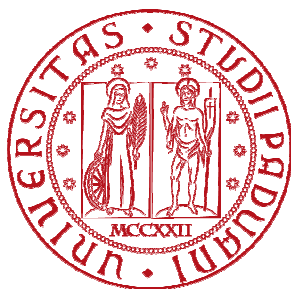


**UNIVERSITA' DEGLI STUDI DI PADOVA**

**CORSO DI LAUREA IN SCIENZE STATISTICHE  
ECONOMICHE FINANZIARIE E AZIENDALI**



**APPROSSIMAZIONI ASINTOTICHE DI ORDINE ELEVATO  
PER VERIFICHE D'IPOTESI BAYESIANE:  
UNO STUDIO PER DATI DI SOPRAVVIVENZA**

Relatore: Ch.ma Prof.ssa Laura Ventura

Dipartimento di Scienze Statistiche

Laureanda: Anna Pinto

ANNO ACCADEMICO 2011/2012



# INDICE

<b>Introduzione</b> .....	1
<b>1 Metodi di analisi della sopravvivenza</b> .....	3
1.1 Introduzione all'analisi della sopravvivenza .....	3
1.2 Alcune definizioni .....	5
1.3 Metodi non parametrici .....	7
1.3.1 Il metodo delle tavole di sopravvivenza .....	8
1.3.2 Lo stimatore di Kaplan-Meier .....	9
1.3.3 Il Log-rank test .....	11
1.4 Modelli parametrici di regressione .....	12
1.5 Il modello di regressione di Cox .....	15
1.6 Test disponibili per la valutazione dei modelli .....	18
1.7 Conclusioni .....	19
<b>2 Analisi Bayesiana della sopravvivenza</b> .....	21
2.1 Introduzione .....	21
2.2 Scelta della distribuzione a priori .....	22
2.3 Metodi parametrici per l'analisi bayesiana della sopravvivenza .....	25
2.4 Alcune riflessioni .....	31
<b>3 La misura di evidenza di Pereira-Stern e uno studio di simulazione</b> .....	33
3.1 Introduzione .....	33
3.2 Il fattore di Bayes .....	34

3.3	La procedura di Pereira-Stern .....	36
3.3.1	<i>Tail area approximation</i> .....	38
3.4	Approssimazioni asintotiche per la procedura di Pereira-Stern .....	41
3.4.1	Esempio: <i>matching priors</i> .....	42
3.5	Studi di simulazione e un esempio .....	44
3.5.1	L'analisi .....	45
 <b>Conclusioni</b> .....		51
 <b>Bibliografia</b> .....		53

## INTRODUZIONE

L'analisi della sopravvivenza si riferisce a tutti quei metodi statistici che studiano la distribuzione del *tempo di comparsa di un evento*.

Essa può riguardare diversi ambiti. In quello epidemiologico, ad esempio, l'evento d'interesse può essere non solo la morte di un paziente, ma anche la ricaduta in seguito a una malattia o la risposta di un paziente a un trattamento a cui è stato sottoposto. Si parla di analisi di sopravvivenza anche in casi in cui si vuole studiare, ad esempio, il tempo necessario a ex carcerati per commettere un crimine una volta che sono stati rilasciati, oppure, molto semplicemente, quando si vuole studiare la durata di una certa qualità di lampadine.

In generale, nella letteratura dell'analisi della sopravvivenza, la morte, come nell'ambito epidemiologico, o il guasto, sono considerati l'"evento d'interesse".

Il primo passo di un'analisi della sopravvivenza è il calcolo del "tempo di sopravvivenza", determinato dalla differenza tra il tempo in cui è avvenuto l'evento e il momento dell'ingresso nello studio dell'unità statistica.

Gli assunti che vengono fatti sul tempo di comparsa dell'evento permettono di classificare le procedure di analisi della sopravvivenza in metodi non parametrici, metodi semi-parametrici e metodi parametrici.

In questa tesi, l'attenzione è focalizzata sui metodi di regressione parametrici per l'analisi bayesiana della sopravvivenza. In tale contesto, è di interesse affrontare il problema della verifica di ipotesi di nullità di un singolo coefficiente di regressione allo scopo di valutare l'effetto di un regressore sulla sopravvivenza.

E' noto che, in ambito bayesiano, la verifica di un siffatto sistema d'ipotesi tramite il fattore di Bayes può condurre al paradosso di Jeffreys-Lindley (Liseo, 2008). In questa tesi si presenta un'alternativa al fattore di Bayes, data dalla *misura di evidenza di Pereira-Stern* (Pereira e Stern, 1999). Essa permette di misurare e quantificare l'evidenza a favore di una certa ipotesi nulla, a partire dalla sola distribuzione a posteriori per il parametro d'interesse.

Obiettivo del presente lavoro è discutere l'utilizzo di recenti sviluppi asintotici di ordine elevato per le aree nelle code (Reid, 2003, Ventura et al., 2011) per ottenere una semplice e accurata approssimazione per la misura di evidenza di Pereira-Stern. Il principale

vantaggio di tali procedure asintotiche è che non è richiesta alcuna assunzione sulla *prior*. Inoltre, sono procedure semplici da applicare, basate su quantità di verosimiglianza, e forniscono risultati accurati anche per numerosità campionarie piccole.

La tesi è articolata in tre capitoli.

Il Capitolo 1 fornisce una panoramica dell'analisi della sopravvivenza e passa in rassegna i modelli parametrici di regressione, i metodi semi-parametrici, di cui fa parte il modello di Cox, e i metodi non parametrici.

Il Capitolo 2 si concentra sull'analisi bayesiana della sopravvivenza, limitando l'attenzione ai metodi parametrici, e introduce le limitazioni legate all'uso di tale approccio.

Il Capitolo 3 mostra i limiti dovuti all'utilizzo del fattore di Bayes nell'ambito della verifica d'ipotesi e presenta una procedura alternativa nota come *misura di evidenza di Pereira-Stern*. Inoltre, il capitolo propone e discute una procedura di approssimazione di ordine elevato delle probabilità sulle code della distribuzione a posteriori.

Il capitolo si conclude con uno studio di simulazione per un modello di Weibull allo scopo di supportare la validità delle procedure di approssimazione presentate precedentemente. Lo studio viene effettuato per dati di sopravvivenza mediante la libreria HOA del software statistico R.

# CAPITOLO 1

## METODI DI ANALISI DELLA SOPRAVVIVENZA

### 1.1 Introduzione all'analisi della sopravvivenza

L'importanza della statistica nella ricerca biomedica sta nel fatto che essa consente di trarre delle conclusioni o fare inferenze facendo riferimento ad un campione piuttosto che a tutte le unità costituenti una popolazione. Una volta estratto un campione da una popolazione, è possibile testare su di esso una o più ipotesi attraverso l'uso congiunto della statistica descrittiva e della statistica inferenziale. La statistica descrittiva consente di descrivere il campione. Per esempio, un campione di persone può essere descritto sulla base delle sue caratteristiche, come l'età, il sesso, il fumo e la statura. Questi fattori possono essere a loro volta riassunti, generalmente, attraverso la media o la mediana oppure tramite misure di frequenza (proporzione di maschi, percentuale di fumatori, etc). L'analisi della sopravvivenza è parte della statistica inferenziale e la sua peculiarità consiste nel mettere in rapporto un certo evento con il fattore tempo (Tripepi e Catalano, 2004). Ad esempio, se si vuole studiare la sopravvivenza a due anni di un campione di persone (coorte dello studio), si deve seguire ciascuno per 24 mesi. Alla fine di questo periodo di osservazione ciascun paziente viene descritto da una coppia di valori, indicata generalmente con  $(\delta, t)$ , dove  $\delta$  è la condizione (0=non morto, 1=morto) e  $t$  è la durata dell'osservazione. E' necessario che, per tutti i soggetti reclutati, l'evento che individua il momento iniziale sia uguale e identificato nello stesso modo, utilizzando, per esempio, gli stessi criteri diagnostici (Di Orio, 1994, Cap.5). La caratteristica principale delle coorti è la loro dinamicità: infatti, seguire una coorte di persone per un certo tempo implica la possibilità che la sua numerosità vari nel tempo in ragione del verificarsi o meno dell'evento di interesse (per esempio l'evento morte).

L'analisi della sopravvivenza è un insieme di metodologie usate per descrivere e studiare il tempo necessario affinché uno specifico evento si verifichi. Nelle analisi della

sopravvivenza si usa il termine “guasto” per definire il verificarsi dell’evento d’interesse (anche nel caso in cui quest’ultimo costituisca un “successo”).

Un assunto fondamentale di questo tipo di analisi è che le unità statistiche sotto osservazione entrano ed escono dallo studio in modo del tutto casuale.

In uno studio della sopravvivenza sono da considerarsi cruciali, ai fini di ottenere la migliore validità possibile dei risultati, i seguenti elementi (si veda, ad esempio, Marubini e Valsecchi, 1995, Cap. 4):

- 1) *momento iniziale (starting point)*: è definito dal verificarsi di un evento ben individuabile nel tempo. Usualmente, potenziali momenti iniziali, negli studi clinici, sono eventi medici come diagnosi, inizio del trattamento, prima ricaduta, inizio del trattamento per la ricaduta, inizio della chemioterapia. In altri casi, la definizione del momento iniziale è molto complessa, come, ad esempio, nello studio della storia naturale delle malattie. In questi casi, l’introduzione nello studio di pazienti in momenti diversi del decorso della malattia espone il clinico a rischi di errore di valutazione della sopravvivenza e dei fattori prognostici.
- 2) *Criteri di diagnosi e di ammissione*: devono essere chiaramente esplicitati. Nell’adottare i criteri, il clinico deve, da una parte far riferimento a procedure dotate di una certa affidabilità quanto più elevata possibile nelle condizioni dello studio, dall’altra deve essere attento a non indurre selezioni rilevanti a causa, ad esempio, di una bassa accettabilità per il paziente.
- 3) *Modalità di reclutamento dei malati*: tutti i pazienti esaminati dagli osservatori devono essere inclusi nello studio, precisando se all’interno di una struttura sono coinvolti uno o più ricercatori. Il numero e le caratteristiche degli esclusi devono essere attentamente considerati e precisati. E’ fondamentale considerare ed esplicitare le vie di accesso dei malati all’osservazione e le possibili distorsioni indotte da meccanismi di filtro.
- 4) *Controllo* di tutti i pazienti per un periodo minimo prestabilito tenendo presente che una perdita di informazioni superiore al 20% diminuisce decisamente il valore clinico dello studio.
- 5) *Evento terminale (end point)*: evento prognosticamente rilevante che deve essere preliminarmente definito in termini obiettivi, affidabili e quanto più possibile legati alla pratica diagnostica corrente, oppure decesso del paziente.



Il tempo che intercorre tra il momento iniziale e l'evento terminale viene rappresentato da una variabile casuale  $T$  ( $T \geq 0$ ) definita "tempo di sopravvivenza", anche se l'evento che si osserva non è un decesso.

Nell'analisi statistica della sopravvivenza emergono due difficoltà: la prima è connessa con il fatto che la gran parte delle funzioni di densità di probabilità della variabile casuale  $T$  è fortemente asimmetrica positiva. Tale asimmetria suggerisce l'uso di trasformazioni o il ricorso a tecniche non parametriche per ridurre l'influenza di eventuali tempi di sopravvivenza particolarmente lunghi. La seconda difficoltà è dovuta alla presenza di dati censurati.

## 1.2 Alcune definizioni

Il termine "tempo di sopravvivenza" va usato in senso estensivo perché si applica anche a eventi diversi dalla morte. L'analisi della sopravvivenza riguarda, infatti, tutti quegli studi in cui si vuole analizzare l'incidenza di un determinato evento in un certo arco temporale (studi di coorte), ovvero il numero di pazienti per i quali si è verificato l'evento (Tripepi e Catalano, 2004). Perciò, il tempo di sopravvivenza assume significati diversi in relazione al tipo di evento a cui si è interessati. Il tempo di sopravvivenza  $T$  può essere il tempo che intercorre tra l'inizio dello studio e la morte, l'incidenza di un evento cardiovascolare (come l'infarto o l'ictus), il tempo di insorgenza di una patologia o il tempo in cui si verifica un episodio di rigetto renale (in una coorte di pazienti portatori di trapianto di rene). Ovviamente, il tempo di sopravvivenza, per i pazienti senza l'evento di interesse, corrisponde a quello compreso tra lo *starting point* e la fine dell'osservazione.

Si conosce il tempo esatto di sopravvivenza solo per quegli individui che presentano l'evento d'interesse durante il periodo di osservazione. Per gli altri individui, ovvero per coloro cui non si è manifestato l'evento alla fine del periodo osservato e per coloro di cui non si hanno più informazioni da un certo momento in poi nel periodo in esame, tutto quello che si può dire è che non mostrano l'evento d'interesse nel periodo studiato. Questi individui costituiscono le cosiddette "osservazioni censurate".

Una caratteristica interessante delle analisi della sopravvivenza è data dal fatto che sono in grado di tener conto delle informazioni derivanti dalle osservazioni censurate, la cui presenza non può essere trascurata nella fase di analisi dei dati.

Si possono individuare i seguenti tipi di censura:

- *censura di primo tipo*: al tempo 0 vengono messi sotto studio  $n$  individui e si termina l'osservazione dopo un tempo prefissato  $t^* > 0$ . Per gli  $m$  individui per i quali si è verificato l'evento d'interesse, è possibile osservare il tempo all'evento, per gli  $(n-m)$  individui rimanenti si saprà soltanto che essa è superiore a  $t^*$ .
- *Censura di secondo tipo*: al tempo 0 vengono messi sotto osservazione  $n$  individui e si termina l'osservazione quando per un numero prefissato di  $m$  unità si è osservato l'evento d'interesse, dove  $m$  è un numero prefissato all'inizio dello studio.

Inoltre la censura può essere dovuta a:

- *Censura variabile*: costituisce una generalizzazione della censura del primo tipo, in cui il tempo di osservazione prefissato  $t^*$  varia da unità a unità.
- *Censura casuale*: i tempi di osservazione, che variano da soggetto a soggetto in modo non controllabile dall'osservatore, sono realizzazioni di variabili casuali.

Un'altra peculiarità degli studi della sopravvivenza, che si osserva più frequentemente negli studi osservazionali, è il troncamento. Esso si distingue in:

- *troncamento a sinistra*: un soggetto è troncato a sinistra se entra a far parte della popolazione a rischio in un qualche momento dopo l'inizio del periodo osservato.
- *Troncamento a destra*: un soggetto è troncato a destra se esce dalla popolazione a rischio in un qualche momento dopo l'inizio dello studio.

Il collegamento tra l'analisi della sopravvivenza e la "probabilità condizionata" è intuitivo: infatti, la probabilità che ha un paziente di sopravvivere dopo  $x$  giorni dall'ingresso in uno studio (probabilità indicata generalmente con  $P$ ) è condizionata dal fatto che il paziente sia sopravvissuto negli  $x-1$  giorni precedenti. Questa probabilità viene anche definita probabilità cumulativa o sopravvivenza cumulativa. Siano  $p_1$  la probabilità che ha il paziente di sopravvivere il primo giorno,  $p_2$  la probabilità di sopravvivere il secondo giorno e  $p_3$  la probabilità di sopravvivere il terzo giorno. La sopravvivenza cumulativa è data dal prodotto di queste singole probabilità:  $P=p_1 \cdot p_2 \cdot p_3$ . Per cui se la probabilità di sopravvivere il primo giorno è 0.80, quella di sopravvivere il secondo giorno è 0.68 e quella di sopravvivere il terzo giorno è 0.55, la probabilità cumulativa è  $0.80 \cdot 0.68 \cdot 0.55$ , ovvero 0.30 (cioè il 30%).

