d\theta, \quad i = 0, 1,$$

dove  $\pi(\theta | x)$  è la probabilità a posteriori, dopo aver osservato il risultato sperimentale  $x$ , ed è data dalla formula seguente:

$$\pi(\theta | x) = \frac{L(\theta; x)}{\int_{\Theta_i} \pi(\theta) L(\theta; x) d\theta} \pi(\theta).$$

L'ipotesi nulla  $H_0$  viene rifiutata se questo valore di probabilità risulta minore di una soglia critica fissata. Il test risulta simmetrico, in quanto non predilige alcuna delle due ipotesi formulate.

### 2.3 Due ipotesi semplici

Si consideri inizialmente il caso più semplice. Ossia si consideri il caso in cui sia l'ipotesi nulla che l'ipotesi alternativa sono ipotesi puntuali. In questo tipo di problema si capisce chiaramente il ruolo dell'informazione a priori e della funzione di verosimiglianza. Date le ipotesi

$$H_0 : \theta = \theta_0$$

$$H_1 : \theta = \theta_1,$$

siano date le seguenti probabilità a priori:

$$P(H_0) = \pi_0 \quad \text{e} \quad P(H_1) = \pi_1 = 1 - \pi_0.$$

Confrontando le probabilità a posteriori delle due ipotesi si ottiene il seguente rapporto:

$$\frac{P(H_0 | x)}{P(H_1 | x)} = \frac{\pi(\theta_0 | x)}{\pi(\theta_1 | x)} = \frac{\pi_0}{1 - \pi_0} \frac{L(\theta_0; x)}{L(\theta_1; x)}.$$

Gli *odds* finali risultano dal prodotto tra  $\pi_0/(1-\pi_0)$ , il quale rappresenta gli *odds* iniziali, ovvero il peso relativo delle due ipotesi prima di osservare i dati, e  $L(\theta_0; x)/L(\theta_1; x)$ , il quale viene indicato con  $B_{01}$  e altro non è che il *Fattore di Bayes*. Scrivendo il rapporto a posteriori con la simbologia introdotta nel § 2.1, si ha

$$O(H_0 | x) = B_{01}O(H_0) \quad .$$

Ancora, risulta chiaro definire il fattore di Bayes come il fattore moltiplicativo che trasforma gli *odds* iniziali in *odds* finali, alla luce dei dati osservati.  $B_{01}$  rappresenta una misura della veridicità dell'ipotesi nulla, rispetto all'ipotesi alternativa, sulla base dei risultati sperimentali. Volendo ricavare dall'equazione sopra  $B_{01}$  si ottiene:  $B_{01} = O(H_0 | x)/O(H_0)$ , si arriva così alla seguente definizione.

**Definizione.** Si chiama fattore di Bayes, e si indica con  $B_{01}$ , il rapporto tra gli *odds* a posteriori e gli *odds* a priori.

- Il caso in cui  $B_{01} > 1$  corrisponde alla situazione in cui l'ipotesi nulla è supportata maggiormente dai dati rispetto ad  $H_1$  e il rapporto a posteriori è maggiore del rapporto a priori.
- Se invece  $B_{01} < 1$ , accade il contrario, ovvero è l'ipotesi alternativa ad essere supportata maggiormente dai dati e il rapporto a priori è maggiore del rapporto a posteriori.
- Infine se  $B_{01} = 1$  significa che le due ipotesi hanno ottenuto lo stesso supporto sperimentale.

Jeffreys (1961, app. B) propose una diversa interpretazione del fattore di Bayes, utilizzando la scala logaritmica in base 10, come si può vedere nella seguente tabella:

$\log_{10}(B_{01})$	$B_{01}$	Prova contro $H_1$
$0 \text{ to } \frac{1}{2}$	$1 \text{ to } 3.2$	Molto debole
$\frac{1}{2} \text{ to } 1$	$3.2 \text{ to } 10$	Sostanziale
$1 \text{ to } 2$	$10 \text{ to } 100$	Forte
$> 2$	$> 100$	Decisiva

Queste categorie non sono scelte su base teorica, ma sono state determinate attraverso numerose prove sperimentali.

Si utilizza più spesso, però, la trasformazione logaritmica con base naturale, anziché base 10, per essere sulla stessa scala del test rapporto di verosimiglianza. Infatti la trasformazione  $2\log_e(B_{01})$  corrisponde esattamente alla statistica test rapporto di verosimiglianza.

$2\log_e(B_{01})$	$B_{01}$	Prova contro $H_1$
$0 \text{ to } 2$	$1 \text{ to } 3$	Molto debole
$2 \text{ to } 6$	$3 \text{ to } 20$	Sostanziale
$6 \text{ to } 10$	$20 \text{ to } 150$	Forte
$> 10$	$> 150$	Decisiva

Kass e Raftery (1995) sostengono che queste classi forniscono una regola di decisione appropriata.

*Esempio (Kass e Wasserman, 1997)*

Nel 1919, durante un'eclisse solare, l'astronomo Eddington effettuò il seguente esperimento: da due posizioni diverse egli misurò il grado di piegatura della luce emessa in funzione della posizione intorno al sole. Egli effettuò  $n_A = 5$  misurazioni nel sito A e  $n_B = 7$  misurazioni nel sito B. La teoria di Newton, denotata qui con  $H_0$ , prevede una deflazione di luce di circa 0.875 secondi di arco. Al contrario, la teoria della relatività generale di Einstein, denotata con  $H_1$ , conduce ad una previsione di 1.75 secondi di arco. Le osservazioni

portarono ad un valore medio pari a  $\bar{x}_A = 1.98$  secondi, con un errore standard pari a  $s_A = 0.16$  per il punto di osservazione A, e  $\bar{x}_B = 1.61$  con errore standard pari a  $s_B = 0.40$  per il punto di osservazione B. Si può supporre che i dati raccolti in entrambi i siti abbiano distribuzione normale con media  $\mu$  e varianza  $\sigma^2$  incognite. Sia inoltre, a priori,  $\pi_0 = \pi_1 = 1/2$ , ovvero si da la stessa probabilità alle due teorie.

Si considera il primo esperimento in cui  $n_A = 5$ . Si assume che  $s_A$  sia una buona stima puntuale di  $\sigma/\sqrt{n_A}$ , quindi la distribuzione dei dati è  $N(\mu, 0.16^2)$ . Le ipotesi sono così specificate:

$$\begin{aligned} H_0 : \mu &= 0.875 \\ H_1 : \mu &= 1.75 . \end{aligned} \quad (3)$$

Il fattore di Bayes vale:

$$B_{01}^A = \frac{\frac{1}{0.16\sqrt{2\pi}} \exp\left\{-\frac{(1.98 - 0.875)^2}{2 * 0.16^2}\right\}}{\frac{1}{0.16\sqrt{2\pi}} \exp\left\{-\frac{(1.98 - 1.75)^2}{2 * 0.16^2}\right\}} = 1.23 * 10^{-10} .$$

Da questo valore si deduce che sono sufficienti solamente 5 osservazioni per avere un'evidenza inconfutabile in favore di  $H_1$  (odds pari a 1 su un milione); il fattore di Bayes però non afferma che l'ipotesi alternativa è vera, ma che i dati supportano tale ipotesi con un'evidenza un milione di volte superiore rispetto ad  $H_0$ .

Si considerano ora i dati relativi al secondo punto di osservazione. Le ipotesi sono ancora specificate come in (3).

Il fattore di Bayes vale:

$$B_{01}^B = \frac{\frac{1}{0.40\sqrt{2\pi}} \exp\left\{-\frac{(1.61 - 0.875)^2}{2 * 0.40^2}\right\}}{\frac{1}{0.40\sqrt{2\pi}} \exp\left\{-\frac{(1.61 - 1.75)^2}{2 * 0.40^2}\right\}} = 0.197 .$$

In questo caso la situazione è molto più incerta, in quanto il rapporto tra *odds* è solo di 5 a 1 in favore dell'ipotesi alternativa.

In conclusione si vogliono fare alcune considerazioni. Nel caso in cui le ipotesi specificate sono equiprobabili, come nell'esempio appena presentato, il fattore di Bayes è equivalente al rapporto delle probabilità a posteriori, il quale "rappresenta la "vera" risposta bayesiana al problema di verifica d'ipotesi" (Liseo,2004). Inoltre, nel caso in cui le due ipotesi sono semplici, esso corrisponde al rapporto delle verosimiglianze e non coinvolge nessuna componente soggettiva. Tale rapporto viene infatti utilizzato nella risoluzione di problemi di verifica d'ipotesi anche nell'impostazione di Neyman e Pearson.

#### 2.4 Ipotesi generali

Si vuole trattare il problema della verifica d'ipotesi in maniera più generale. Si considerino due sottoinsiemi  $\Theta_0$  e  $\Theta_1$  dello spazio dei parametri  $\Theta$ , i quali sono non vuoti e hanno le seguenti caratteristiche:  $\Theta_0 \cup \Theta_1 = \Theta$  e  $\Theta_0 \cap \Theta_1 = \emptyset$ . Le ipotesi a confronto sono così specificate:

$$H_0 : \theta \in \Theta_0$$

$$H_1 : \theta \in \Theta_1 ,$$

con probabilità a priori:

$$\pi_0 = P(H_0) = \int_{\Theta_0} \pi(\theta) d\theta$$

$$\pi_1 = P(H_1) = \int_{\Theta_1} \pi(\theta) d\theta$$

e densità di  $\theta$  condizionata a  $\theta \in \Theta_i$  pari a:

$$g_i(\theta) = \frac{\pi(\theta)}{\int_{\Theta_i} \pi(\theta) d\theta} , \quad i = 0,1 ,$$

dove con  $\pi(\theta)$  si denota la densità complessiva iniziale su tutto lo spazio dei parametri  $\Theta$ .

Nella situazione descritta, il fattore di Bayes risulta:

$$B_{01} = \frac{P(x|H_0)}{P(x|H_1)} = \frac{\int_{\Theta_0} \pi_0(\theta|x) d\theta}{\int_{\Theta_1} \pi_1(\theta|x) d\theta} = \frac{\int_{\Theta_0} L(\theta) g_0(\theta) d\theta}{\int_{\Theta_1} L(\theta) g_1(\theta) d\theta} .$$

Da questa formula si nota che il fattore di Bayes non dipende affatto dalle probabilità a priori  $\pi_0$  e  $\pi_1$ , bensì dipende in modo marginale dalla densità complessiva iniziale, la quale è coinvolta nel calcolo delle densità condizionate  $g_0(\theta)$  e  $g_1(\theta)$ , ed è estranea all'identificazione del rapporto tra le probabilità a priori relative alle due ipotesi.

Questo aspetto rende il fattore di Bayes uno strumento inferenziale “oggettivo” utile a risolvere problemi di verifica d'ipotesi senza introdurre a priori vincoli troppo impegnativi (Piccinato, 1997).

Inoltre si vuol far notare che il fattore di Bayes non ha più la forma del rapporto tra verosimiglianze, ma si tratta di verosimiglianze integrate (si veda il Capitolo 4), le quali possono introdurre delle difficoltà computazionali notevoli (si veda il Capitolo 3).

## 2.5 Ipotesi alternativa composta

Si propone ora il caso in cui all'ipotesi nulla semplice  $H_0 : \theta = \theta_0$  viene contrapposta l'ipotesi alternativa composta  $H_1 : \theta \neq \theta_0$ . Questo schema corrisponde al classico test di significatività, menzionato nel § 2.2.

Si assume che i dati siano realizzazioni di  $n$  variabili casuali  $X_1, \dots, X_n$  con densità di probabilità  $p(x; \theta)$ . Sia  $\pi_0$  la probabilità a priori assegnata all'ipotesi nulla  $H_0$ , la quale, essendo  $H_0$  semplice, è concentrata sul punto  $\theta_0$ . La probabilità a priori assegnata all'ipotesi alternativa  $H_1$ , invece, deve essere distribuita su tutti i valori di  $\theta$  diversi dal punto  $\theta_0$ . Volendo indicare la distribuzione iniziale complessiva si introduce la formula seguente:

$$\pi(\theta) = \pi_0 I_{\theta=\theta_0}(\theta) + (1 - \pi_0) g(\theta) I_{\theta \neq \theta_0}(\theta) ,$$

nella quale  $I_{\theta=\theta_0}(\theta)$  è una funzione indicatrice che dà probabilità 1 al punto  $\theta_0$  e vale 0 altrimenti,  $g(\theta)$  rappresenta la densità di probabilità a priori condizionata all'ipotesi  $H_1$ ,  $I_{\theta \neq \theta_0}(\theta)$  è una funzione indicatrice che vale 1 per valori di  $\theta$  diversi da  $\theta_0$  e vale 0 altrimenti.

Il rapporto delle probabilità a posteriori è:

$$\frac{P(H_0 | x)}{P(H_1 | x)} = \frac{\pi_0}{1 - \pi_0} \frac{L(\theta_0; x)}{\int_{\theta \neq \theta_0} L(\theta; x) g(\theta) d\theta} ,$$

e quindi il fattore di Bayes è pari a :

$$B_{01} = \frac{P(H_0 | x)}{P(H_1 | x)} \bigg/ \frac{\pi_0}{1 - \pi_0} = \frac{L(\theta_0; x)}{\int_{\theta \neq \theta_0} L(\theta; x) g(\theta) d\theta} . \quad (4)$$

Con riguardo alla (4) si possono porre alcuni problemi interpretativi, noti come *paradosso di Lindley-Jeffreys*. Infatti se la legge di probabilità a priori condizionata ad  $H_1$ , ossia  $g(\theta)$ , è diffusa su tutto lo spazio dei parametri  $\Theta - \{\theta_0\}$ , non concentrandosi su alcun punto in particolare, l'ipotesi  $H_0$  viene automaticamente favorita.