I tempi all'evento sono realizzazioni della variabile casuale  $T$ .

La funzione di sopravvivenza è il complementare a 1 della funzione di ripartizione  $F(t) = Pr\{T \leq t\} = \int_0^t f(u) du$ , dove  $f(t)$  è la funzione di densità di  $T$ , il cui differenziale,  $dF(t) = p(t)dt \sim F(t + dt) - F(t)$ , fornisce la probabilità condizionata di subire l'evento

nell'intervallo infinitesimo  $(t, t+dt)$ . Pertanto,  $S(t) = 1 - F(t) = \int_t^{+\infty} f(u)du = Pr\{T > t\}$  definisce la probabilità di sopravvivere oltre il tempo  $t$ , e fornisce informazioni sul rischio nel tempo.

La funzione azzardo  $\lambda(t)$  fornisce informazioni sul rischio corrente. E' definita da

$$\lambda(t) = \lim_{\Delta t \rightarrow 0^+} \frac{Pr\{t < T \leq t + dt \mid T \geq t\}}{\Delta t}. \quad (1.1)$$

Pertanto,  $\lambda(t) \cdot d(t)$  fornisce la probabilità di evento nell'intervallo  $(t, t+dt)$ , dato che il soggetto è vivo al tempo  $t$ . La funzione di azzardo  $\lambda(t)$ , definita in termini di  $f(t)$  e  $S(t)$ , è data da  $\lambda(t) = f(t)/S(t) = -\frac{d}{dt} \log S(t)$ . Integrando entrambi membri, si ottiene  $-\int_0^t \lambda(u) du = \log S(u)|_0^t$  e, dato che  $S(0)=1$ , si può definire la funzione cumulata dell'azzardo come

$$\Lambda(t) = -\log S(t) = \int_0^t \lambda(t) dt. \quad (1.2)$$

Elevando a potenza, si ottiene

$$S(t) = \exp\left[-\int_0^t \lambda(u) du\right] = \exp[-\Lambda(t)]. \quad (1.3)$$

La funzione di azzardo relativo (*HR - Hazard Ratio*) è data da  $h(t) = \frac{\lambda(t)}{\lambda_0(t)}$ , dove  $\lambda_0(t)$  è la funzione rischio arbitraria che si assume essere uguale per tutti i soggetti. Se  $h(t)$  non varia al variare del tempo, si ha  $h(t) = h$ .

I metodi dell'analisi della sopravvivenza sono classificati in metodi non parametrici, metodi semi-parametrici e metodi parametrici, sulla base degli assunti che vengono fatti sulla distribuzione di  $T$ .

### 1.3 Metodi non parametrici

I metodi non parametrici vengono utilizzati quando non si assume una distribuzione teorica per  $F(t)$ . Ci sono essenzialmente tre metodi non parametrici: la tavola di sopravvivenza, il metodo di Kaplan Meier e il Log-rank test.

### 1.3.1 Il metodo delle tavole di sopravvivenza

Il metodo di calcolo delle tavole di sopravvivenza, conosciuto anche come *metodo attuariale* o *metodo di Cutler Ederer* (Cutler ed Ederer, 1958), consiste nella suddivisione dell'asse temporale in intervalli di tempo, ad esempio in numero  $J + 1$ , di uguale ampiezza, tranne l'ultimo che ha ampiezza infinita. Su un campione casuale di  $n$  soggetti, vengono rilevati i tempi di sopravvivenza e si contano quanti individui presentano l'evento d'interesse, o l'osservazione troncata, in ciascun intervallo. Gli elementi necessari alla costruzione della tavola di sopravvivenza sono quindi (per  $j=1, \dots, J+1$ ):

- l'ampiezza,  $h_j$ , dell'intervallo  $j$ -esimo;
- il numero di individui,  $n_j$ , che entrano nell'intervallo;
- il numero di eventi,  $d_j$ , verificatisi nell'intervallo;
- il numero di individui,  $w_j$ , il cui tempo di sopravvivenza è troncato nell'intervallo perché sono usciti vivi dallo studio;
- il numero di individui,  $l_j$ , il cui tempo di sopravvivenza è troncato nell'intervallo perché sono stati persi di vista durante lo studio.

E' possibile definire una stima della probabilità condizionata  $\hat{q}_j$  che un soggetto muoia nell'intervallo  $j$ -esimo, dato che è vivo nell'intervallo immediatamente precedente, come il rapporto tra  $d_j$  ed  $n_j$ . La stima della probabilità di sopravvivenza all'intervallo  $j$ -esimo, data la sopravvivenza all'inizio dello stesso, è data da  $\hat{p}_j$ , complementare di  $\hat{q}_j$ . Il prodotto delle stime delle probabilità condizionate di sopravvivere in ciascuno degli intervalli fino al  $j$ -esimo, indicato con  $P_j$ , definisce la probabilità cumulativa di sopravvivenza.

Un' ulteriore componente fondamentale delle tavole di sopravvivenza è la stima puntuale del tasso di morte, definita nel punto centrale dell'intervallo  $j$ -esimo come

$$\hat{\lambda}_j = \frac{d_j}{h_j(n_j - d_j/2)} = \frac{2\hat{q}_j}{h_j(1 + \hat{p}_j)}, \quad j = 1, \dots, J+1. \quad (1.4)$$

La (1.4) è detta *stima attuariale del rischio di morte* (Kimball, 1960).

La funzione di azzardo istantaneo è definita dalla proporzione della popolazione presente al tempo  $t$  che presenta l'evento d'interesse, per unità di tempo. La funzione di azzardo cumulato al tempo  $t$ ,  $A(t)$ , si ottiene dalla somma degli azzardi fino al tempo  $t$ . La

relazione tra la funzione di azzardo e la funzione di sopravvivenza è definita dall'equazione (1.2).

Lo stimatore Nelson-Aalen della funzione cumulata dell' azzardo al tempo  $t$ , è definito come

$$\widehat{\Lambda}(t) = \sum_{j:t_j \leq t} \frac{d_j}{n_j}, \quad (1.5)$$

dove  $d_j$  è il numero di soggetti per i quali si verifica l'evento terminale al tempo  $j$ -esimo e  $n_j$  è il numero di soggetti a rischio al tempo  $j$ -esimo, ovvero il numero di individui per i quali non si è ancora verificato l'evento terminale all'inizio del periodo.

### 1.3.2 Lo stimatore di Kaplan-Meier

Il metodo non parametrico più diffuso per la stima della probabilità di sopravvivenza è il *metodo del prodotto limite*, più noto come *stimatore di Kaplan - Meier* (Kaplan e Meier, 1958). Questo criterio è spesso definito per piccoli campioni in quanto sfrutta meglio l'informazione disponibile. Consiste nello stimare la probabilità condizionata di sopravvivenza in corrispondenza di ciascuno dei tempi in cui si verifica almeno un evento terminale. A differenza della procedura adottata nel metodo attuariale, il metodo di Kaplan-Meier non implica la suddivisione dell'asse temporale in intervalli di ampiezza prefissata e quindi neppure il conseguente raggruppamento di soggetti.

Si supponga di avere  $n$  soggetti e di osservare i tempi all'evento, in ordine crescente, per gli  $n$  individui:  $t_{(1)} < t_{(2)} < \dots < t_{(J)}$ , con  $J \leq n$ . Sia  $d_j$  il numero di soggetti che presentano l'evento d'interesse al tempo  $t_{(j)}$ , per  $j = 1, \dots, J$ . Nel caso in cui  $d_j$  sia maggiore di uno, significa che più soggetti presentano tempi di evento uguali. Sia, inoltre,  $n_j$  il numero di individui a rischio al tempo  $t_{(j)}$ . E' possibile, a questo punto, definire la probabilità condizionata di morire nell'istante  $t_{(j)}$ , dato che il soggetto in questione è vissuto fino all'istante immediatamente precedente. La sua stima è definita come

$$\hat{q}_j = \frac{d_j}{n_j}. \quad (1.6)$$

Pertanto, la probabilità di sopravvivere al tempo  $t_{(j)}$ , data dal complementare di  $\hat{q}_j$ , è stimata da

$$\hat{p}_j = 1 - \hat{q}_j = \frac{n_j - d_j}{n_j}. \quad (1.7)$$

Moltiplicando tra loro le stime delle probabilità condizionate di sopravvivere, si ottiene la stima della probabilità cumulata di vivere oltre l'istante  $t_{(j)}$ , data da

$$\widehat{S}(t) = \prod_{j|t_{(j)} \leq t} \hat{p}_j, \quad (1.8)$$

che definisce lo stimatore di Kaplan - Meier. Dal momento che gli istanti di tempo, in cui non si verifica l'evento terminale, assumono probabilità pari a uno, si ha che lo stimatore di Kaplan - Meier coinvolge solo i tempi all'evento e varia solo quando se ne verifica almeno uno.

Lo stimatore ottenuto nella (1.8) è non distorto e segue una distribuzione asintotica normale. La varianza asintotica è stimabile tramite la formula di Greenwood, come

$$\text{Var} [\widehat{S}(t)] = \widehat{S}(t)^2 \sum_{j|t_{(j)} \leq t} \frac{d_j}{n_j(n_j - d_j)}. \quad (1.9)$$

Essa è inversamente proporzionale al numero dei soggetti a rischio.

La stima di Kaplan - Meier della probabilità di sopravvivenza viene rappresentata con una curva a gradini, che parte dal valore 1 e decresce nel tempo. L'altezza dei gradini dipende dal numero di eventi e dal numero dei soggetti a rischio. Si veda la Figura 1.1 per un esempio.

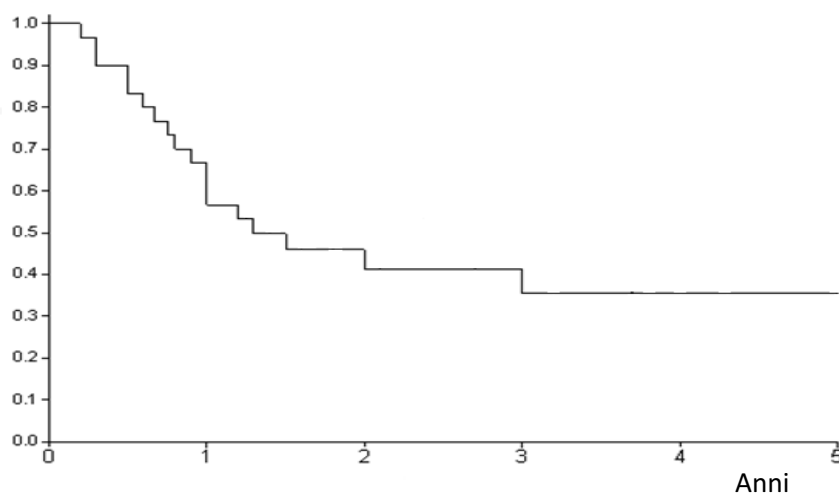


Figura 1.1: Esempio di curva di Kaplan-Meier.

### 1.3.3 Il Log-rank test

Il log-rank test, proposto da Mantel nel 1966 come estensione del test di Mantel e Haenszel (Mantel e Haenszel, 1959), è un test non parametrico, basato sui ranghi, che permette di confrontare curve di sopravvivenza di due (o più) insiemi di soggetti sottoposti a differenti trattamenti o esposti a differenti rischi.

Si supponga, per semplicità, di avere due insiemi di soggetti, di numerosità rispettivamente  $n_A$  ed  $n_B$ , scelti in maniera casuale, appartenenti, rispettivamente, ai due gruppi  $A$  e  $B$ , per un totale di  $n = n_A + n_B$  soggetti. Il log-rank test viene utilizzato per saggiare l'ipotesi nulla  $H_0: \lambda_A(t) = \lambda_B(t)$ , per ogni istante  $t$ . Si vuole, pertanto, verificare l'uguaglianza delle probabilità di sopravvivenza sui due differenti gruppi. Per convenzione, si considera il numero  $d_{Aj}$  di pazienti per i quali si è verificato l'evento terminale al tempo  $t_{(j)}$ , all'interno del gruppo  $A$ , per  $j=1, \dots, J$ . Sotto l'ipotesi nulla, la distribuzione di  $d_{Aj}$  risulta essere ipergeometrica, per cui essa ha media e varianza date, rispettivamente, da

$$E(d_{Aj}) = n_{Aj} \cdot \frac{d_j}{n_j} \quad (1.10)$$

e

$$\text{var}(d_{Aj}) = \left[ n_{Aj} \cdot \frac{d_j}{n_j} \left( 1 - \frac{d_j}{n_j} \right) \right] \left( \frac{n_j - n_{Aj}}{n_j - 1} \right). \quad (1.11)$$

Il test è quindi dato da

$$Q_{M-H} = \frac{\left\{ \sum_{j=1}^J [d_{Aj} - E(d_{Aj})] \right\}^2}{\sum_{j=1}^J \text{var}(d_{Aj})}. \quad (1.12)$$

Tale test, sotto l'ipotesi nulla, si distribuisce asintoticamente come un Chi-quadro con un grado di libertà. Si rifiuta l'ipotesi nulla di uguaglianza delle curve di sopravvivenza relative ai due insiemi di pazienti, per valori grandi di  $Q_{M-H}$ .

## 1.4 Modelli parametrici di regressione

Si supponga di disporre di un campione di tempi di sopravvivenza  $t_i$ ,  $i=1, \dots, n$ . E' spesso d'interesse valutare se, e in che modo, la distribuzione di sopravvivenza è influenzata da variabili concomitanti  $\mathbf{x}=(x_{i1}, \dots, x_{ik})$ , dove  $x_{ir}$  è il valore assunto dalla  $r$ -esima variabile concomitante per l' $i$ -esima unità statistica. La modellazione più semplice si ha specificando un modello parametrico per la distribuzione marginale dei tempi  $t_i$  (si veda, ad esempio, Pace e Salvan, 2001).

I modelli di regressione parametrici vengono usati quando si assume una distribuzione teorica per  $F(t)$ . Alcuni tra i modelli parametrici più usati nell'analisi della sopravvivenza sono riportati in Tabella 1, insieme alla relativa funzione rischio. I modelli parametrici di regressione sono in genere distinti a seconda del comportamento della funzione rischio.

Il modello esponenziale presuppone che il rischio istantaneo di morte non vari al trascorrere del tempo. Questo implica, per esempio in uno studio clinico, che il rischio di morte nel primo anno di osservazione del paziente in esame, è pari al rischio di morte in ciascuno degli anni successivi di osservazione. Un esempio dove questo non accade è dato dall'osservazione del periodo immediatamente successivo all'operazione chirurgica (in genere 30 giorni) per i pazienti che sono stati sottoposti ad amputazione della gamba, periodo in cui si nota un eccesso di mortalità dovuto al forte rischio di infezione. Il rischio di morte dopo questo periodo risulta essere più piccolo. In generale, per distinguere la funzione di azzardo costante nel tempo da quella che invece varia al variare del tempo, si usa la notazione  $\lambda$  in luogo di  $\lambda(t)$ .

Il modello di Weibull presuppone, invece, che la funzione rischio possa essere costante nel tempo, monotona crescente o monotona decrescente, a seconda del valore assunto dal parametro di forma  $\alpha$  della distribuzione.

Il modello log-normale è l'unico ad avere una funzione rischio non monotona, ma che cresce dal valore iniziale nullo ( $t=0$ ) fino a raggiungere un punto di massimo, per poi decrescere verso zero (per  $t \rightarrow \infty$ ).

Incorporando esplicitamente il vettore  $\mathbf{x}$  delle covariate nella formulazione parametrica della funzione rischio, è possibile studiare il loro effetto sulla sopravvivenza. Ad esempio, la funzione rischio studiata da Glasser (1967), ottenuta assumendo che la variabile casuale  $T$ , dato il vettore  $\mathbf{x}$ , sia distribuita in accordo ad un modello esponenziale, prevede



MODELLO	FORMULAZIONE DEL MODELLO	DESCRIZIONE DELLA FUNZIONE RISCHIO
Esponenziale	$f(t) = \lambda \exp(-\lambda t)$ $S(t) = \exp(-\lambda t)$ $\lambda(t) = \lambda$	costante
Weibull si riduce al modello esponenziale per $\alpha = 1$	$f(t) = \lambda \alpha (\lambda t)^{\alpha-1} \exp[-(\lambda t)^\alpha]$ $S(t) = \exp[-(\lambda t)^\alpha]$ $\lambda(t) = \lambda \alpha (\lambda t)^{\alpha-1}$	Per $\alpha < 1$ monotona decrescente Per $\alpha > 1$ monotona crescente Per $\alpha = 1$ costante e pari a $\lambda$
Gamma si riduce al modello esponenziale per $k = 1$	$f(t) = \frac{\lambda(\lambda t)^{k-1} e^{-\lambda t}}{\Gamma(k)}$ $S(t) = 1 - \left[ \frac{1}{\Gamma(k)} \cdot \int_0^t u^{k-1} e^{-u} du \right]$ $\lambda(t) = \frac{f(t)}{S(t)}$ <p>con <math>k &gt; 0, \lambda &gt; 0, \Gamma(k) = \int_0^{+\infty} u^{k-1} e^{-u} du</math></p>	Per $k < 1$ monotona decrescente  Per $k > 1$ monotona crescente  Per $k = 1$ costante e pari a $\lambda$
Log - normale	$f(t) = (2\pi)^{-1/2} (\sigma t)^{-1} \exp\left[-\frac{1}{2} \left(\frac{\ln t - \mu}{\sigma}\right)^2\right]$ $S(t) = 1 - \Phi(z)$ $\lambda(t) = \frac{f(t)}{S(t)}$ <p>con <math>\Phi(z) = \int_{-\infty}^z (2\pi)^{-1/2} e^{-u^2/2} du</math> e <math>z = \frac{\ln t - \mu}{\sigma}</math></p>	Cresce, raggiunge un massimo e poi decresce. Al variare di $\sigma$ la forma di $\lambda(t)$ varia, mantenendo in ogni caso il comportamento descritto.

TABELLA 1 : Modelli parametrici di uso frequente nell'analisi della sopravvivenza.

$$\lambda(t;\mathbf{x})=\lambda \exp(\mathbf{x}^T \boldsymbol{\beta}), \quad (1.13)$$

dove  $\boldsymbol{\beta}$  è un vettore di  $p \times 1$  ignoti parametri di regressione,  $\mathbf{x}$  è un vettore di covariate e  $\lambda$  è il tasso di guasto ( $\lambda > 0$ ).

I coefficienti di regressione  $\boldsymbol{\beta}$  costituiscono una misura quantitativa dell'effetto esercitato da ciascuno dei  $p$  fattori prognostici sul rischio di morte. La stima di  $\boldsymbol{\beta}$ , così come i test sui coefficienti di regressione, ricorrono alla teoria della verosimiglianza.

Il modello di Weibull assume invece

$$\lambda(t;\mathbf{x})=\lambda \alpha (\lambda t)^{\alpha-1} \exp(\mathbf{x}^T \boldsymbol{\beta}). \quad (1.14)$$

Ne segue che il rapporto di azzardi (*hazard ratio*) per due individui con vettori di covariate, rispettivamente,  $\mathbf{x}_1$  e  $\mathbf{x}_2$  non dipende dal tempo poiché si ha

$$\text{HR} = \frac{\lambda(t;\mathbf{x}_1)}{\lambda(t;\mathbf{x}_2)} = \frac{\exp(\mathbf{x}_1^T \boldsymbol{\beta})}{\exp(\mathbf{x}_2^T \boldsymbol{\beta})}. \quad (1.15)$$

Pertanto, l'effetto di ciascuna covariata sull'HR è moltiplicativo.

Nel Capitolo 2 saranno presentati alcuni esempi di funzioni di verosimiglianza per modelli di regressione parametrici, tra i quali il modello esponenziale e il modello di Weibull.

Per mezzo delle relazioni (1.14) e (1.15) è possibile definire una più ampia classe di modelli per i quali la funzione azzardo è scomponibile in due fattori, di cui uno dipendente solo dal tempo e l'altro solo dalle covariate, ossia

$$\lambda(t; \mathbf{x})=\lambda_0(t) \cdot h(\mathbf{x}). \quad (1.16)$$

L'equazione (1.16) caratterizza la classe di modelli che presuppongono la proporzionalità delle funzioni rischio (*proportional hazard models* o *PH models*).

Si ottengono diversi modelli in base alle differenti forme parametriche assunte da  $\lambda_0(t)$ . La scelta di  $h(\mathbf{x})$  dipende dal tipo di dati sotto studio e non deve essere necessariamente di forma esponenziale.

Una terza classe di modelli utili per l'analisi della sopravvivenza è costituita dai modelli con tempi all'evento accelerati (*accelerated failure time models* o *AFT models*). In questa classe di modelli si assume che il vettore  $\mathbf{x}$  agisca moltiplicativamente sul tempo di sopravvivenza: il ruolo di  $\mathbf{x}$  consiste nell'aumentare o diminuire la velocità con cui un individuo procede lungo l'asse dei tempi, accelerando o rallentando il verificarsi dell'evento terminale (si veda, ad esempio, Marubini e Valsecchi, 1995).

Le classi di modelli parametrici sopra brevemente descritte, forniscono una varietà di metodi per l'analisi della sopravvivenza che non è esaurita da quelli presentati in Tabella 1. E' necessario tenere presente che la loro applicabilità è limitata dal fatto che la funzione rischio di ciascuno di essi assume una ben definita forma matematica, mentre in molte situazioni pratiche non vi sono sufficienti giustificazioni per l'adozione di un particolare modello.

Nel 1972 Cox (Cox, 1972) ha presentato un modello che non presuppone nessuna forma parametrica per la funzione  $\lambda_0(t)$ . Il modello è espresso come

$$\lambda(t; \mathbf{x}) = \lambda_0(t) \exp(\mathbf{x}^T \boldsymbol{\beta}), \quad (1.17)$$

dove  $\lambda_0(t)$  rappresenta una funzione rischio arbitraria che si assume essere uguale per tutti i soggetti.

## 1.5 Il modello di regressione di Cox

Il modello di Cox si colloca tra le procedure semi-parametriche per il fatto che non assume alcuna forma specifica circa la distribuzione della variabile casuale  $T$ , ma modella l'effetto delle variabili prognostiche sulla sopravvivenza in modo parametrico (si veda, ad esempio, Parmar e Machin, 1995, Cap. 6).

Dall'equazione (1.17) è immediato vedere che

$$\frac{\lambda(t; \mathbf{x})}{\lambda(t; \mathbf{0})} = \exp(\mathbf{x}^T \boldsymbol{\beta}), \quad (1.18)$$

da cui si ottiene

$$\lambda(t; \mathbf{0}) = \lambda_0(t), \quad (1.19)$$

dove  $\lambda_0(t)$  è la funzione di azzardo del soggetto con tutte le covariate a valore nullo, definita pertanto *azzardo di base* (*baseline*). La stima dei coefficienti  $\beta$  permette di valutare il rischio relativo di presentare l'evento terminale per un soggetto con covariate  $\mathbf{x}$  rispetto ad un soggetto con tutte le covariate a valore nullo, assunto che tale rischio si mantenga costante nel tempo. Non essendo  $\lambda_0(t)$  definita, risulta impossibile stimare i coefficienti di regressione  $\beta$  con una funzione di verosimiglianza ordinaria. Nel suo lavoro originario, Cox (Cox, 1972), stima i coefficienti di regressione  $\beta$  considerando, nella funzione di verosimiglianza,  $\lambda_0(t)$  come funzione di disturbo (*nuisance function*). La funzione di verosimiglianza così ottenuta è definita funzione di verosimiglianza parziale. Dal momento che il tasso di rischio per due soggetti con vettori di covariate fissi  $\mathbf{x}_i$  e  $\mathbf{x}_j$ , definito come

$$\frac{\lambda(t; \mathbf{x}_i)}{\lambda(t; \mathbf{x}_j)} = \frac{\lambda_0(t) \exp(\mathbf{x}_i^T \beta)}{\lambda_0(t) \exp(\mathbf{x}_j^T \beta)} = \frac{\exp(\mathbf{x}_i^T \beta)}{\exp(\mathbf{x}_j^T \beta)}, \quad (1.20)$$

è costante nel tempo, questo modello è anche chiamato “*modello a rischi proporzionali*” (*proportional hazards model*).

Dato  $R(t)$  il numero di soggetti a rischio al tempo  $t$ , la probabilità che un soggetto con covariate  $\mathbf{x}$  subisca l'evento d'interesse nell'intervallo  $(t, t+dt)$  è pari a  $\lambda(t; \mathbf{x})dt$ . Condizionatamente al fatto che il soggetto abbia subito l'evento al tempo  $t_{(j)}$ , la probabilità che si tratti di un individuo con covariate  $\mathbf{x}_j$  è data da

$$\frac{\lambda(t_{(j)}; \mathbf{x}_j) \cdot dt}{\sum_{k \in R_{t_{(j)}}} [\lambda(t_{(j)}; \mathbf{x}_k) \cdot dt]} = \frac{\exp(\mathbf{x}_j^T \beta)}{\sum_{k \in R_{t_{(j)}}} \exp(\mathbf{x}_k^T \beta)}. \quad (1.21)$$

Non essendo  $\lambda_0(t)$  specificata, si nota che nessuna informazione utile per la stima di  $\beta$  sia ottenibile dagli intervalli di tempo  $(t_{(j)}, t_{(j+1)})$ , in cui non si verifica alcun evento.

Pertanto, la verosimiglianza su cui si basa la stima di  $\beta$ , alla cui formulazione contribuiscono solo le probabilità condizionali (1.21) calcolate in corrispondenza di ogni tempo di risposta, è definita come

$$L_{PA}(\beta) = \prod_{j=1}^J \frac{\exp(\mathbf{x}_j^T \beta)}{\sum_{k \in R_{t_{(j)}}} \exp(\mathbf{x}_k^T \beta)}. \quad (1.22)$$

Questa funzione prende il nome di *funzione di verosimiglianza parziale*. Pur non essendo una funzione di verosimiglianza in senso proprio, la verosimiglianza parziale ha, approssimativamente, le stesse proprietà della verosimiglianza propria.

La stima del vettore  $\beta$  implica la massimizzazione della funzione di verosimiglianza parziale, che equivale a massimizzare il suo logaritmo

$$l_{PA}(\beta) = \log L_{PA}(\beta) = \sum_{j=1}^J \{(\mathbf{x}_j^T \beta) - \log[\sum_{k \in R_{t(j)}} \exp(\mathbf{x}_k^T \beta)]\}. \quad (1.23)$$

Per ottenere la SMV di  $\beta$  è necessario usare un processo iterativo, ad esempio il metodo di Newton-Raphson. Le stime di  $\beta$  ottenute permettono di studiare il ruolo, significativo o meno, di ciascuna covariata. Tuttavia, dal punto di vista clinico è fondamentale studiare anche la curva di sopravvivenza  $S(t; \mathbf{x})$  di un paziente, con vettore di covariate  $\mathbf{x}$ , dati particolari valori assunti dalle variabili prognostiche rilevate sullo stesso.

La funzione cumulativa di rischio di tale soggetto, essendo  $\Lambda_0(t)$  la funzione rischio cumulativa del soggetto con tutte le covariate nulle, è data da

$$\Lambda(t; \mathbf{x}) = \int_0^t \lambda_0(u) \exp(\mathbf{x}^T \beta) du = \exp(\mathbf{x}^T \beta) \cdot \Lambda_0(t). \quad (1.24)$$

Si ha, pertanto,

$$S(t; \mathbf{x}) = \exp[-\Lambda(t; \mathbf{x})] = S_0(t) \exp(-\mathbf{x}^T \beta). \quad (1.25)$$

E' fondamentale disporre di uno stimatore di  $S_0(t)$ . Una possibilità è lo stimatore proposto da Breslow (1980), definito come

$$\widehat{S}_0(t) = \prod_{t(j) \leq t} \left( 1 - \frac{d_j}{\sum_{k \in R_{t(j)}} \exp(\mathbf{x}_k^T \beta)} \right), \quad (1.26)$$

con  $\widehat{S}_0(0) = 1$ .

## 1.6 Test disponibili per la valutazione dei modelli

Ci sono tre test in grado di valutare e comparare differenti modelli: il test log-rapporto di verosimiglianza, il test alla Wald e il test score (si veda, ad esempio, Marubini e Valsecchi, 1995). Tali test sono applicabili sia nell'ambito dei modelli di regressione parametrici, facendo ricorso alla funzione di verosimiglianza, sia nell'ambito del modello di Cox, utilizzando la funzione di verosimiglianza parziale. Infatti quest'ultima, pur non essendo una funzione di verosimiglianza in senso proprio, gode tuttavia di tutte le proprietà di una funzione di verosimiglianza genuina (si veda, ad esempio, Parmar e Machin, 1995).

Dato il vettore  $\beta$  di  $p$  parametri, in genere si hanno due scenari: o il ricercatore è interessato a vedere se le  $p$  covariate sono determinanti ai fini dello studio, oppure è interessato a valutare il ruolo di un insieme  $g < p$  di variabili indipendenti. Nel primo caso si va a verificare l'ipotesi nulla  $H_0: \beta=0$  contro l'ipotesi alternativa  $H_1: \beta \neq 0$  (intercetta esclusa). Nel secondo caso, invece, si vuole verificare l'ipotesi nulla  $H_0: \beta = (\beta_1, \dots, \beta_{p-g}, 0, \dots, 0)$ . Si valuta così l'effetto di  $g$  variabili sulla sopravvivenza. Il *test log-rapporto di verosimiglianza* calcola la verosimiglianza sotto due differenti ipotesi e le compara. Sia  $\hat{\beta}$  la stima di massima verosimiglianza del vettore  $\beta$ , a  $p$  componenti, ottenuta considerando il modello di regressione completo. Sia, invece  $\tilde{\beta}$  la stima nel modello ridotto. Il test log-rapporto di verosimiglianza è dato da

$$W = 2[l(\hat{\beta}) - l(\tilde{\beta})], \quad (1.27)$$

dove  $l(\hat{\beta})$  è la funzione di log-verosimiglianza massimizzata sotto il modello completo e  $l(\tilde{\beta})$  è la funzione di log-verosimiglianza massimizzata sotto il modello ridotto. Sotto  $H_0$  tale statistica test ha distribuzione asintotica  $\chi^2$  con  $g$  gradi di libertà.