Per risolvere questo problema, Lindley e Jeffreys proposero due diverse soluzioni: Lindley sostiene di ricavare la regione di confidenza bayesiana del parametro con livello di significatività pari all' $(1 - \alpha)\%$  e di accettare l'ipotesi nulla se il valore  $\theta_0$  cade nella regione; Jeffreys, invece, osserva che in alcune situazioni il valore  $\theta_0$  ha una particolare importanza applicativa rispetto a tutti i rimanenti valori di  $\theta$  compresi nello spazio parametrico  $\Theta - \{\theta_0\}$ , come ad esempio il valore di controllo di un esperimento. Per questo motivo propone di scomporre la probabilità a priori in due parti: una probabilità discreta  $\pi(\theta_0)$  su  $\theta_0$  e una densità a priori continua  $\pi(\theta)$  per  $\theta \neq \theta_0$ . Poi, attraverso le osservazioni sperimentali, si calcolano le probabilità a posteriori  $\pi(\theta_0 | x)$  e

$\pi(\theta | x)$  per  $\theta \neq \theta_0$ . La regola per accettare o rifiutare l'ipotesi nulla si basa sul valore dell'*odds*:

$$\frac{\pi(\theta_0 | x)}{\int_{\theta \neq \theta_0} \pi(\theta | x) d\theta}$$

(Liseo, 2004 ; Cipollini e Stefanini, 1995).

## 2.6 Problema della scelta tra modelli

Il problema della scelta tra modelli è sostanzialmente un caso di verifica d'ipotesi e per questo motivo il metodo usato per risolvere tale situazione è basato sull'uso del fattore di Bayes.

Si suppone di confrontare due modelli per i dati  $x$  :

$$H_0 : p(x | \theta_0; M_0), \theta_0 \in \Theta_0$$

$$H_1 : p(x | \theta_1; M_1), \theta_1 \in \Theta_1 \quad .$$

Il fattore di Bayes per il modello  $M_0$  contro il modello  $M_1$  è dato dal rapporto:

$$B_{01} = \frac{m(x | M_0)}{m(x | M_1)} \quad ,$$

dove  $m(x | M_0)$  e  $m(x | M_1)$  rappresentano le leggi di probabilità dei risultati possibili condizionati al modello e sono definite nel modo seguente:

$$\begin{aligned} m(x | M_i) &= \int_{\Theta_i} p(x | \theta_i; M_i) \pi(\theta_i | M_i) d\theta_i = \\ &= \int_{\Theta_i} L(\theta_i | M_i; x) \pi(\theta_i | M_i) d\theta_i \quad , \quad i = 0, 1 \quad , \quad (5) \end{aligned}$$

dove  $\pi(\theta_i | M_i)$ , con  $i = 0, 1$  , rappresenta la legge di probabilità a priori dei due parametri  $\theta_0$  e  $\theta_1$  condizionatamente al modello e la funzione  $L(\theta_i | M_i; x)$  è la funzione di verosimiglianza associata al generico risultato sperimentale  $x$  quando si assume il modello  $M_i$  come valido (Piccinato, 1997).



Nei problemi di scelta tra modelli si pone un ostacolo relativo all'incertezza sul valore da assegnare ai parametri, non avendo la certezza neppure sul modello da adottare. Per aggirare l'ostacolo gli studiosi bayesiani propongono una procedura basata sul *campione di prova* e sul *fattore di Bayes parziale*. Il metodo consiste nel dividere il campione, teoricamente o effettivamente, in due parti, una delle quali è il cosiddetto campione di prova. Quest'ultimo ha lo scopo di "aggiornare" la distribuzione iniziale impropria del parametro, per renderla propria. Si ricorda che una distribuzione iniziale è propria, se l'integrale sul suo dominio è pari a 1, ossia

$$\int_{\Theta} p(\theta) d\theta = 1 .$$

Questo è il caso in cui si posseggono informazioni a priori sui parametri di interesse. Invece, con il termine di distribuzione di probabilità impropria si designa la distribuzione con "funzione di densità" il cui integrale, esteso a tutto il supporto, diverge (Cifarelli e Muliere, 1989). Questo tipo di distribuzioni vengono utilizzate quando non si hanno informazioni a priori.

Si vuole dare, ora, una breve spiegazione della procedura basata sul campione di prova e sul fattore di Bayes parziale.

Sia dato un campione casuale  $x$ , composto da  $n$  elementi. Esso viene diviso in due sottocampioni denominati  $y$  e  $z$  di dimensioni, rispettivamente,  $\ell$  e  $n - \ell$ . Il sottocampione  $y$  è il campione di prova. A questo punto vengono calcolati diversi tipi di fattori di Bayes, in base a quali sottocampioni e a quali distribuzioni di probabilità vengono presi in considerazione. In particolare:

- BF basato su tutti i dati  $\rightarrow B_{01}(x) = \frac{m(x | M_0)}{m(x | M_1)}$
- BF basato sul campione di prova  $\rightarrow B_{01}(y) = \frac{m(y | M_0)}{m(y | M_1)}$

- BF parziale, condizionato a  $y \rightarrow B_{01}(z | y) = \frac{m(z | y; M_0)}{m(z | y; M_1)}$ ,

dove la legge di probabilità dei risultati possibili condizionati al modello è data da:

$$m(z | y, M_i) = \int_{\Theta_i} L(\theta_i | M_i; x) \pi(\theta_i | y, M_i) d\theta_i$$

che viene messa a confronto con la formula di  $m(x | M_i)$  (5). Si nota la presenza della legge  $\pi(\theta_i | y, M_i)$  al posto della distribuzione iniziale  $\pi(\theta_i | M_i)$ . Il campione di prova  $y$  ha permesso di aggiornare quest'ultima in modo da ottenere  $\pi(\theta_i | y, M_i)$  presente nella formula di  $m(z | y, M_i)$ .

Per i tre fattori di Bayes introdotti sopra vale la seguente proprietà:

$$B_{01}(x) = B_{01}(y) B_{01}(z | y)$$

(Piccinato, 1997).

A questo punto però sorge un ulteriore problema legato alla scelta del campione di prova  $y$  e alla relativa dimensione  $\ell$ . Per eliminare la dipendenza dei risultati dalla suddivisione del campione e dalla scelta del campione di prova, gli studiosi bayesiani introducono due estensioni del fattore di bayes classico: il fattore di Bayes intrinseco e il fattore di Bayes frazionario. Questi argomenti saranno trattati nel prossimo capitolo.

## Capitolo 3

# Estensioni e aspetti computazionali legati al fattore di Bayes

### 3.1 Estensioni del Fattore di Bayes classico

Nel capitolo precedente si è visto come problemi di scelta tra modelli possono essere trattati allo stesso modo di problemi di verifica d'ipotesi. Sorgono però diverse problematiche che gli studiosi cercano di risolvere proponendo nuove forme del fattore di Bayes: il *fattore di Bayes intrinseco* e il *fattore di Bayes frazionario*. In questo paragrafo verranno brevemente illustrate tali procedure.