L'inferenza sul vettore  $\beta$  può essere eseguita anche tramite il test di Wald o il test di Rao (o test score). Nel caso in cui  $H_0$  imponga una sola restrizione  $g=1$ , ovvero nel caso in cui si voglia verificare l'ipotesi nulla  $H_0: \beta_j=0$  contro l'ipotesi alternativa  $H_1: \beta_j \neq 0$ , allora la statistica di Wald è definita come

$$z = \frac{\hat{\beta}_j}{se(\hat{\beta}_j)}, \quad (1.28)$$

dove  $Se(\beta_j)$  è lo *standard error* relativo al coefficiente di regressione  $\beta_j$ .

Essa ha distribuzione normale standard sotto l'ipotesi nulla.

Infine, il test Rao consente di verificare l'ipotesi nulla  $H_0: \boldsymbol{\beta} = \boldsymbol{\beta}_0 = (\beta_1, \dots, \beta_{p-g}, 0, \dots, 0)$ .

Sia  $U(\boldsymbol{\beta})$  il vettore con componente generica  $U(\beta_i) = \frac{\partial l(\boldsymbol{\beta})}{\partial \beta_i}$  e  $U(\boldsymbol{\beta}_0)$  il valore in corrispondenza di  $\boldsymbol{\beta}_0$ . La statistica di Rao è pertanto data da

$$Q = (U(\boldsymbol{\beta}_0))^T I^{-1}(\boldsymbol{\beta}_0) (U(\boldsymbol{\beta}_0)), \quad (1.29)$$

dove  $I(\boldsymbol{\beta}_0)$  è la matrice d'informazione attesa calcolata in  $\boldsymbol{\beta}_0$ . La statistica  $Q$  ha distribuzione  $\chi^2$  con  $g$  gradi di libertà.

## 1.7 Conclusioni

Il capitolo successivo introduce l'approccio bayesiano come alternativa all'approccio frequentista all'analisi della sopravvivenza. L'attenzione è focalizzata sui modelli parametrici di cui al Paragrafo 1.4.





## CAPITOLO 2

### ANALISI BAYESIANA DELLA SOPRAVVIVENZA

#### 2.1 Introduzione

Nell'approccio frequentista, si esclude l'eventualità che ci siano informazioni pregresse sul fenomeno in esame. Secondo questo approccio, si assume che ciò che si osserva è un campione estratto da una variabile  $Y$  avente funzione di densità  $f(y/\theta)$ , con  $\theta \in \Theta$ , dove  $\Theta$  indica lo spazio parametrico. Si considera, pertanto, che i dati siano realizzazioni di una variabile aleatoria e che il parametro  $\theta$  sia un valore incognito su cui si vuole fare inferenza. L'elemento soggettivo nell'analisi di  $y$  è limitato alla scelta del modello statistico che meglio rappresenta i dati,  $\mathcal{F}$  (Pace e Salvan, 2001).

Nell'approccio bayesiano, invece, i dati osservati sono considerati costanti e il parametro ignoto  $\theta$  è una variabile casuale con distribuzione definita sullo spazio parametrico  $\Theta$ . Questo approccio si avvale sia della verosimiglianza che di una distribuzione a priori  $\pi(\theta)$ , che descrive le aspettative del ricercatore circa i possibili valori che il parametro può assumere. Essendo l'informazione a priori antecedente all'osservazione del campione, essa non dipende da questo. Pertanto, tutte le informazioni sul parametro, che sono a disposizione del ricercatore prima di effettuare le osservazioni campionarie, sono rappresentate dalla distribuzione a priori  $\pi(\theta)$ .

Nel corso degli anni sono via via aumentate le applicazioni statistiche in cui l'esigenza di tener conto di informazioni extra-sperimentali emergeva con chiarezza. In ambito epidemiologico, ad esempio, quando si valuta la probabilità che un paziente sia affetto da una certa patologia sulla base di un test diagnostico, le informazioni a priori sono nient'altro che le informazioni sulla prevalenza di quella malattia in quel contesto specifico (Liseo, 2008).

Tramite il Teorema di Bayes, le informazioni a priori sul parametro vengono aggiornate sulla base delle informazioni provenienti dal campione estratto, contenute nella funzione

di verosimiglianza  $L(\theta; y)$ . Tale processo di aggiornamento conduce alla formazione della *distribuzione a posteriori*, data da

$$\pi(\theta/y) = \frac{\pi(\theta)L(\theta;y)}{\int_{\Theta} L(\theta;y)\pi(\theta)d\theta}, \quad (2.1)$$

dove il denominatore indica la costante di normalizzazione. Equivalentemente, si ha,

$$\pi(\theta/y) \propto \pi(\theta)L(\theta; y). \quad (2.2)$$

La (2.2) riassume tutte le informazioni che si hanno sul parametro  $\theta$ :  $L(\theta; y)$  rappresenta il contributo proveniente dai dati osservati, mentre  $\pi(\theta)$  quantifica il contributo proveniente dall'informazione pre-sperimentale. Il risultato sperimentale, contenuto in  $L(\theta; y)$ , porta a un arricchimento dell'opinione e dell'informazione su  $\theta$ .

Il procedimento di determinazione della distribuzione a posteriori può, in sintesi, essere riassunto nei seguenti passi:

- individuazione dello spazio parametrico  $\Theta$ ;
- scelta di una distribuzione a priori che rifletta le opinioni pre-sperimentali,  $\pi(\theta)$ , con supporto  $\Theta$ ;
- specificazione di una funzione di verosimiglianza che apporti l'informazione proveniente dai dati osservati;
- combinazione della distribuzione a priori con la funzione di verosimiglianza attraverso il Teorema di Bayes;
- determinazione della distribuzione a posteriori, con supporto lo spazio parametrico  $\Theta$ , da cui è possibile ricavare stime puntuali, intervallari ed effettuare verifiche d'ipotesi.

A parità di verosimiglianza, le distribuzioni finali, o distribuzioni a posteriori, possono più o meno differire a seconda delle distribuzioni iniziali.

## 2.2 Scelta della distribuzione a priori

La distribuzione a priori rappresenta da un lato un'opportunità, in quanto permette di includere formalmente informazioni che non provengono dalle osservazioni campionarie; da un altro lato, rappresenta un problema, in quanto implica l'inclusione di informazioni che non provengono dall'esperimento. Requisito della distribuzione a priori è che abbia

per supporto lo spazio parametrico  $\Theta$ . La distribuzione a posteriori ha, qualunque sia la verosimiglianza, un supporto che è un sottoinsieme proprio di quello della distribuzione a priori.

La scelta della distribuzione a priori rappresenta il punto più caratterizzante, ma allo stesso tempo il più delicato e controverso dell'approccio bayesiano (Borra e Di Ciaccio, 2004). Se l'informazione a priori proviene da altre indagini ed è ampiamente condivisa, ossia è oggettivamente disponibile, il suo utilizzo non è in discussione in quanto trascurarla porterebbe a risultati non ottimali. Il problema sorge nel caso di informazione a priori soggettiva, ovvero frutto di valutazioni individuali, che possono differire tra ricercatore e ricercatore. L'uso di informazioni soggettive non è un'eventualità rara.

La scelta della distribuzione iniziale viene influenzata dalla quantità di informazione pre-sperimentale disponibile sul parametro  $\theta$ . Quando quest'informazione è disponibile, si cerca di definire una distribuzione a priori che apporti tale conoscenza al procedimento inferenziale, assegnando una probabilità maggiore ai valori ritenuti più probabili. In questo caso, si parla di distribuzione a priori informativa o soggettiva. Nel caso, invece, di informazioni a priori molto vaghe, vengono identificate delle distribuzioni a priori non informative o oggettive, che hanno un effetto trascurabile sulla distribuzione a posteriori, rispetto a quello proveniente dai dati osservati (Borra e Di Ciaccio, 2004). In questi casi, l'obiettivo dell'analisi è quello di "far parlare i dati" apportando, pertanto, scarsa informazione a quella fornita dal campione, in modo tale da limitare il peso dell'informazione pre-sperimentale (Cifarelli e Muliere, 1989). Si può usare, a tal fine, la distribuzione uniforme o una costante, della forma  $\pi(\theta) \propto \text{costante}$ , per  $\theta \in \Theta$ . Nel caso in cui  $\Theta = [c, +\infty)$  oppure  $\Theta = (-\infty, +\infty)$ , si è in presenza di una distribuzione a priori impropria.

Le distribuzioni a priori improprie sono tali che l'integrale della loro funzione di densità, esteso a tutto il supporto di  $\theta$ , diverge. L'inferenza è valida purché la distribuzione a posteriori che ne deriva sia propria. Se  $\theta \in R^+$ , Jeffreys suggerisce di assumere per  $\pi(\theta)$  una distribuzione uniforme della forma  $\pi(\theta) \propto 1/\theta$  (si veda, ad esempio, Liseo, 2008, Cap.5).

Tra i diversi metodi utilizzati per determinare le distribuzioni iniziali non informative (Cifarelli e Muliere, 1989), il più noto è quello di Jeffreys, che ha proposto di scegliere come distribuzione non informativa per un parametro scalare la funzione

$$\pi(\theta) \propto \sqrt{i(\theta)}, \quad (2.3)$$

dove  $i(\theta)$  indica l'informazione attesa di Fisher. Nel caso in cui  $\theta$  sia un vettore, si può considerare

$$\pi(\theta) \propto \sqrt{|i(\theta)|}, \quad (2.4)$$

dove  $|i(\theta)|$  indica il determinante dell'informazione attesa di Fisher. Altre distribuzioni a priori automatiche sono definite dalle *reference priors* (si veda, ad esempio, Liseo, 2008) o dalle *matching priors* (si veda, ad esempio, Datta e Mukerjee, 2004), che saranno considerate nel prossimo capitolo.

Nella scelta di distribuzioni a priori informative, si fa generalmente riferimento alle classi di distribuzione coniugate (Borra e Di Ciaccio, 2004, Cap.15). Una classe di distribuzioni si definisce coniugata al modello considerato se, per ogni distribuzione a priori all'interno della classe, anche la distribuzione a posteriori appartiene alla medesima classe. Questo significa che la distribuzione a priori e la distribuzione a posteriori hanno la stessa forma funzionale e differiscono solo per il valore assunto dai parametri. Date la funzione di verosimiglianza  $L(\theta; y)$  e data la classe  $R$  di densità iniziali per  $\theta \in \Theta$ , si dice che  $R$  è coniugata al modello se  $\pi(\theta|y) \in R$ . I parametri della distribuzione a priori vengono detti iperparametri per distinguerli da quelli della verosimiglianza. Gli iperparametri vengono scelti arbitrariamente in base alle informazioni disponibili (Liseo, 2008).

L'influenza della distribuzione a priori sulla distribuzione a posteriori diminuisce con l'aumentare della dimensione del campione (Borra e Di Ciaccio, 2004). Pertanto, l'evidenza empirica, quando è supportata da una numerosità campionaria elevata, è in grado di modificare le aspettative iniziali rispetto al valore del parametro.

Siano  $a_n = \hat{\theta} + aj(\hat{\theta})^{-1/2}$  e  $b_n = \hat{\theta} + bj(\hat{\theta})^{-1/2}$ , dove  $\hat{\theta}$  è la stima di massima verosimiglianza di  $\theta$  e  $j(\hat{\theta})$  è l'informazione osservata, calcolata in  $\hat{\theta}$ . Sotto opportune condizioni di regolarità del modello, simili a quelle richieste per la normalità asintotica dello stimatore di massima verosimiglianza, e sotto opportune condizioni sulla distribuzione a priori, la distribuzione a posteriori è asintoticamente normale (si veda, ad esempio, Brazzale et al., 2007). In particolare, vale

$$\int_{a_n}^{b_n} \pi(\theta|y) d\theta \xrightarrow{p} \Phi(b) - \Phi(a), \quad (2.5)$$

per  $n \rightarrow \infty$ . Ne consegue che per  $\pi(\theta|y)$  vale l'approssimazione asintotica  $N(\hat{\theta}, j(\hat{\theta})^{-1})$ . Tale approssimazione vale anche per  $\theta$  multidimensionale.

Inoltre, minore è la variabilità della distribuzione a priori, maggiore è la sua influenza sulla distribuzione a posteriori. La variabilità della distribuzione a priori indica il livello di informazione disponibile a priori: una grande variabilità implica una grande incertezza a priori; viceversa, una distribuzione a priori che assegna probabilità nulla ad alcuni valori del parametro  $\theta$ , implica che anche la distribuzione a posteriori attribuisca probabilità nulla a quei valori.

All'interno dell'approccio bayesiano, si individuano quindi due filosofie, una soggettiva e una oggettiva. L'impostazione soggettiva si basa sulla considerazione che nulla è veramente oggettivo: l'evidenza sperimentale, attraverso l'applicazione del Teorema di Bayes, serve a modificare le opinioni sulle informazioni pre-sperimentali di ciascun individuo. Una difficoltà che si ha nell'utilizzo dell'impostazione soggettiva è data dall'impossibilità, in alcuni casi, di esprimere la valutazione iniziale attraverso una singola distribuzione di probabilità. L'impostazione oggettiva parte dalla considerazione della debolezza applicativa del metodo soggettivo. Le procedure inferenziali dipendono esclusivamente dal modello prescelto e dal risultato campionario. Quest'impostazione si prefigge, da un lato, di conservare l'oggettività delle procedure inferenziali, e dall'altro, come è proprio dell'approccio bayesiano, di usare una legge iniziale che produca conclusioni inferenziali nel linguaggio probabilistico.

Un'ulteriore differenza tra la statistica frequentista e la statistica bayesiana, sta nelle procedure inferenziali. Mentre nella prima esse si classificano in stima puntuale, stima intervallare e verifica d'ipotesi, nella seconda questa distinzione ha meno importanza, dal momento che il risultato dell'inferenza è la distribuzione a posteriori e le procedure inferenziali non sono altro che metodi per riassumere la distribuzione a posteriori.

### **2.3 Metodi parametrici per l'analisi bayesiana della sopravvivenza**

Anche l'approccio bayesiano riconosce la classificazione delle procedure di analisi della sopravvivenza in modelli parametrici, modelli semi-parametrici e modelli non parametrici

(Ibrahim et al, 2001). In questo elaborato, l'attenzione è focalizzata e limitata all'uso dei modelli parametrici.

Come evidenziato nel Paragrafo 1.4, nell'ambito dell'analisi della sopravvivenza, i modelli parametrici più usati sono il modello esponenziale, il modello Weibull, il modello log-normale e il modello Gamma (Ibrahim et al, 2001).

Si supponga di disporre di un campione di tempi di sopravvivenza di dimensione  $n$   $\mathbf{t} = (t_1, \dots, t_n)$ , realizzazioni empiriche della variabile casuale  $T$ , definita nel primo capitolo "tempo di sopravvivenza". Sia  $f(t; \theta)$  la funzione di densità di  $T$ . Si vuole valutare se e come la distribuzione dei tempi di sopravvivenza è influenzata da variabili concomitanti  $(x_{i1}, \dots, x_{ik})$ , dove  $x_{ir}$  è il valore assunto dalla  $r$ -esima variabile concomitante per l' $i$ -esima unità statistica. Come viene mostrato negli esempi successivi, si procede specificando, in primo luogo, un modello parametrico per la distribuzione dei tempi di sopravvivenza. La funzione di verosimiglianza che va a moltiplicare la distribuzione a priori,  $\pi(\theta)$ , deve tenere conto sia della funzione di densità  $f(t; \theta)$ , sia della funzione di sopravvivenza  $S(t; \theta)$ ,  $\theta \in \Theta$ . L'eventuale presenza di censura è indicizzata dal vettore  $\mathbf{v} = (v_1, \dots, v_n)$ , dove  $v_i$  assume valore 0 in presenza di censura e assume valore 1 altrimenti. La funzione di verosimiglianza per  $\theta$  è quindi data da (Pace e Salvan, 2001)

$$L(\theta; \mathbf{t}) = \prod_{i=1}^n f(t_i; \theta)^{v_i} S(t_i; \theta)^{(1-v_i)} . \quad (2.6)$$

Scelta un'opportuna distribuzione a priori, la (2.2) diventa

$$\pi(\theta | \mathbf{t}) \propto \pi(\theta) L(\theta; \mathbf{t}) = \pi(\theta) \cdot \prod_{i=1}^n f(t_i; \theta)^{v_i} S(t_i; \theta)^{(1-v_i)} . \quad (2.7)$$

Negli esempi che seguono la distribuzione a priori,  $\pi(\theta)$ , che va a moltiplicare la (2.6) è scelta essere parametrica e propria (Ibrahim et al, 2001). Nel capitolo successivo, invece, verranno considerate distribuzioni a priori non informative.

### § Esempio: il modello esponenziale

Si supponga di avere un vettore  $\mathbf{t} = (t_1, \dots, t_n)$  di tempi di sopravvivenza, indipendenti e identicamente distribuiti, ciascuno dei quali abbia distribuzione esponenziale con tasso di guasto  $\lambda$ . La funzione di verosimiglianza, definita a partire dalla funzione di densità  $f(t_i; \lambda) = \lambda \exp(-\lambda t_i)$ ,  $i = 1, \dots, n$ , è data da

$$L(\lambda;t)=\prod_{i=1}^n f(t_i; \lambda)^{v_i} S(t_i; \lambda)^{(1-v_i)} = \lambda^d \exp(-\lambda \sum_{i=1}^n t_i), \quad (2.8)$$

dove  $d=\sum_{i=1}^n v_i$  e  $S(t;\lambda)$  è la funzione di sopravvivenza. Utilizzando la distribuzione a priori coniugata naturale al modello esponenziale, ovvero la distribuzione Gamma con parametro di forma  $\alpha>0$  e parametro di scala  $\gamma>0$ , data da

$$\pi(\lambda) \propto \lambda^{\alpha-1} \exp(-\gamma\lambda), \quad (2.9)$$

si ottiene la seguente distribuzione a posteriori per  $\lambda$

$$\pi(\lambda/t) \propto \lambda^{\alpha+d-1} \exp\{-\lambda(\gamma + \sum_{i=1}^n t_i)\}. \quad (2.10)$$

La (2.10) è una distribuzione Gamma di parametri  $(\alpha+d, \gamma+\sum_{i=1}^n t_i)$ .

Nel caso di un modello di regressione è necessario introdurre un vettore di covariate  $\mathbf{x}$ , attraverso la funzione  $g(\mathbf{x}^T \boldsymbol{\beta})$ , dove  $\boldsymbol{\beta}$  è un vettore  $p \times 1$  di coefficienti di regressione e per  $g(\mathbf{x}^T \boldsymbol{\beta})$  si assume, per semplicità, la funzione esponenziale (Ibrahim et al, 2001). La funzione di verosimiglianza per  $\boldsymbol{\beta}$  diventa allora

$$L(\boldsymbol{\beta};t)=\exp\{\sum_{i=1}^n v_i \mathbf{x}_i^T \boldsymbol{\beta}\} \exp\{-\sum_{i=1}^n t_i \exp(\mathbf{x}_i^T \boldsymbol{\beta})\}, \quad (2.11)$$

dove  $\mathbf{x}_i^T$  rappresenta la riga  $i$ -esima della matrice di covariate  $X$  di dimensione  $n \times p$ .

Distribuzioni a priori comuni per  $\boldsymbol{\beta}$  sono la distribuzione a priori impropria, l'uniforme e la distribuzione a priori normale (Ibrahim et al., 2001). Ad esempio, si supponga che il vettore dei coefficienti di regressione  $\boldsymbol{\beta}$  abbia distribuzione a priori normale  $N_p(\mu_0, \Sigma_0)$ , dove  $\mu_0$  indica la media della distribuzione a priori e  $\Sigma_0$  indica la matrice di varianze e covarianze. La distribuzione a posteriori per  $\boldsymbol{\beta}$  è quindi data da

$$\pi(\boldsymbol{\beta}/t) \propto L(\boldsymbol{\beta};t) \pi(\boldsymbol{\beta}; \mu_0, \Sigma_0), \quad (2.12)$$

dove  $\pi(\boldsymbol{\beta}; \mu_0, \Sigma_0)$  è la densità normale multivariata con media  $\mu_0$  e matrice di varianze e covarianze  $\Sigma_0$ . Si ottiene, quindi, la distribuzione a posteriori

$$\pi(\boldsymbol{\beta}/t) \propto \lambda^{\alpha-1} \exp\left\{\sum_{i=1}^n v_i \mathbf{x}_i^T \boldsymbol{\beta} - \sum_{i=1}^n t_i \exp(\mathbf{x}_i^T \boldsymbol{\beta}) - \frac{1}{2}(\boldsymbol{\beta} - \boldsymbol{\mu}_0) \boldsymbol{\Sigma}_0^{-1}(\boldsymbol{\beta} - \boldsymbol{\mu}_0)\right\}. \quad (2.13)$$

### § Esempio: il modello Weibull

Il modello Weibull è forse il modello parametrico più usato per l'analisi di dati di sopravvivenza. Sia dato il campione  $\mathbf{t}=(t_1, \dots, t_n)$  di tempi di sopravvivenza, indipendenti e identicamente distribuiti, ciascuno dei quali ha distribuzione di Weibull  $W(\alpha, \gamma)$ , con parametro di forma  $\alpha > 0$ , parametro di scala  $\gamma > 0$ . Sia  $\lambda = \log \gamma$ .

La funzione di verosimiglianza per  $(\alpha, \lambda)$ , definita a partire dalla funzione di densità  $f(t; \alpha, \lambda) = \alpha t^{\alpha-1} \exp(\lambda - \exp(\lambda) t^\alpha)$ , è data da

$$\begin{aligned} L(\alpha, \lambda; \mathbf{t}) &= \prod_{i=1}^n f(t_i; \alpha, \lambda)^{v_i} S(t_i; \alpha, \lambda)^{(1-v_i)} \\ &= \alpha^d \exp\{d\lambda + \sum_{i=1}^n (v_i(\alpha - 1) \log(t_i) - \exp(\lambda) t_i^\alpha)\}, \end{aligned} \quad (2.14)$$

con  $S(y; \alpha, \lambda) = \exp(-\exp(\lambda) y^\alpha)$  funzione di sopravvivenza.

Quando  $\alpha$  è noto, la distribuzione a priori coniugata per  $\gamma = \exp(\lambda)$  è la Gamma. Quando, invece, sia  $\alpha$  che  $\lambda$  sono ignoti, un'usuale funzione a priori congiunta si ottiene assumendo che i due parametri siano indipendenti e, in particolare, che  $\alpha$  abbia distribuzione Gamma  $Ga(\alpha_0, k_0)$ , e che  $\lambda$  abbia distribuzione normale  $N(\mu_0, \sigma_0^2)$  (Ibrahim et al., 2001). La distribuzione a posteriori per  $(\alpha, \lambda)$  è data dal prodotto della (2.14) per le due distribuzioni a priori, rispettivamente, di  $\alpha$  e di  $\lambda$ . Si ottiene

$$\begin{aligned} \pi(\alpha, \lambda/t) &\propto \alpha^{\alpha_0 + d - 1} \cdot \exp\{d\lambda + \sum_{i=1}^n v_i(\alpha - 1) \log(t_i) - \exp(\lambda) t_i^\alpha - \\ &\quad - k_0 \alpha - \frac{1}{2\sigma_0^2} (\lambda - \mu_0)^2\}. \end{aligned} \quad (2.15)$$

Il modello di regressione di Weibull, si ottiene introducendo le covariate. Assumendo per  $\boldsymbol{\beta}$  una distribuzione a priori  $N_p(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$  e per  $\alpha$  una distribuzione Gamma indipendenti, si ottiene la distribuzione a posteriori



$$\begin{aligned} \pi(\boldsymbol{\beta}, \alpha/t) \propto & \alpha^{\alpha_0+d-1} \exp\left\{\sum_{i=1}^n (v_i \mathbf{x}_i^T \boldsymbol{\beta} + v_i(\alpha - 1) \log(t_i) - t_i^\alpha \exp(\mathbf{x}_i^T \boldsymbol{\beta})) - k_0 \alpha - \right. \\ & \left. - \frac{1}{2} (\boldsymbol{\beta} - \boldsymbol{\mu}_0)^T \boldsymbol{\Sigma}_0^{-1} (\boldsymbol{\beta} - \boldsymbol{\mu}_0)\right\}. \end{aligned} \quad (2.16)$$

### § Esempio: il modello log-normale

Il modello log-normale assume che i logaritmi dei tempi di sopravvivenza abbiano distribuzione normale. Data la funzione di sopravvivenza  $S(t; \mu, \sigma) = 1 - \Phi\left(\frac{\log(t) - \mu}{\sigma}\right)$ , con  $\Phi(\cdot)$  la funzione di densità della distribuzione normale standard, la funzione di verosimiglianza, definita a partire dalla funzione di densità  $f(y; \mu, \sigma) = (2\pi)^{-1/2} (t\sigma)^{-1} \exp\left\{-\frac{1}{2\sigma^2} (\log(t) - \mu)^2\right\}$ , è

$$\begin{aligned} L(\mu, \sigma; t) &= \prod_{i=1}^n f(t_i; \mu, \sigma)^{v_i} S(t_i; \mu, \sigma)^{(1-v_i)} \\ &= (2\pi\sigma^2)^{-d/2} \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^n v_i (\log(t_i) - \mu)^2\right\} \cdot \\ &\quad \cdot \prod_{i=1}^n t_i^{-v_i} \left(1 - \Phi\left(\frac{\log(t_i) - \mu}{\sigma}\right)\right)^{1-v_i}. \end{aligned} \quad (2.17)$$

Posto  $\tau = 1/\sigma^2$ , nel caso in cui entrambi i parametri  $(\mu, \tau)$  siano ignoti, un'usuale distribuzione a priori congiunta prevede che  $\mu/\tau$  abbia distribuzione normale  $N\left(\mu_0, \frac{1}{\tau\tau_0}\right)$ , con  $\tau \sim Ga\left(\frac{\alpha_0}{2}, \frac{\lambda_0}{2}\right)$ . Sotto queste assunzioni, la distribuzione a posteriori congiunta si ottiene come

$$\begin{aligned} \pi(\mu, \tau/t) \propto & L(\mu, \sigma; t) \pi(\mu, \tau/\mu_0, \tau_0, \alpha_0, \lambda_0) \\ \propto & \tau^{\frac{\alpha_0+d}{2}-1} \exp\left\{-\frac{\tau}{2} \left[\sum_{i=1}^n v_i (\log(t_i) - \mu)^2 + \tau_0 (\mu - \mu_0)^2 + \lambda_0\right]\right\} \cdot \\ & \cdot \prod_{i=1}^n y_i^{-v_i} \left(1 - \Phi\left(\tau^{1/2} (\log(t_i) - \mu)\right)\right)^{1-v_i}. \end{aligned} \quad (2.18)$$

Posto  $\boldsymbol{\mu}_i = \mathbf{x}_i^T \boldsymbol{\beta}$  e assunta per  $\boldsymbol{\beta}/\tau$  la distribuzione a priori normale  $N_p(\boldsymbol{\mu}_0, \tau^{-1} \boldsymbol{\Sigma}_0)$ , la distribuzione a posteriori, che tiene conto del vettore di covariate  $\mathbf{x}$ , è

$$\begin{aligned} \pi(\boldsymbol{\beta}, \tau | t) \propto & \tau^{\frac{\alpha_0 + d}{2} - 1} \exp\left\{-\frac{\tau}{2} \left[ \sum_{i=1}^n v_i (\log(t_i) - \mathbf{x}_i^T \boldsymbol{\beta})^2 + (\boldsymbol{\beta} - \boldsymbol{\mu}_0)^T \boldsymbol{\Sigma}_0^{-1} (\boldsymbol{\beta} - \boldsymbol{\mu}_0) + \lambda_0 \right]\right\} \cdot \\ & \cdot \prod_{i=1}^n t_i^{-v_i} \left(1 - \Phi\left(\tau^{1/2} (\log(t_i) - \mathbf{x}_i^T \boldsymbol{\beta})\right)\right)^{1-v_i}. \end{aligned} \quad (2.19)$$

### § Esempio: il modello Gamma

Il modello Gamma è una generalizzazione del modello esponenziale. Definita la funzione di sopravvivenza  $S(t; \alpha, \lambda) = -\frac{1}{\Gamma(\alpha)} \int_0^{t \exp(\lambda)} u^{\alpha-1} \exp(-u) du$ , la funzione di verosimiglianza, definita a partire dalla funzione di densità  $f(t; \alpha, \lambda) = \frac{\lambda^\alpha}{\Gamma(\alpha)} t^{\alpha-1} \exp\{-\lambda t\}$ , è data da

$$\begin{aligned} L(\alpha, \lambda; t) &= \prod_{i=1}^n f(t_i | \alpha, \lambda)^{v_i} S(t_i; \alpha, \lambda)^{(1-v_i)} \\ &= \frac{1}{(\Gamma(\alpha))^d} \exp\{d\alpha\lambda + \sum_{i=1}^n v_i (\alpha \log(t_i) - t_i \exp(\lambda))\} \cdot \\ & \cdot \prod_{i=1}^n t_i^{-v_i} (S(t_i; \alpha, \lambda))^{1-v_i}. \end{aligned} \quad (2.20)$$

Una volta definite le distribuzioni a priori  $\alpha \sim Ga(\alpha_0, k_0)$  e  $\lambda \sim N(\mu_0, \sigma_0^2)$ , nel caso in cui entrambi i parametri  $(\alpha, \lambda)$  siano ignoti e indipendenti, si ottiene la distribuzione a posteriori congiunta

$$\begin{aligned} \pi(\alpha, \lambda | t) \propto & L(\alpha, \lambda; t) \pi(\alpha, \lambda | \alpha_0, k_0, \mu_0, \sigma_0) \\ &= \frac{\alpha^{\alpha_0-1}}{(\Gamma(\alpha))^d} \exp\{d\alpha\lambda + \sum_{i=1}^n v_i (\alpha \log(t_i) - t_i \exp(\lambda))\} \cdot \\ & \cdot \prod_{i=1}^n t_i^{-v_i} (S(t_i; \alpha, \lambda))^{1-v_i} \cdot \exp\left(k_0 \alpha - \frac{1}{2\sigma_0^2} (\lambda - \mu_0)^2\right). \end{aligned} \quad (2.21)$$

Introducendo nel modello Gamma le covariate e utilizzando una distribuzione a priori normale  $\boldsymbol{\beta} \sim N_p(\boldsymbol{\mu}_0, \boldsymbol{\sigma}_0)$ , si ha

$$\begin{aligned} \pi(\alpha, \lambda/t) &\propto \frac{\alpha^{\alpha_0-1}}{(\Gamma(\alpha))^d} \exp\left\{\sum_{i=1}^n v_i [\alpha(\mathbf{x}_i^T \boldsymbol{\beta} + \log(t_i)) - t_i \exp(\mathbf{x}_i^T \boldsymbol{\beta})]\right\} \cdot \\ &\cdot \prod_{i=1}^n t_i^{-v_i} (S(t_i|\alpha, \lambda))^{1-v_i} \cdot \exp\left(k_0 \alpha - \frac{1}{2}(\boldsymbol{\beta} - \boldsymbol{\mu}_0)^T \boldsymbol{\Sigma}_0^{-1}(\boldsymbol{\beta} - \boldsymbol{\mu}_0)\right). \end{aligned} \quad (2.22)$$

In tutti questi esempi è stata utilizzata una distribuzione a priori parametrica propria. E' possibile, in caso di scarsa informazione iniziale circa il parametro d'interesse, utilizzare distribuzioni a priori non informative, quale la distribuzione a priori di Jeffreys.