#### 3.1.1 Fattore di Bayes intrinseco

I problemi di scelta tra modelli vengono risolti grazie all'utilizzo di una procedura basata sul campione di prova (vedi § 2.6). Il campione casuale  $x$ , composto da  $n$  elementi, viene suddiviso in due sottocampioni denominati  $y$ , di dimensione  $\ell$ , e  $z$ , di dimensione  $n - \ell$ . Il sottocampione  $y$  è il cosiddetto campione di prova, il quale ha lo scopo di convertire la distribuzione iniziale impropria del parametro in distribuzione propria, utilizzata poi per calcolare il fattore di Bayes per i dati rimanenti ( $B_{01}(z|y)$ ). Questo risultato però dipende dal campione di prova scelto e dalla relativa dimensione. Per eliminare questa

dipendenza e per incrementare la stabilità dei risultati, una soluzione possibile è quella di considerare tutte le  $\binom{n}{\ell}$  suddivisioni possibili del campione  $x$ , calcolare il fattore di Bayes  $B_{01}(z|y)$  relativo ad ogni campione di prova ottenuto ed infine calcolarne una media. In questo modo si ottiene il *fattore di Bayes intrinseco* (IBF). Nella maggior parte dei casi viene calcolata la media aritmetica, ottenendo così AIBF, definito come

$$B_{01}^{AI} = B_{01}(x) \frac{1}{L} \sum_{\ell=1}^L B_{10}(y) ,$$

dove  $L = \binom{n}{\ell}$ . Oppure viene utilizzata la mediana, ottenendo così MIBF, definito da

$$B_{01}^{MI} = B_{01}(x) \text{Med}[B_{10}(y)] .$$

Generalmente si pone il modello più “complesso” a numeratore. Nelle precedenti definizioni si è assunto che  $M_0$  fosse il modello più complicato, di conseguenza viene definito  $B_{10}$  come:

$$B_{10} = \frac{1}{B_{01}} ,$$

essendo il fattore di Bayes un rapporto di verosimiglianze.

### 3.2.2 Fattore di Bayes frazionario

Un'alternativa al fattore di Bayes intrinseco è il fattore di *Bayes frazionario*. L'idea è di utilizzare una frazione della funzione di verosimiglianza, anziché una parte dei dati, per trasformare la distribuzione a priori del parametro da impropria a propria. Questo metodo si basa sull'equivalenza approssimata

$$[p(y|\theta_i, M_i)]^{1/\ell} \cong [p(x|\theta_i, M_i)]^{1/n} , \quad i = 0, 1 ,$$

dove  $\ell$  rappresenta la dimensione del campione di prova  $y$  e  $n$  la dimensione del campione casuale  $x$ . Successivamente si ottiene

$$p(y | \theta_i, M_i) \cong [p(x | \theta_i, M_i)]^b$$

con  $b = \ell/n$ .

Per costruire il fattore di Bayes viene utilizzata la frazione  $b$  di verosimiglianza, ottenendo così

$$B_{01}^b(x) = \frac{\int_{\Theta_0} [p(x | \theta_0, M_0)]^b \pi(\theta_0 | M_0) d\theta_0}{\int_{\Theta_1} [p(x | \theta_1, M_1)]^b \pi(\theta_1 | M_1) d\theta_1} \cong B_{01}(y).$$

Dall'equazione

$$B_{01}(x) = B_{01}(y) B_{01}(z | y)$$

si ricava

$$B_{01}(z | y) = \frac{B_{01}(x)}{B_{01}^b(x)}.$$

Ed essendo  $B_{01}(y) \cong B_{01}^b(x)$ , si ottiene come approssimazione del fattore di Bayes parziale  $B_{01}(z | y)$ , la quantità

$$B_{01}^F = \frac{B_{01}(x)}{B_{01}^b(x)},$$

chiamata fattore di *Bayes frazionario* (FBF).

### 3.2 Aspetti computazionali

Una delle maggiori difficoltà che si incontra nell'approccio bayesiano è senza dubbio l'aspetto computazionale. Nell'impostazione bayesiana ci si trova a svolgere elaborazioni computazionali tutt'altro che banali. Infatti nella maggior parte dei casi non si riesce ad ottenere, in una forma esplicita, la distribuzione

finale del parametro, o dei parametri, d'interesse ponendo così dei limiti allo sviluppo e alla diffusione delle tecniche bayesiane.

Fino agli anni '80 dello scorso secolo, problemi di questo tipo venivano aggirati usando delle approssimazioni analitiche, basate sul teorema del limite centrale, al posto di modelli eccessivamente semplificati, inadatti a rappresentare fenomeni complessi. Successivamente, grazie all'introduzione dei computer, si sono sviluppate, e tutt'ora si stanno sviluppando, tecniche di integrazione numerica adatte a risolvere i problemi computazionali dell'approccio bayesiano.

Ma la vera rivoluzione nella pratica statistica si è compita negli ultimi 15 anni (Liseo, 2004). Tecniche di tipo *Monte Carlo* e metodi *Monte Carlo basati sulle proprietà delle catene di Markov* sono i nomi di questa rivoluzione.

Nell'impostazione bayesiana il problema principale è quello di determinare la distribuzione finale del parametro d'interesse in situazioni particolarmente complesse, che in termini matematici coinvolge la valutazione di integrali del tipo:

$$I = \int_{\Theta} g(\theta)\pi(\theta)d\theta \quad ,$$

come si è visto nel Capitolo 2 per il calcolo del fattore di Bayes. I metodi *Monte Carlo* permettono di risolvere integrali come  $I$  tramite simulazione.

### 3.2.1 Il metodo Monte Carlo

Il metodo Monte Carlo prevede di calcolare l'integrale di una opportuna funzione di interesse, rispetto ad una determinata funzione di densità di probabilità, tramite la simulazione di valori pseudo-casuali indipendenti ed identicamente distribuiti da tale densità.

Si consideri l'integrale  $I$  e si supponga che  $\pi(\theta)$  sia una distribuzione di probabilità. Per definizione si ha che

$$I = E_{\pi}[g(\theta)] \quad .$$

Ovvero l'integrale  $I$  corrisponde al valore atteso della funzione  $g(\theta)$ , con  $\theta$  avente funzione di densità  $\pi(\theta)$ . E' quindi ragionevole stimare  $I$  con un suo

corrispondente campionario. A questo punto si simulano  $N$  valori, con  $N$  elevato,  $\theta_1, \dots, \theta_N$  dalla distribuzione  $\pi(\theta)$  e si calcola la seguente media campionaria:

$$\hat{I} = \frac{1}{N} \sum_{i=1}^N g(\theta_i) ,$$

ottenendo così una stima non distorta di  $I$ , per la legge forte dei grandi numeri.

### 3.2.2 Il metodo Monte Carlo Importance Sampling

Il metodo Monte Carlo Importance Sampling è una generalizzazione del metodo Monte Carlo e viene utilizzato nel caso in cui non è possibile simulare direttamente da  $\pi(\theta)$ . Tale metodo prevede di riscrivere l'integrale  $I$  nella forma:

$$I = \int_{\Theta} \frac{g(\theta)}{f(\theta)} f(\theta) \pi(\theta) d\theta ,$$

in cui  $f(\theta)$  è una funzione di densità di probabilità definita in un intervallo i cui estremi coincidono con gli estremi di integrazione di  $I$ . In questo caso il campione di valori pseudo-casuali  $\theta_1, \dots, \theta_N$  viene simulato dalla distribuzione  $f(\theta)$  e un'approssimazione di  $I$  si ottiene con la formula seguente:

$$\hat{I} = \frac{1}{N} \sum_{i=1}^N \frac{g(\theta_i)}{f(\theta_i)} \pi(\theta_i) .$$

Può accadere, però, che si abbiano delle difficoltà ad ottenere la distribuzione finale  $\pi(\theta)$ , oppure a simulare il campione di valori pseudo-casuali perché la funzione  $f(\theta)$  non risulta essere affidabile. In queste situazioni si ricorre al metodo Monte Carlo basato sulle proprietà delle catene di Markov, le quali danno la certezza che il campione sia effettivamente simulato dalla densità di probabilità  $f(\theta)$  (si rimanda al § 3.2.3).