### § Esempio: distribuzione non informativa di Jeffreys

Si consideri la funzione di verosimiglianza per un modello parametrico

$$\begin{aligned} L(\boldsymbol{\beta}; t) &\propto \prod_{i=1}^n [\lambda_o(t_i) \exp(\mathbf{x}_i^T \boldsymbol{\beta})]^{v_i} \left( S_o(t_i)^{\exp(\mathbf{x}_i^T \boldsymbol{\beta})} \right) \\ &= \prod_{i=1}^n [\lambda_o(t_i) \exp(\mathbf{x}_i^T \boldsymbol{\beta})]^{v_i} \exp\left\{-\sum_{i=1}^n \exp(\mathbf{x}_i^T \boldsymbol{\beta}) \Lambda_o(t_i)\right\}, \end{aligned} \quad (2.23)$$

dove  $S_o(t)$  è la funzione di sopravvivenza *baseline*  $S_o(t) = \exp\left(-\int_0^t \lambda_o(u) du\right) = \exp(-\Lambda_o(t))$ , con  $\Lambda_o(t)$  nota,  $\mathbf{x}$  è un vettore  $p \times 1$  di covariate e  $\boldsymbol{\beta}$  è il vettore  $p \times 1$  dei coefficienti di regressione.

La distribuzione non informativa di Jeffreys, nel caso multidimensionale, è data da  $\pi(\boldsymbol{\beta}) \propto \sqrt{|i(\boldsymbol{\beta})|}$ . La distribuzione a posteriori è quindi data da

$$\pi(\boldsymbol{\beta}; t) \propto \sqrt{|i(\boldsymbol{\beta})|} \cdot \prod_{i=1}^n [\lambda_o(t_i) \exp(\mathbf{x}_i^T \boldsymbol{\beta})]^{v_i} \left( S_o(t_i)^{\exp(\mathbf{x}_i^T \boldsymbol{\beta})} \right). \quad (2.24)$$

## 2.4 Alcune riflessioni

Nel prossimo capitolo si considera il problema della verifica della nullità di un singolo coefficiente di regressione, nell'ambito dei modelli parametrici di regressione per dati di sopravvivenza, presentati in questo capitolo.

L'approccio bayesiano al test d'ipotesi può essere criticato nel caso in cui si voglia verificare la significatività di un singolo coefficiente  $\beta_j$  nell'ambito di un modello di regressione. Tale problema si traduce nel voler verificare l'ipotesi nulla  $H_0: \beta_j = 0$  contro

l'ipotesi alternativa  $H_1: \beta_j \neq 0$ . Come sarà illustrato nel capitolo successivo, la verifica di un siffatto sistema d'ipotesi tramite il fattore di Bayes, può condurre al paradosso di Jeffreys-Lindley, fonte di un notevole dibattito (Liseo, 2008). Da qui, la decisione di ricorrere alla misura di evidenza di Pereira-Stern (Pereira e Stern, 1999).

## CAPITOLO 3

### APPROSSIMAZIONI ASINTOTICHE PER LA MISURA DI EVIDENZA DI PEREIRA-STERN

#### 3.1 Introduzione

La procedura bayesiana più nota per la verifica d'ipotesi è quella basata sul fattore di Bayes. Tale procedura è stata proposta da Jeffreys con lo scopo di quantificare l'evidenza in favore di una determinata ipotesi nulla  $H_0$  (si veda, ad esempio, Kass e Raftery, 1995). Tale criterio, tuttavia, non risulta applicabile quando si utilizzano distribuzioni a priori improprie o distribuzioni a priori vaghe. In tal caso, alternativamente, è possibile ricorrere ad una procedura proposta da Pereira e Stern (Pereira e Stern, 1999).

Sia data la generica distribuzione a posteriori  $\pi(\theta/\mathbf{t}) \propto \pi(\theta)L(\theta; \mathbf{t})$  per un parametro di interesse  $\theta$ , e sia  $\mathbf{t}=(t_1, \dots, t_n)$  un campione casuale semplice di tempi di sopravvivenza, di dimensione  $n$ . Nel seguito, per semplicità di esposizione, ci si limita a presentare la situazione di  $\theta$  scalare. Inoltre, tale assunzione non è limitativa, essendo di interesse lo studio di un singolo coefficiente di regressione. Nel caso di un modello di regressione, si vuole valutare se e come la distribuzione dei tempi di sopravvivenza è influenzata da variabili concomitanti  $(x_{i1}, \dots, x_{ik})$ , dove  $x_{ir}$  è il valore assunto dalla  $r$ -esima variabile concomitante per l' $i$ -esima unità statistica (si veda il Paragrafo 2.3). Si vuole, in questo contesto, verificare l'ipotesi di nullità di un singolo coefficiente di regressione  $\beta_j=0$  contro l'ipotesi alternativa  $\beta_j \neq 0$  ai fini di studiare l'evidenza campionaria del regressore  $x_j$  sul tempo di sopravvivenza.

Di seguito, si richiamano le procedure di verifica d'ipotesi di Jeffreys e di Pereira-Stern. Per il calcolo di quest'ultima, viene illustrata una procedura di approssimazione di ordine elevato con particolare interesse per l'uso della distribuzione a priori "matching", e viene svolto uno studio di simulazione per studiare l'accuratezza di tale procedura.

### 3.2 Il fattore di Bayes

La procedura proposta da Jeffreys è basata sul fattore di Bayes, che è definito come il rapporto tra le densità a posteriori, calcolate sotto le due differenti ipotesi  $H_0$  e  $H_1$ . Sia  $\theta$  parametro di interesse scalare e sia  $\pi_0$  la probabilità a priori assegnata all'ipotesi nulla  $H_0: \theta = \theta_0$ . La probabilità assegnata all'ipotesi alternativa  $H_1$  è, invece, distribuita su tutti i valori di  $\theta$  diversi da  $\theta_0$ . E' quindi possibile scrivere la distribuzione a priori complessiva come

$$\pi(\theta) = \pi_0 I_{\theta = \theta_0}(\theta) + (1 - \pi_0) g(\theta) I_{\theta \neq \theta_0}(\theta), \quad (3.1)$$

dove  $I_{\theta = \theta_0}(\theta)$  è la funzione indicatrice che vale 1 per  $\theta = \theta_0$  e 0 altrimenti, mentre  $I_{\theta \neq \theta_0}(\theta)$  vale 1 per tutti i valori di  $\theta$  diversi da  $\theta_0$  e 0 per  $\theta = \theta_0$ , e  $g(\theta)$  è la densità a priori su  $\theta$  condizionata all'ipotesi alternativa  $H_1$ .

Il fattore di Bayes è definito come

$$BF = \frac{L(\theta_0; \mathbf{t})}{\int_{\theta \neq \theta_0} L(\theta; \mathbf{t}) g(\theta) d\theta}, \quad (3.2)$$

dove  $L(\theta_0; \mathbf{t})$  è la verosimiglianza sotto  $H_0$ .

Il fattore di Bayes è a favore dell'ipotesi nulla nel caso in cui esso assume valori elevati. In particolare, se  $BF > 1$  si è in presenza di evidenza a favore di  $H_0$  e se  $BF < 1$  c'è evidenza contro l'ipotesi nulla in favore dell'ipotesi alternativa (si veda, ad esempio, Kass e Raftery, 1995).

Una decisione basata su una procedura di inferenza bayesiana è determinata da una funzione di perdita ad essa associata. La funzione di perdita, essenzialmente, riflette le conseguenze negative associate alla decisione "sbagliata". Nell'ambito del problema di verifica d'ipotesi, assunte le funzioni di perdita  $L(\text{Accetto } H_0, \theta) = a I_{\theta \neq \theta_0}(\theta)$  e  $L(\text{Rifiuto } H_0, \theta) = b I_{\theta = \theta_0}(\theta)$ , con  $a, b > 0$ , si decide in favore di  $H_0$  se il rischio a posteriori di accettare l'ipotesi nulla è molto piccolo. Secondo questo approccio, pertanto, si decide in favore di  $H_0$  se

$$P(H_0 | \mathbf{t}) > \frac{a}{a+b}. \quad (3.3)$$

Esiste, naturalmente, una relazione tra il fattore di Bayes definito nella (3.2) e la probabilità finale assegnata all'ipotesi nulla  $Pr(H_0/\mathbf{t})$ . Infatti, per  $Pr(H_0/\mathbf{t})$  si ha

$$Pr(H_0/\mathbf{t}) = \frac{\pi_0 L(\theta_0; \mathbf{t})}{\pi_0 L(\theta_0; \mathbf{t}) + (1 - \pi_0) \int_{\theta \neq \theta_0} L(\theta; \mathbf{t}) g(\theta) d\theta}. \quad (3.4)$$

Dividendo la (3.4) al il numeratore e al denominatore per  $\pi_0 L(\theta_0; \mathbf{t})$ , si ottiene

$$Pr(H_0/\mathbf{t}) = \left( 1 + \frac{1 - \pi_0}{\pi_0} BF^{-1} \right)^{-1}. \quad (3.5)$$

E' ben noto che, il fattore di Bayes, nel caso si usino distribuzioni a priori improprie o distribuzioni a priori proprie ma vaghe, porta al paradosso di Jeffreys-Lindley (si veda, ad esempio, Robert, 1993). Con riguardo alla (3.2) accade che, se la densità a priori  $g(\theta)$  condizionata all'ipotesi alternativa  $H_1$  è diffusa su tutto lo spazio dei parametri  $\Theta - \{\theta_0\}$ , non concentrandosi su alcun valore in particolare, allora l'ipotesi nulla viene automaticamente accettata. Il paradosso di Jeffreys-Lindley viene spesso usato per argomentare l'impossibilità di usare distribuzioni a priori improprie o vaghe in un problema di verifica d'ipotesi. Ad esempio, si consideri la legge a priori impropria, della forma

$$g(\theta) \propto k \cdot h(\theta), \quad (3.6)$$

con  $h(\theta)$  funzione non integrabile su  $\Theta$  che rende la costante  $k$  non esattamente determinabile. Il fattore di Bayes è dato da

$$BF = \frac{L(\theta_0; \mathbf{t})}{k \int_{\theta \neq \theta_0} L(\theta; \mathbf{t}) h(\theta) d\theta}, \quad (3.7)$$

e, poiché dipende dalla costante indeterminata  $k$ , è anch'esso indeterminato. Dal momento che la probabilità finale associata all'ipotesi nulla  $H_0$ , data da  $Pr(H_0/\mathbf{t})$ , dipende dal fattore di Bayes, come illustrato nella (3.5), risulta anch'essa indeterminata in presenza di distribuzioni a priori improprie o vaghe.

Questo problema impedisce l'uso di distribuzioni improprie o vaghe nelle procedure bayesiane di verifica d'ipotesi, in quanto il fattore di Bayes risente della presenza di costanti di normalizzazione non eliminabili (si veda, ad esempio, Liseo, 2008, cap.6).

Una possibile soluzione consiste nel suddividere il campione in due parti. La prima parte viene utilizzata per determinare, a partire dalla distribuzione a priori impropria o vaga, la corrispondente distribuzione a posteriori. Tale distribuzione viene considerata come distribuzione a priori propria e utilizzata nel calcolo della verosimiglianza media, condizionata alle osservazioni relative alla seconda parte del campione. Il fattore di Bayes è, pertanto, calcolato solo sulla seconda parte del campione e viene chiamato fattore di Bayes parziale (Piccinato, 1999).

Un'altra soluzione, come precedentemente detto, è rappresentata dalla procedura di evidenza di Pereira e Stern. Tale soluzione viene presentata nel prossimo paragrafo.

### 3.3 La procedura di Pereira-Stern

Pereira e Stern hanno introdotto una misura di evidenza in favore dell'ipotesi nulla  $H_0$ , nota come *misura di evidenza di Pereira-Stern* (Pereira e Stern, 1999).

Sia  $\Theta_0$  un sottoinsieme di  $\Theta$ , avente misura di Lebesgue nulla. Dal momento che si vuole verificare l'ipotesi nulla  $H_0: \theta = \theta_0$ , si ha che  $\Theta_0$  coincide con  $\theta_0$ . Sia, inoltre,  $\pi(\theta|\mathbf{t})$  la densità a posteriori di  $\theta$ , dato il campione di tempi di sopravvivenza  $\mathbf{t}$ , sia  $\pi(\theta_0|\mathbf{t})$  la densità a posteriori nel punto  $\theta_0$ , e sia  $T(\mathbf{t}) = \{\theta: \pi(\theta|\mathbf{t}) > \pi(\theta_0|\mathbf{t})\}$ . La *misura di evidenza di Pereira-Stern* è definita come (si veda anche la Figura 3.1)

$$EV(\theta_0, \mathbf{t}) = 1 - Pr_{\pi} [\theta \in T(\mathbf{t}) | \mathbf{t}]. \quad (3.8)$$

La procedura di Pereira-Stern mostra evidenza in favore dell'ipotesi nulla  $H_0$  se  $EV(\theta_0, \mathbf{t})$  è grande. Essa considera in favore di una precisa ipotesi tutti i punti dello spazio parametrico i cui valori della densità a posteriori sono, al massimo, grandi come  $\theta_0$ . Considera, pertanto, tutti i punti che sono "meno probabili" rispetto a  $\theta_0$ . Un valore grande di  $EV(\theta_0, \mathbf{t})$  implica che  $\theta_0$  si trovi in una regione di  $\Theta$  ad alta probabilità e, pertanto, la distribuzione a posteriori  $\pi(\theta|\mathbf{t})$  supporta l'ipotesi nulla. Viceversa, un valore piccolo di  $EV(\theta_0, \mathbf{t})$  indica che  $\theta_0$  si trova in una regione di  $\Theta$  a bassa probabilità, quindi  $\pi(\theta|\mathbf{t})$  mostra evidenza contro l'ipotesi nulla a favore dell'ipotesi alternativa.