### 3.2.3 Metodo Monte Carlo basato sulle catene di Markov

Si consideri nuovamente un integrale del tipo

$$I = \int_{\Theta} g(\theta) \pi(\theta) d\theta$$

e si supponga di non essere in grado né di risolverlo analiticamente, né di risolverlo con i metodi Monte Carlo e Monte Carlo Importance Sampling, perché risulta impossibile simulare valori pseudo-casuali da  $\pi(\theta)$  o da una densità  $f(\theta)$  che somigli a  $\pi(\theta)$ . Vengono utilizzati, perciò, metodi Monte Carlo basati sulle proprietà delle catene di Markov (MCMC). I metodi MCMC permettono di simulare valori pseudo-casuali, i quali possono essere considerati, in maniera approssimata, delle realizzazioni indipendenti dalla densità di probabilità  $\pi(\theta)$ . Deve essere costruita, quindi, una catena di Markov  $\theta_1, \dots, \theta_N$ , con distribuzione limite  $\pi(\theta)$  e velocità di convergenza verso tale limite più alta possibile, da cui simulare  $N$  valori da considerare approssimativamente realizzazioni quasi indipendenti di una variabile casuale con distribuzione  $\pi(\theta)$ . Di conseguenza è possibile costruire lo stimatore

$$\hat{I} = \frac{1}{N} \sum_{i=1}^N g(\theta_i)$$

che risulta essere uno stimatore consistente per  $I$ .

Per una esauriente e precisa trattazione di Catene di Markov e metodi Monte Carlo basati sulla proprietà delle catene di Markov si rimanda a Liseo (2004), Casella e Robert (1999) e Barbieri (1996).



## Capitolo 4

# La verosimiglianza integrata

### 4.1 Introduzione

Nei problemi di verifica di ipotesi si è visto che il fattore di Bayes prende la forma di rapporto tra verosimiglianze di tipo integrato (si veda § 2.4). L'utilizzo di questo tipo di verosimiglianze deriva dalla necessità di fare inferenza solamente sul parametro d'interesse, nei casi in cui ci sia la presenza anche di parametri di disturbo. Questi ultimi, infatti, rendono difficile qualsiasi tipo di approccio inferenziale sul parametro d'interesse. L'impostazione bayesiana prevede, quindi, l'utilizzo di verosimiglianze integrate per eliminare i parametri di disturbo. Si consideri, ad esempio, un modello statistico in cui il parametro  $\theta$  è così partizionato:  $\theta = (\psi, \lambda)$ . Il parametro  $\psi$  rappresenta il parametro d'interesse, mentre  $\lambda$  rappresenta il parametro di disturbo. Inoltre il modello è rappresentato dalla funzione di verosimiglianza  $L(\psi, \lambda)$  e dalla probabilità a priori congiunta  $\pi(\psi, \lambda)$ . La probabilità a posteriori risulta essere:

$$\pi(\psi, \lambda | y) \propto \pi(\psi, \lambda) L(\psi, \lambda)$$

e scomponendo la priori si ottiene

$$\pi(\psi, \lambda | y) \propto \pi(\psi) \pi(\lambda | \psi) L(\psi, \lambda) ,$$

dove  $\pi(\lambda|\psi)$  rappresenta un'appropriata probabilità a priori di  $\lambda$  dato  $\psi$ .  
 Volendo fare inferenza solo sul parametro d'interesse  $\psi$ , è necessario eliminare il parametro di disturbo  $\lambda$ , dalla funzione di verosimiglianza completa  $L(\psi, \lambda)$ . Nell'approccio bayesiano ciò avviene integrando in  $\lambda$  la funzione descritta poc'anzi, ovvero

$$\pi(\psi | y) \propto \int_{\Lambda} \pi(\psi) \pi(\lambda | \psi) L(\psi, \lambda) d\lambda .$$

Poiché la probabilità a priori  $\pi(\psi)$  non dipende da  $\lambda$ , può essere portata fuori dall'integrale e si ottiene

$$\pi(\psi | y) \propto \pi(\psi) \int_{\Lambda} \pi(\lambda | \psi) L(\psi, \lambda) d\lambda$$

e la verosimiglianza integrata è data da

$$L_I(\psi) = \int_{\Lambda} \pi(\lambda | \psi) L(\psi, \lambda) d\lambda \quad (6) .$$

La (6) viene trattata come una vera e propria funzione di verosimiglianza e la posteriori per  $\psi$  può essere scritta come

$$\pi(\psi | y) \propto \pi(\psi) L_I(\psi) . \quad (7)$$

L'utilizzo della verosimiglianza integrata, però, presenta due principali difficoltà: la scelta di un'opportuna probabilità a priori  $\pi(\lambda|\psi)$  ed il calcolo dell'integrale (6) quando la dimensione del parametro di disturbo  $\lambda$  è elevata, anche se per quest'ultimo punto vengono in aiuto i metodi Monte Carlo e i metodi Monte Carlo basati sulle proprietà delle catene di Markov (si veda il § 3.2).

Nell'approccio frequentista, invece, il parametro di disturbo viene eliminato dalla funzione di verosimiglianza completa  $L(\psi, \lambda)$  passando a delle appropriate funzioni di pseudo-verosimiglianza. Una funzione di pseudo-verosimiglianza,

$L_{ps}(\psi)$ , è una funzione del solo parametro d'interesse  $\psi$  e dei dati, con proprietà simili a quelle di una funzione di verosimiglianza propria. Alcuni esempi di pseudo-verosimiglianze sono la verosimiglianza marginale, la verosimiglianza condizionata, la verosimiglianza profilo e la quasi verosimiglianza (si veda Pace e Salvani, 1996, Capitolo 4).

Un recente filone di ricerca prevede la possibilità di utilizzare appropriate funzioni di pseudo-verosimiglianza, al posto della verosimiglianza di tipo integrato, nell'inferenza bayesiana, permettendo di scrivere la probabilità a posteriori nel modo seguente:

$$\pi_{ps}(\psi | y) \propto \pi(\psi) L_{ps}(\psi) .$$

Questo modo di procedere, però, non è del tutto appropriato in ambito bayesiano, specialmente se si deve fare inferenza su un parametro d'interesse  $\psi$  con probabilità a priori di tipo non informativo, in quanto  $L_{ps}(\psi)$  non è derivata da una densità:

$$L_{ps}(\psi) \neq p(y; \theta) ,$$

quindi non vale il teorema di Bayes.

## 4.2 Un'applicazione

Si vuole ora provare ad usare il fattore di Bayes con un'opportuna  $L_{ps}(\psi)$  in un problema di scelta tra modelli.