Nella Figura 3.1 viene fornita la rappresentazione grafica della (3.8), ponendo in evidenza le regioni che costituiscono  $EV(\theta_0, \mathbf{t})$ :  $T(\mathbf{t})$  è rappresentato dalla porzione di ascissa evidenziata, compresa tra  $\theta_0$  e  $\theta_1$ , in quanto è l'insieme dei  $\theta$  per i quali  $\pi(\theta|\mathbf{t}) > \pi(\theta_0|\mathbf{t})$ . E' evidente che  $EV(\theta_0, \mathbf{t})$  può essere scritta come

$$EV(\theta_0, \mathbf{t}) = Pr_{\pi}(\theta \leq \theta_0|\mathbf{t}) + 1 - Pr_{\pi}(\theta \leq \theta_1|\mathbf{t}). \quad (3.9)$$

La (3.9) vale se  $\theta_0$  è inferiore alla moda della distribuzione a posteriori. Se così non fosse, si avrebbe  $EV(\theta_0, \mathbf{t}) = Pr_{\pi}(\theta \leq \theta_1|\mathbf{t}) + 1 - Pr_{\pi}(\theta \leq \theta_0|\mathbf{t})$ . Per semplicità di esposizione, si considera solo il caso in cui  $\theta_0$  sia maggiore della moda a posteriori.

Si indichi con  $LP$  il sistema di funzioni di perdita  $L(\text{Accetto } H_0, \theta) = b + cI(\theta \in T(\mathbf{t}))$  e  $L(\text{Rifiuto } H_0, \theta) = a [1 - I(\theta \in T(\mathbf{t}))]$ , con  $a, b, c > 0$ . Tali funzioni di perdita dipendono dal campione di osservazioni  $\mathbf{t}$ . Esse "puniscono pesantemente" colui che deve effettuare la verifica d'ipotesi, nel caso in cui accetti l'ipotesi nulla quando è più probabile che  $\theta$  sia, invece, diverso da  $\theta_0$ , ovvero quando  $\theta$  appartiene a  $T(\mathbf{t})$ .

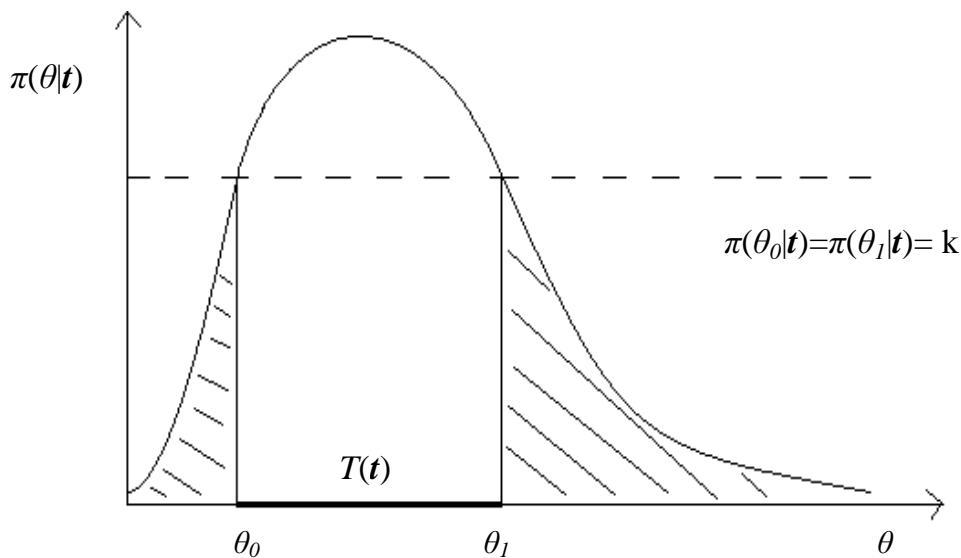


Figura 3.1: Misura di Evidenza di Pereira e Stern  $EV(\theta_0, \mathbf{t})$ : parte tratteggiata.

Il rischio a posteriori di accettare l'ipotesi nulla è allora dato da

$$E_{\pi}[L(\text{Accetto } H_0, \theta)|\mathbf{t}] = \int_{\theta} [b + cI(\theta \in T(\mathbf{t}))]\pi(\theta|\mathbf{t})d\theta = b + c(1 - EV(\theta_0, \mathbf{t})). \quad (3.10)$$

Il rischio a posteriori di rifiutare l'ipotesi nulla è, invece, dato da

$$E_{\pi}[L(\text{Rifiuto } H_0, \theta)|\mathbf{t}] = \int_{\theta} a[1 - I(\theta \in T(\mathbf{t}))]\pi(\theta|\mathbf{t})d\theta = aEV(\theta_0, \mathbf{t}). \quad (3.11)$$

Si ha, pertanto, evidenza a favore di  $H_0$  se e solo se il rischio a posteriori di accettare l'ipotesi nulla è minore rispetto al rischio a posteriori di rifiutare la stessa. Si decide, pertanto, in favore dell'ipotesi nulla se

$$EV(\theta_0, \mathbf{t}) > \frac{b+c}{a+c}. \quad (3.12)$$

Si nota che se  $a < b$ , allora non si decide mai in favore dell'ipotesi nulla, dal momento che  $EV(\theta_0, \mathbf{t})$  assume valori compresi nell'intervallo  $[0, 1]$ . Questo avviene se per il decisore il rifiuto dell'ipotesi nulla è più plausibile rispetto all'accettazione della stessa. Viceversa, se  $a$  è molto più grande di  $b$  e se  $c$  è piccolo, allora l'accettazione dell'ipotesi nulla non richiede un valore alto di  $EV(\theta_0, \mathbf{t})$ . Questo accade se l'accettazione dell'ipotesi nulla è più plausibile rispetto al suo rifiuto.

La procedura di Pereira e Stern è a tutti gli effetti una procedura bayesiana in quanto implica l'utilizzo della distribuzione a posteriori  $\pi(\theta|\mathbf{t})$  (Madruga et al., 2001).

Dalla (3.9) e dalla Figura 3.1 è evidente che, per determinare  $EV(\theta_0, \mathbf{t})$ , è necessario calcolare probabilità a posteriori della forma  $Pr_{\pi}(\theta \leq \theta_0|\mathbf{t})$ . Nel prossimo paragrafo, si presenta un'approssimazione, accurata anche per piccoli campioni, della probabilità d'interesse. Questo procedimento prende il nome di approssimazione delle probabilità sulle code, ossia *tail area approximation* (si vedano, ad esempio, Brazzale et al., 2007 e Reid, 2003). Tali approssimazioni saranno quindi usate per calcolare la misura di evidenza di Pereira e Stern.

### 3.3.1 Tail area approximation

Di seguito, si presenta l'approssimazione dell'area definita da  $Pr_{\pi}(\theta \leq \theta_0|\mathbf{t})$ .

Si riprenda il risultato asintotico del primo ordine presentato nel Paragrafo 2.2 definito nella (2.5). Si assuma, per semplicità,  $\theta_0$  minore della media a posteriori. Secondo tale approssimazione, la misura di Pereira-Stern può essere approssimata semplicemente come

$$EV_N(\theta_0, \mathbf{t}) = 2 \cdot Pr_{\pi}(\theta \leq \theta_0 | \mathbf{t}) \cong 2 \cdot Pr_{\pi}\left(\frac{\theta - \hat{\theta}}{\sqrt{j(\hat{\theta})^{-1}}} \leq \frac{\theta_0 - \hat{\theta}}{\sqrt{j(\hat{\theta})^{-1}}}\right), \quad (3.13)$$

dove  $j(\hat{\theta})$  è l'informazione osservata in  $\hat{\theta}$ .

La tecnica di approssimazione di Laplace (si veda, ad esempio, Pace e Salvan, 1996, Cap.9) rappresenta un utile strumento ai fini dell'approssimazione della distribuzione a posteriori per  $\theta$ . In particolare, per  $\theta$  scalare, l'espansione in serie di Taylor dell'integrale presente al denominatore della (2.1), porta all'approssimazione

$$\pi(\theta | \mathbf{t}) = \{\sqrt{(2\pi)}\}^{-1} |j(\hat{\theta})|^{1/2} \exp\{l(\theta) - l(\hat{\theta})\} \{\pi(\theta)/\pi(\hat{\theta})\} \{1 + O(n^{-1})\}. \quad (3.14)$$

A partire dalla (3.14), l'area d'interesse nella coda è approssimabile nel seguente modo (si veda, ad esempio, Brazzale et al., 2007)

$$\int_{-\infty}^{\theta_0} \pi(\theta | \mathbf{t}) d\theta \doteq \int_{-\infty}^{r_0} c\phi(r) \left(\frac{r}{q_B}\right) dr = \Phi(r_0) + \phi(r_0) \left(\frac{1}{r_0} - \frac{1}{q_{B0}}\right), \quad (3.15)$$

dove  $r_0 = r(\theta_0) = \text{sign}(\hat{\theta} - \theta_0) \left[2(l(\hat{\theta}) - l(\theta_0))\right]^{1/2}$  è la radice con segno del log-rapporto di verosimiglianza,  $\phi(\cdot)$  è la funzione di densità della distribuzione normale standard, e  $q_{B0} = q_B(\theta_0)$  è definito come

$$q_B(\theta) = l'(\theta) j(\hat{\theta})^{-1/2} \frac{\pi(\hat{\theta})}{\pi(\theta)}. \quad (3.16)$$

Nella (3.15) il simbolo “ $\doteq$ ” indica che l'ordine di approssimazione è  $O(n^{-3/2})$ .

La (3.15) può essere equivalentemente espressa come

$$\int_{-\infty}^{\theta_0} \pi(\theta | \mathbf{t}) d\theta \doteq \Phi(r_B^*(\theta_0)), \quad (3.17)$$

dove

$$r_B^*(\theta) = r(\theta) + \frac{1}{r(\theta)} \log \frac{q_B(\theta)}{r(\theta)}. \quad (3.18)$$

Si supponga ora che il parametro complessivo  $\theta$  sia partizionato come  $\theta=(\psi, \lambda)$ , dove il parametro  $\psi$  è un parametro scalare di interesse, mentre  $\lambda$ , di dimensione  $(d-1)$ , è il parametro di disturbo. Partendo dalla distribuzione marginale della densità a posteriori per  $\psi$  e utilizzando l'approssimazione di Laplace (si veda, ad esempio, Brazzale et al., 2007) si trova

$$\begin{aligned} \pi_m(\psi|\mathbf{t}) &= \int \pi(\theta|\mathbf{t}) d\lambda = \frac{\int L(\theta;\mathbf{t})\pi(\theta)d\lambda}{\int L(\theta;\mathbf{t})\pi(\theta)d\theta} \\ &\doteq \frac{1}{\sqrt{2\pi}} |j_p(\hat{\psi})|^{1/2} \exp \{l_p(\psi) - l_p(\hat{\psi})\} \rho(\psi, \hat{\psi}) \frac{\pi(\psi, \hat{\lambda}_\psi)}{\pi(\hat{\psi}, \hat{\lambda})}, \end{aligned} \quad (3.19)$$

con

$$\rho(\psi, \hat{\psi}) = \left[ \frac{|j_{\lambda\lambda}(\hat{\psi}, \hat{\lambda})|}{|j_{\lambda\lambda}(\psi, \hat{\lambda}_\psi)|} \right]^{1/2}, \quad (3.20)$$

$j_p(\psi)$  informazione osservata profilo e  $l_p(\psi) = l(\psi, \hat{\lambda}_\psi)$  funzione di log-verosimiglianza profilo relativa a  $\psi$ , con  $\hat{\lambda}_\psi$  valore di  $\lambda$  che rende massima la log-verosimiglianza per un fissato valore di  $\psi$ . Nella (3.20),  $j_{\lambda\lambda}(\psi, \lambda)$  identifica l'elemento di posizione  $(\lambda, \lambda)$  nella matrice d'informazione osservata  $j(\psi, \lambda)$ . L'approssimazione (3.19) è derivata utilizzando la (3.14) per il denominatore e una versione analoga alla (3.14), con  $\psi$  tenuto fisso, al numeratore.

A partire dalla (3.19), si ottiene la seguente *tail area approximation* (si veda, ad esempio, Reid, 2003)

$$\begin{aligned} Pr_{\pi}(\psi \leq \psi_0|\mathbf{t}) &= \int_{-\infty}^{\psi_0} \pi_m(\psi|\mathbf{t}) d\psi \\ &\doteq \Phi \left( r_p(\psi_0) + \frac{1}{r_p(\psi_0)} \log \frac{q_B(\psi_0)}{r_p(\psi_0)} \right) \\ &= \Phi(r_{Bp}^*(\psi_0)), \end{aligned} \quad (3.21)$$

dove  $r_p(\psi) = \text{sign}(\hat{\psi} - \psi) [2(l_p(\hat{\psi}) - l_p(\psi))]^{1/2}$  è la radice con segno del log-rapporto di verosimiglianza profilo, e

$$q_B(\psi) = l'_p(\psi) j_p(\hat{\psi})^{-1/2} \rho(\psi, \hat{\psi})^{-1} \frac{\pi(\hat{\psi}, \hat{\lambda})}{\pi(\psi, \hat{\lambda}_\psi)}. \quad (3.22)$$

Un importante vantaggio dell'approssimazione (3.21) è dato dal fatto che essa fornisce risultati accurati anche nel caso in cui la dimensione di  $\lambda$  sia elevata, con la dimensione del campione  $n$  piccola.

Inoltre, a partire dalla (3.21), volendo calcolare la mediana a posteriori  $\psi_{ME}$ , quello che si ottiene è che  $\psi_{ME}$  deve essere tale per cui  $\Phi(r_{Bp}^*(\psi_{ME})) = 0.5$ . Quest'equazione è soddisfatta per  $r_{Bp}^*(\psi_{ME}) = 0$ .

Infine, volendo calcolare un intervallo di credibilità HPD (*highest posterior density*), ossia ad alta probabilità a posteriori, per  $\psi$  con livello approssimato  $(1-\alpha)$  si può considerare l'insieme

$$\text{HPD} = \left\{ \psi : |r_{Bp}^*(\psi)| \leq z_{1-\alpha/2} \right\}, \quad (3.23)$$

dove  $z_{1-\alpha/2}$  è il quantile  $(1 - \alpha/2)$  della distribuzione normale standard.