Si consideri il seguente modello di regressione

$$y = X\beta + \sigma\varepsilon ,$$

dove  $X$  è la matrice dei regressori, di dimensione  $n \times p$  e di rango  $p$ ,  $\beta \in \mathbb{R}^p$  è il vettore dei coefficienti di regressione,  $\sigma > 0$  è il parametro di scala e  $\varepsilon$  rappresenta il vettore degli errori. Si assume che  $\varepsilon_1, \dots, \varepsilon_n$  siano indipendenti ed

identicamente distribuiti da una distribuzione  $p_k(.)$  nota. Il parametro  $k$  indicizza i diversi tipi di distribuzione, ad esempio:

- $k = 0$  normale
- $k = 1$  t-Student
- $k = 2$  valori estremi
- $k = 3$  gamma
- $k = 4$  Cauchy

e così via.

Si pone quindi un importante problema inferenziale, ovvero selezionare la distribuzione degli errori alla luce dei dati osservati. Tale problema si traduce, ad esempio, in una verifica d'ipotesi del tipo:

$$H_0 : k = 0$$

$$H_1 : k = 1 .$$

Si deduce che il parametro d'interesse è  $k$ , mentre  $\beta$  e  $\sigma$  rappresentano i parametri di disturbo.

Nell'impostazione frequentista vengono utilizzate due diverse procedure per affrontare problemi di verifica d'ipotesi come questo: il test rapporto di verosimiglianza (LR) e il test più potente tra gli invarianti (MPI). Il test LR è basato sul rapporto delle verosimiglianze specificate dalle ipotesi, nelle quali i parametri ignoti sono sostituiti con le loro stime di massima verosimiglianza. Questa procedura ha il vantaggio di essere facilmente applicabile in tutte le situazioni di interesse pratico ed inoltre non implica difficoltà computazionali. Mentre, il test MPI è basato sul rapporto delle verosimiglianze marginali, le quali sono rappresentate da integrali multidimensionali che possono portare difficoltà di calcolo.

In ambito bayesiano, per risolvere questi test di verifica d'ipotesi, si ricorre al calcolo del fattore di Bayes, che prende la forma del rapporto tra verosimiglianze integrate:

$$B_{01} = \frac{L_I(k=0)}{L_I(k=1)} . \quad (8)$$

La verosimiglianza di tipo integrato, però, può essere sostituita da una pseudo-verosimiglianza la quale ha le stesse proprietà e che permette di giungere agli stessi risultati. Il rapporto (8) coincide, per una specifica scelta della distribuzione a priori, ossia

$$\pi(\beta, \sigma | k) \propto \frac{1}{\sigma} ,$$

con il rapporto di queste nuove pseudo-verosimiglianze, chiamate verosimiglianze marginali  $L_m(k)$ , date da:

$$L_m(k) = \int_{\mathcal{R}^+} \int_{\mathcal{R}^p} \frac{1}{\sigma} L(k, \beta, \sigma) d\beta d\sigma \quad (9) .$$

La probabilità a posteriori può essere quindi scritta come:

$$\pi(k | y) \propto \pi(k) L_m(k)$$

ed è stato dimostrato che essa corrisponde esattamente alla (7), ossia alla distribuzione a posteriori ottenuta utilizzando la verosimiglianza integrata (6), basata sulla distribuzione completa dei dati  $y$  e sulla probabilità a priori condizionata e non informativa  $\pi(\beta, \sigma | k) = 1/\sigma$ .

In altre parole, il fattore di Bayes può essere scritto come rapporto tra verosimiglianze marginali

$$B_{01} = \frac{L_m(k=0)}{L_m(k=1)} .$$

### 4.3 Un'applicazione con dati reali

#### *4.3.1 Descrizione del problema*

Si considera il data set oggetto di studio di Sen e Srivastava (1990, pag. 32) riguardante i costi delle case. Esso è composto da  $n = 26$  osservazioni raccolte su alcune variabili d'interesse:

- *price* ( $y$ ): prezzo di vendita in migliaia di dollari;
- *bdroom* ( $x_1$ ): numero di camere da letto;
- *floor* ( $x_2$ ): dimensione della casa in piedi quadrati;
- *rooms* ( $x_3$ ): numero di stanze totali;

Si consideri il seguente modello per spiegare la variabile risposta *price*:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \sigma \varepsilon_i, \quad i = 1, \dots, 26. \quad (10)$$

Il problema che si pone ora è quello di determinare la distribuzione  $p_k(\cdot)$  degli errori  $\varepsilon_i$  alla luce dei dati osservati. Due possibili alternative da porre a confronto sono la distribuzione *normale* e la distribuzione *t di Student* con, ad esempio, due gradi di libertà. Questo schema si traduce nel seguente test di verifica d'ipotesi:

$$\begin{aligned} H_0 : k = 0 & \quad \text{Normale} \\ H_1 : k = 1 & \quad \text{t-Student con 2 g.d.l.} \end{aligned} \quad (11)$$

Le ipotesi (11) possono essere viste come due ipotesi semplici, di conseguenza il fattore di Bayes prende la forma di rapporto tra funzioni di verosimiglianza. In questa situazione è sembrato opportuno calcolare il fattore di Bayes per l'ipotesi  $H_1$  contro l'ipotesi nulla  $H_0$ , ovvero  $B_{10}$ , ma nulla sarebbe cambiato se si fosse calcolato  $B_{01}$ , come è stato fatto fin'ora, in quanto si ricorda che vale la relazione:

$$B_{10} = \frac{1}{B_{01}}.$$

#### 4.3.2 Descrizione delle procedure impiegate

Per ottenere il fattore di Bayes si è proceduto sia nel modo classico, ovvero utilizzando la verosimiglianza integrata, sia coinvolgendo due tipi di pseudo-verosimiglianze: la profilo e la marginale.

Inoltre si vuole aggiungere che in situazioni pratiche è più semplice lavorare con la funzione di log-verosimiglianza, anziché con la verosimiglianza. Si definisce quindi la log-verosimiglianza per i parametri  $(\beta, \sigma, k)$  relativa al modello (10):

$$l(\beta, \sigma, k) = -n \log \sigma + \sum_{i=1}^n p_k(\beta, \sigma) \quad , \quad k = 0, 1 \quad .$$

Si passa ora a descrivere i tre approcci elencati poc'anzi.

- *BF con verosimiglianza integrata.*

Il fattore di Bayes è dato dal rapporto delle verosimiglianze integrate, come la (6):

$$B_{10}^I = \frac{L_I(k=1)}{L_I(k=0)} \quad .$$

Scegliendo la priori pari a  $1/\sigma$  si ottiene:

$$B_{10}^I = \frac{L_I(k=1)}{L_I(k=0)} = \frac{\int_{\mathfrak{R}^+} \int_{\mathfrak{R}^p} \frac{1}{\sigma} \exp(l(\beta, \sigma, 1)) d\beta d\sigma}{\int_{\mathfrak{R}^+} \int_{\mathfrak{R}^p} \frac{1}{\sigma} \exp(l(\beta, \sigma, 0)) d\beta d\sigma} \quad . \quad (12)$$

- *BF con pseudo-verosimiglianza marginale.*

Il fattore di Bayes ha la seguente forma:

$$B_{10}^m = \frac{L_m(k=1)}{L_m(k=0)} \quad . \quad (13)$$

Con riferimento alla (9) si ottiene:

$$B_{10}^m = \frac{\int_{\mathfrak{R}^+} \int_{\mathfrak{R}^p} \frac{1}{\sigma} \exp(l(\beta, \sigma, 1)) d\beta d\sigma}{\int_{\mathfrak{R}^+} \int_{\mathfrak{R}^p} \frac{1}{\sigma} \exp(l(\beta, \sigma, 0)) d\beta d\sigma} \quad . \quad (14)$$

Avendo stabilito la probabilità a priori pari a  $1/\sigma$ , la (14) coincide con la (12) (si veda § 4.2).

Per risolvere gli integrali a numeratore e denominatore della (14) è stato utilizzato il metodo analitico di *Laplace*, ottenendo così la seguente approssimazione:

$$L_m^\diamond(k) = \frac{(2\pi)^{(p+1)/2}}{\hat{\sigma}_k |j(\hat{\beta}_k, \hat{\sigma}_k)|^{-(1/2)}} \exp(l_p(k)) \quad , \quad k = 0,1 \quad (15)$$

dove con  $|j(\hat{\beta}_k, \hat{\sigma}_k)|$  si indica il determinante della matrice d'informazione. La (15), inoltre, viene interpretata come verosimiglianza profilo modificata (cfr. Pace, et al., 2005). Il fattore di Bayes può quindi essere riscritto nel modo seguente:

$$B_{10}^\diamond = \frac{L_m^\diamond(k=1)}{L_m^\diamond(k=0)} \quad . \quad (16)$$

- *BF con pseudo-verosimiglianza profilo.*

Il fattore di Bayes è così specificato:

$$B_{10}^p = \frac{L_p(k=1)}{L_p(k=0)} = \frac{\exp(l_p(k=1))}{\exp(l_p(k=0))} = \exp(l_p(k=1) - l_p(k=0)) \quad , \quad (17)$$

dove  $l_p(k)$  rappresenta la log-verosimiglianza profilo ed è così espressa:

$$l_p(k) = l(\hat{\beta}_k, \hat{\sigma}_k, k) \quad , \quad k = 0,1 \quad . \quad (18)$$

Nella (18)  $\hat{\beta}_k$  e  $\hat{\sigma}_k$  sono rispettivamente le stime di massima verosimiglianza dei parametri  $\beta$  e  $\sigma$  con  $k$  fissato.



La regola decisionale scelta per rispondere al quesito posto dal test è quella descritta nel Capitolo 2 per  $B_{01}$ . Riadattandola per interpretare il fattore di Bayes dell'ipotesi  $H_1$  contro l'ipotesi nulla  $H_0$ , ovvero  $B_{10}$ , si ottiene:

$2\log_e(B_{10})$	$B_{10}$	Prova contro $H_0$
0 to 2	1 to 3	Molto debole
2 to 6	3 to 20	Sostanziale
6 to 10	20 to 150	Forte
> 10	> 150	Decisiva

#### 4.3.3 Risultati ottenuti

Il fattore di Bayes ottenuto come rapporto delle verosimiglianze marginali approssimate con il metodo analitico di *Laplace*, (16), e il fattore di Bayes ottenuto come rapporto delle verosimiglianze profilo, (17), sono stati calcolati utilizzando delle funzioni scritte ad hoc in R (si rimanda all' Appendice A2).

I risultati possono essere riassunti nella seguente tabella:

$2\log_e(B_{10}^p)$	$2\log_e(B_{10}^\diamond)$	$2\log_e(B_{10}^I)$
2.78	3.96	5.41

I tre valori ottenuti vanno nella stessa direzione, ovvero cadono tutti nella seconda classe decisionale, che porta a rifiutare l'ipotesi nulla di distribuzione normale in modo sostanziale.

Si vuol aggiungere, però, che il fattore di Bayes ottenuto dal rapporto delle verosimiglianze integrate, calcolate utilizzando la libreria *adapt* di R, restituisce un valore numericamente instabile (si veda Pace, Salvan, Ventura, 2005), in quanto la dimensione  $p$  dello spazio parametrico è pari a quattro, un valore elevato e che non permette di approssimare gli integrali nel modo migliore.

Confrontando, quindi, solo  $B_{10}^P$  e  $B_{10}^\diamond$  si nota che entrambi rifiutano l'ipotesi nulla, ma  $B_{10}^\diamond$  lo fa con maggiore evidenza.

## Appendice

### A1 I dati

Data set “pricedata”:

	price	bdroom	floor	rooms	front
1	53	2	967	5	39
2	55	2	815	5	33
3	56	3	900	5	35
4	58	3	1007	6	24
5	64	3	1100	7	50
6	44	4	897	7	25
7	49	5	1400	8	30
8	70	3	2261	6	29
9	72	4	1290	8	33
10	82	4	2104	9	40
11	85	8	2240	12	50
12	45	2	641	5	25
13	47	3	862	6	25
14	49	4	1043	7	30
15	56	4	1325	8	50
16	60	2	782	5	25
17	62	3	1126	7	30
18	64	4	1226	8	37
19	66	2	929	5	30
20	35	4	1137	7	25
21	38	3	743	6	25
22	43	3	596	5	50
23	46	2	803	5	27
24	46	2	696	4	30
25	50	2	691	6	30
26	65	3	1023	7	30

## A2 Le funzioni in R

Verranno illustrate, in questa sezione, le funzioni in R scritte per risolvere il problema di verifica d'ipotesi posto nel § 4.3 e verranno elencati i comandi utilizzati per giungere ai risultati ottenuti.

La libreria *adapt* di R permette di risolvere con metodi numerici gli integrali come quelli presenti nella formula (12).

```
library(adapt)
```

```
dati<-read.table("c:/pricedata.txt",header=T)
attach(dati)
```

Si costruisce il modello lineare definito dalla (10):

```
reg <- lm(price ~ bdroom + floor, x=T)
```

Identificazione della variabile risposta:

```
ydat <- price
```

e della matrice dei regressori:

```
xmat <- reg$x
```

```
n <- nrow(xmat)
```

Calcolo di  $(X^T X)^{-1}$ :

```
xinv <- solve(t(xmat)%*%xmat)
```

Calcolo di  $(X^T X)^{-1} X^T$ :

```
xmat1 <- xinv%*%t(xmat)
```

Dimensione dello spazio dei parametri:

```
p <- ncol(xmat)
```

Gradi di libertà della distribuzione t-Student:

```
v <- 2
```

- *Funzioni per  $H_0$*

Identificazione di  $\hat{\beta}$ :

```
hatbeta <- reg$coef
```

Funzione che permette di ottenere  $\hat{\sigma}$ :

```
hats.h0 <- function(b,y){  
  sqrt((1/n)*sum((y-xmat%*%b)^2)) }
```

```
ss0 <-hats.h0(hatbeta,ydat)
```

Funzione per calcolare il determinante della matrice d'informazione:

```
detjh0 <- function(s){  
  out <- array(0,dim=c(p+1,p+1))  
  for(i in 1:p){ for(j in 1:p) {  
    out[i,j] <- (1/s^2)*sum(xmat[,i]*xmat[,j]) }}  
  for(i in 1:p){  
    out[i,p+1] <- 0  
    out[p+1,i] <- 0 }  
  out[p+1,p+1]<- 2*n/s^2  
  det(out) }
```

Funzione per il calcolo della log-verosimiglianza profilo:

```
lp.h0 <- function(s){  
  -n*log(s)-n/2 }
```

Funzione per il calcolo della log-verosimiglianza marginale:

```
lm.h0 <- function(s){  
  lp <- lp.h0(s)  
  det <- detjh0(s)  
  lp-log(s)-0.5*log(det) }
```

Funzione per il calcolo della log-verosimiglianza integrata:

```
lint.0 <- function(s) {  
  -(n-1)*log(s) }
```

- *Funzioni per  $H_1$*

Funzione che permette di ottenere la funzione obiettivo per la distribuzione t-Student con due gradi di libertà:

```
fun.ob <- function(par,y){
  b <- par[1:p]
  s <- par[p+1]
  n*log(s)+0.5*(v+1)*sum(log(1+(1/v)*((y-
xmat%%b)/s)^2)) }
```

Funzione per calcolare le stime di massima verosimiglianza dei parametri:

```
maxxx <- function(p0,y) {
  fun.ob1 <- function(par){fun.ob(par,y)}
  opt <- optim(p0,fun.ob1)
  opt$par }
```

Funzione per ottenere il determinante della matrice d'informazione:

```
detjh1 <- function(b,s,y){
  out <- array(0,dim=c(p+1,p+1))
  a <- (y-xmat%%b)/s
  a1 <- 1/(1+(a^2)/v)
  c1 <- (1/s^2)*((v+1)/v)
  for(i in 1:p){ for(j in 1:p) {
    out[i,j] <- c1*sum(a1*xmat[,i]*xmat[,j]) -
(2/v)*c1*sum(a1^2*a^2*xmat[,i]*xmat[,j]) }}
  for(i in 1:p){
    out[i,p+1] <- -(2/v)*c1*sum(a1^2*a^3*xmat[,i])
    out[p+1,i] <- -(2/v)*c1*sum(a1^2*a^3*xmat[,i]) }
  out[p+1,p+1] <- 2*n/s^2-(2/v)*c1*sum(a1^2*a^4)
  det(out) }
```

Funzione per il calcolo della log-verosimiglianza profilo:

```
lp.h1 <- function(b,s,y){
  p <- c(b,s)
  -fun.ob(p,y) }
```

Funzione per il calcolo della log-verosimiglianza marginale:

```
lm.h1 <- function(b,s,y){
  lp <- lp.h1(b,s,y)
  det <- detjh1(b,s,y)
  lp-log(s)-0.5*log(det) }
```

Funzione per il calcolo della log-verosimiglianza integrata:

```
lint.l <- function(par,y) {  
  c1 <- log(par[p+1])  
  c2 <- fun.ob(par,y)  
  exp(-c2-c1) }
```

- *Calcoli che conducono alla costruzione dei fattori di Bayes*

Stima di massima verosimiglianza,  $\hat{\beta}$ , sotto l'ipotesi nulla:

```
betah0 <- reg$coef
```

Stima di massima verosimiglianza,  $\hat{\sigma}$ , sotto l'ipotesi nulla:

```
sigh0 <- hats.h0(betah0,ydat)
```

Calcolo del determinante della matrice d'informazione sotto  $H_0$ :

```
jh0 <- detjh0(sigh0)
```

Calcolo del valore della log-verosimiglianza profilo sotto  $H_0$ :

```
lp0 <- lp.h0(sigh0)
```

Calcolo del valore della log-verosimiglianza marginale sotto  $H_0$ :

```
lm0 <- lm.h0(sigh0)
```

Calcolo del valore della log-verosimiglianza integrata sotto  $H_0$ :

```
esl <- lint.0(sigh0)
```

Definizione dei valori iniziali utili al calcolo delle stime di massima verosimiglianza sotto  $H_1$ :

```
ini <- c(betah0,sigh0)
```

Calcolo delle stime di massima verosimiglianza dei parametri sotto  $H_1$ :

```
mlep <- maxxx(ini,ydat)
mlebh1 <- mlep[1:p]
mlesh1 <- mlep[p+1]
```

Calcolo del determinante della matrice d'informazione sotto  $H_1$ :

```
jph1 <- detjh1(mlebh0,mlesh0,ydat)
```

Calcolo del valore della log-verosimiglianza profilo sotto  $H_1$ :

```
lpph1 <- lp.h1(mlebh0,mlesh0,ydat)
```

Calcolo del valore della log-verosimiglianza marginale sotto  $H_1$ :

```
lmph1 <- lm.h1(mlebh0,mlesh0,ydat)
```

Calcolo del valore della log-verosimiglianza integrata sotto  $H_1$ :

```
es11 <- adapt(p+1,c(20,-540,-
50,0.01),c(80,50,50,50),100,5000,lint.1,y=ydat)
```

Calcolo dei fattori di Bayes nella forma:

$$\begin{aligned}
 B_{10}^{trasf} &= 2 \log_e \left( \frac{L_*(k=1)}{L_*(k=0)} \right) \\
 &= 2 [\log_e (L_*(k=1)) - \log_e (L_*(k=0))] \\
 &= 2 [l_*(k=1) - l_*(k=0)]
 \end{aligned}$$

dove con l'asterisco si indicano i diversi tipi di verosimiglianze impiegate: profilo, marginale e integrata.



Fattore di Bayes con log-verosimiglianza profilo:

```
tp0 <- 2*(lpph1-lp0)
```

Fattore di Bayes con log-verosimiglianza marginale:

```
tm0 <- 2*(lmph1-lm0)
```

Fattore di Bayes con log-verosimiglianza integrata:

```
te0 <- 2*(log(es11$value)-es1)
```

## Bibliografia

Barbieri, M. M. (1996), *Metodi MCMC nell'inferenza statistica*, CISU, Roma.

Berger, J., Pericchi, L. (2001), Objective Bayesian methods for model selection: introduction and comparison, *In Model Selection*, P. Lahiri, ed., Institute of Mathematical Statistics Lecture Notes – Monograph Series, vol. 38, Beachwood Ohio, 135–207.

Casella, G., Robert, C. (1999), *Monte Carlo statistical methods*, Springer, New York.

Cifarelli, D., Muliebre, P. (1989), *Statistica bayesiana*, Giapichelli, Torino.

Cipollini, F., Stefanini, F. M. (1995), *Appunti di inferenza statistica bayesiana*, [www.ds.unifi.it](http://www.ds.unifi.it) .

Kass, R. E., Raftery, A. E. (1995), Bayes Factors, *Journal of the American Statistical Association*, vol. 90, No. 430, 773-795.

Liseo, B. (2004), *Introduzione alla statistica bayesiana*, manoscritto, Roma.

Pace, L., Salvan, A. (1996), *Teoria della statistica*, Cedam, Padova.

Pace, L., Salvan, A., Ventura, L. (2005), Likelihood based discrimination between separate scale and regression models, *Journal of Statistical Planning and Inference*.

Piccinato, L. (1997), *Il fattore di Bayes come strumento pratico di statistica applicata*, Dipartimento di Statistica, Probabilità e Statistiche applicate, Università degli studi "La Sapienza", Roma.

Robert, C. (2001), *The Bayesian choice*, Springer, New York.

Sen, A., Srivastava, M. (1990), *Regression Analysis: Theory, Methods and Applications*, Springer-Verlag, New York.

[www.diee.unica.it](http://www.diee.unica.it), *Stimatori Bayesiani*.