### 3.4 Approssimazioni asintotiche per la procedura di Pereira-Stern

Con riferimento alla Figura 3.1, è di interesse applicare le approssimazioni (3.17) e (3.21) alla misura di evidenza di Pereira-Stern. Si consideri il caso di  $\theta$  scalare e sia  $\theta_0$  minore della media a posteriori. Allora, a partire dalla (3.17), si trova

$$\begin{aligned} EV(\theta_0, \mathbf{t}) &= Pr_\pi(\theta \leq \theta_0 | \mathbf{t}) + 1 - Pr_\pi(\theta \leq \theta_1 | \mathbf{t}) \\ &\doteq \Phi(r_B^*(\theta_0)) + 1 - \Phi(r_B^*(\theta_1)). \end{aligned} \quad (3.24)$$

Nel caso, invece, in cui  $\theta = (\psi, \lambda)$ , con  $\psi$  parametro scalare d'interesse e  $\lambda$  parametro di disturbo, si trova

$$\begin{aligned} Pr_\pi(\psi \geq \psi_1 | \mathbf{t}) &= 1 - Pr_\pi(\psi \leq \psi_1 | \mathbf{t}) \\ &\doteq 1 - \Phi(r_{Bp}^*(\psi_1)). \end{aligned} \quad (3.25)$$

La misura di evidenza di Pereira-Stern è, quindi, esprimibile come

$$EV(\psi_0, \mathbf{t}) = Pr_{\pi}(\psi \leq \psi_0 | \mathbf{t}) + 1 - Pr_{\pi}(\psi \leq \psi_1 | \mathbf{t}) \\ \doteq \Phi(r_{Bp}^*(\psi_0)) + 1 - \Phi(r_{Bp}^*(\psi_1)). \quad (3.26)$$

E' fondamentale precisare che il procedimento sopra illustrato, su cui si basano le approssimazioni nelle probabilità nelle code, è valido, a differenza della procedura basata sul fattore di Bayes, qualunque sia la distribuzione a priori  $\pi(\theta)$  utilizzata nella distribuzione a posteriori. Pertanto, vale anche per distribuzioni a priori improprie o vaghe. La misura di evidenza di Pereira-Stern, invece, richiede la scelta di una distribuzione a priori che conduca a una distribuzione a posteriori propria.

Si nota, inoltre, che se la distribuzione a posteriori, così come mostrata in Figura 3.1, è simmetrica nella moda a posteriori, allora si trova che le due aree nelle code sono uguali. La misura di evidenza di Pereira-Stern è, in questo contesto, semplicemente data da

$$EV(\theta_0, \mathbf{t}) = 2\Phi(r_B^*(\theta_0)) \quad (3.27)$$

o data da

$$EV(\psi_0, \mathbf{t}) = 2\Phi(r_{Bp}^*(\psi_0)), \quad (3.28)$$

in presenza di parametri di disturbo.

### 3.4.1 Esempio: *matching priors*

Si consideri la classe delle distribuzioni a priori “*matching*” nel calcolo della (3.21) e, quindi, della (3.26). Tali distribuzioni a priori sono, per definizione, tali per cui

$$Pr_{\pi}\{\theta \leq \theta^{(1-\alpha)}(t)\} = Pr\{\theta^{(1-\alpha)}(t) \geq \theta\}\{1 + O(n^{-1})\}, \quad (3.29)$$

dove  $O(n^{-1})$  è l'ordine di errore (si veda, ad esempio, Datta e Mukerjee, 2004).

Per  $\theta$  scalare, la distribuzione a priori “*matching*” è definita come

$$\pi(\theta) \propto i(\theta)^{1/2}, \quad (3.30)$$

dove  $i(\theta)$  è l'informazione attesa di Fisher. In corrispondenza di tale distribuzione a priori, si ha

$$r^*(\theta) = r(\theta) + \frac{1}{r(\theta)} \log \frac{q_B(\theta)}{r(\theta)}, \quad (3.31)$$

con

$$q(\theta) = l'(\theta) j(\hat{\theta})^{-1/2} \frac{i(\hat{\theta})^{1/2}}{i(\theta)^{1/2}}. \quad (3.32)$$

Nel caso di presenza di parametri di disturbo, la distribuzione a priori “matching” è esprimibile come (si vedano, ad esempio, Ventura et al., 2009, e Ventura e Racugno, 2011)

$$\pi(\psi) \propto i_{\psi\psi,\lambda}(\psi, \hat{\lambda}_\psi)^{1/2}, \quad (3.33)$$

con  $i_{\psi\psi,\lambda}(\psi, \lambda) = i_{\psi\psi}(\psi, \lambda) - i_{\psi\lambda}(\psi, \lambda) i_{\lambda\lambda}(\psi, \lambda)^{-1} i_{\lambda\psi}(\psi, \lambda)$  informazione parziale definita a partire dall'informazione attesa di Fisher  $i(\psi, \lambda)$ .

In corrispondenza di tale distribuzione a priori, si giunge alla (3.21), con  $q_B(\psi) = q_F(\psi)$  dato da

$$q_F(\psi) = \frac{l'_p(\psi)}{j_p(\hat{\psi})^{1/2}} \frac{i_{\psi\psi,\lambda}(\hat{\psi}, \hat{\lambda})^{1/2}}{i_{\psi\psi,\lambda}(\psi, \hat{\lambda}_\psi)^{1/2}} \frac{|l_{\lambda;\hat{\lambda}}(\psi, \hat{\lambda}_\psi)|}{|j_{\lambda\lambda}(\hat{\psi}, \hat{\lambda})|^{1/2} |j_{\lambda\lambda}(\psi, \hat{\lambda}_\psi)|^{1/2}}, \quad (3.34)$$

dove  $l_{\lambda;\hat{\lambda}}(\psi, \lambda) = \frac{\partial l(\psi, \lambda)}{(\partial \lambda \partial \hat{\lambda}^T)}$ . Tale espressione è stata anche utilizzata da Barndorff-Nielsen e Chamberlin (Barndorff-Nielsen e Chamberlin, 1994) nella definizione frequentista di una versione modificata di  $r_p(\psi)$ , data da

$$r_p^*(\psi) = r_p(\psi) + \frac{1}{r_p(\psi)} \log \frac{q_F(\psi)}{r_p(\psi)}. \quad (3.35)$$

Pertanto, in tal caso, l'utilizzo di una distribuzione a priori appartenente alla classe delle “*matching priors*” garantisce l'equivalenza tra un  $p$ -value frequentista e la probabilità individuata dalla *tail area* (coda) della distribuzione a posteriori definita nell'approccio bayesiano. Un risultato importante è dato dal fatto che l'intervallo HPD, così come definito nell'approccio bayesiano, e l'intervallo di confidenza basato su  $r_p^*(\psi)$ , definito nell'approccio frequentista, coincidono.

Inoltre, usando i risultati in Ventura et al. (2011), usando la *matching prior* (3.33), la distribuzione marginale a posteriori può essere scritta come

$$\pi_m(\psi|\mathbf{t}) \propto \exp\left(-\frac{1}{2}r_p^*(\psi)^2\right) \cdot \left|\frac{r_p(\psi)}{s_p(\psi)}\right| O(n^{-1}), \quad (3.36)$$

dove  $s_p(\psi) = \frac{l'_p(\psi)}{j_p(\hat{\psi})^{1/2}}$  è la statistica *score*. Ma, poiché  $r_p(\psi) = s_p(\psi) + o_p(1)$ , la (3.36), al primo ordine di approssimazione, diventa semplicemente

$$\pi_m(\psi|\mathbf{t}) \propto \exp\left(-\frac{1}{2}r_p^*(\psi)^2\right). \quad (3.37)$$

Un vantaggio delle approssimazioni (3.36) e (3.37) è che includono automaticamente la distribuzione a priori *matching* definita dalla (3.33) senza richiederne il calcolo esplicito.

Nel prossimo paragrafo viene studiata l'accuratezza dell'approssimazione asintotica della misura di evidenza di Pereira-Stern illustrata nel presente capitolo, tramite studi di simulazione Monte Carlo. A questo scopo, viene utilizzata la libreria HOA del software statistico R nel caso di un modello di regressione per dati di sopravvivenza.

### 3.5 Studi di simulazione e un esempio

HOA (*higher order asymptotics*) è una libreria del software statistico R che implementa l'inferenza di ordine elevato per tre classi di modelli ampiamente utilizzate nelle applicazioni: i modelli logistici e loglineari, i modelli lineari non normali e i modelli di



regressione non lineari eteroschedastici (Brazzale, 2005). Il codice corrispondente è organizzato in tre pacchetti, chiamati rispettivamente `cond`, `margin` e `nlreg`.

Per poter utilizzare il pacchetto `hoa` è stata utilizzata la versione 2.14.2 di R per Windows.

Lo studio per dati di sopravvivenza presentato in questo paragrafo è stato effettuato tramite il pacchetto `margin`. La possibilità di ricorrere a questi pacchetti è legata all'utilizzo delle *matching priors* nelle distribuzioni a posteriori, e all'uso delle corrispondenti approssimazioni asintotiche per la misura di evidenza di Pereira-Stern.

### 3.5.1 *L'analisi*

Obiettivo è studiare l'accuratezza delle approssimazioni di ordine elevato presentate nel Paragrafo 3.4 nell'ambito di un modello di Weibull (Paragrafo 2.3).

Si consideri il campione di tempi di sopravvivenza  $t = (65, 156, 100, 134, 16, 108, 121, 4, 39, 143, 56, 26, 22, 1, 1, 5, 65)$  misurati in settimane, rilevati per  $n=17$  pazienti affetti da leucemia. Al momento della diagnosi viene rilevato, per ciascun paziente, il numero di globuli bianchi  $x$  (Cox e Snell, 1981, esempio U). Si osserva  $\mathbf{x} = (3.36, 2.88, 3.63, 3.41, 3.78, 4.02, 4, 4.23, 3.73, 3.85, 3.97, 4.51, 4.54, 5, 5, 4.72, 5)$ . Si consideri, per i tempi di sopravvivenza, un modello di Weibull con parametro di forma  $k$  e  $\lambda_i = \beta_1 \exp(\beta_2(x_i - \bar{x}))$ , per  $i=1, \dots, n$ . Per  $y_i = \log t_i$ , si dimostra che  $y_i = x_i^T \beta + \sigma \varepsilon_i = \log \beta_1 + \beta_2(x_i - \bar{x}) + \sigma \varepsilon_i$ , segue un modello di regressione e scala, ovvero identifica un modello lineare non normale, dove  $\sigma = 1/k$  e  $\varepsilon_i \sim f(\varepsilon_i) = \exp(\varepsilon_i - \exp(\varepsilon_i))$  è una variabile casuale log-Weibull, chiamata anche variabile casuale di Gumbel.

Si vuole verificare la nullità di un singolo coefficiente di regressione utilizzando i risultati asintotici di cui al Paragrafo 3.4. In particolare, si vuole verificare l'ipotesi nulla  $H_0: \psi = \log \sigma = 0$  in un primo momento e, in un secondo momento, l'ipotesi nulla  $H_0: \psi = \beta_2 = 0$  per studiare la significatività della variabile concomitante (il numero di globuli bianchi nell'esempio) sulla distribuzione della sopravvivenza.

A questo scopo si riprende l'approssimazione (3.13) che definisce la misura di evidenza  $EV_N(\psi_0, \mathbf{y})$ , calcolata a partire dalla distribuzione marginale (2.5), e si riprende la misura di evidenza definita dalla (3.26), indicata, per semplicità, con  $EV_*(\psi_0, \mathbf{y})$ , data da

$$EV_*(\psi_0, \mathbf{y}) = 2\Phi(r_p^*(\psi_0)), \tag{3.38}$$

assumendo  $\psi_0$  minore della media a posteriori.

C'è evidenza a favore dell'ipotesi nulla se la misura di evidenza considerata è maggiore di un certo valore soglia  $\alpha$ .

La Figura 3.2 mette a confronto per  $\psi = \log\sigma$  la distribuzione a posteriori marginale definita nella (3.37) con la distribuzione a posteriori marginale definita dalla (2.5).

Quest'ultima è di seguito indicata con la notazione  $\pi_m^{fo}(\psi|\mathbf{y})$  per semplicità.

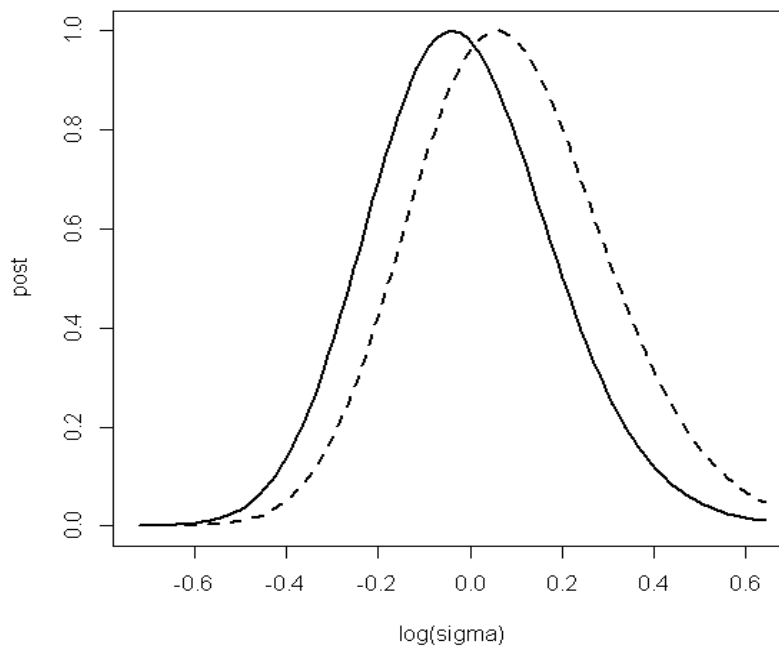


Figura 3.2: Distribuzione a posteriori marginale  $\pi_m(\psi|\mathbf{y})$  (linea continua) e distribuzione a posteriori  $\pi_m^{fo}(\psi|\mathbf{y})$  (linea tratteggiata) per  $\psi = \log\sigma$ .

Dalla Figura 3.2 si vede che il valore di  $\psi$  che massimizza la funzione a posteriori marginale  $\pi_m(\psi|\mathbf{y})$  è  $-0.039$ , mentre il valore di  $\psi$  che massimizza  $\pi_m^{fo}(\psi|\mathbf{y})$  è  $0.058$ . Si osserva, infatti, una traslazione verso destra della  $\pi_m^{fo}(\psi|\mathbf{y})$  rispetto alla distribuzione  $\pi_m(\psi|\mathbf{y})$ . Si ha, pertanto che, nel primo caso, la moda a posteriori per  $\sigma$  è pari a  $0.96$  con un intervallo di credibilità asintotico HPD di livello  $95\%$  dato da  $\{0.728; 1.664\}$ . Nel secondo caso la moda a posteriori di  $\sigma$  è pari a  $1.06$  con intervallo di credibilità asintotico HPD dato da  $\{0.655; 1.408\}$ .

Volendo verificare l'ipotesi nulla  $H_0: \psi = \log\sigma=0$ , ovvero  $H_0: \sigma=1$ , si ottiene

$$EV_*(\psi_0, \mathbf{y}) = 2 \cdot \Phi(r_p^*(\psi_0)) = 0.0001506473 . \quad (3.39)$$

e

$$EV_N(\psi_0, \mathbf{y}) = 1.133507e-05 , \quad (3.40)$$

determinate a partire, rispettivamente, da  $\pi_m(\psi|\mathbf{y})$  e da  $\pi_m^{fo}(\psi|\mathbf{y})$ . Entrambi i risultati mostrano una netta evidenza contro l'ipotesi nulla a favore dell'ipotesi alternativa  $H_1: \log\sigma \neq 0$ , ovvero  $H_1: \sigma \neq 1$ .

Allo scopo di mostrare la validità, anche per piccoli campioni, delle procedure precedentemente illustrate, viene svolto uno studio di simulazione, basato su 5000 repliche, prima per  $n=10$ , poi per  $n=17$  e, infine, per  $n=34$ .

La Tabella 3.1 e la Tabella 3.2 riportano i risultati dello studio di simulazione per  $\psi = \log\sigma$ , per  $\alpha=0.05$  e per  $\alpha=0.1$ , rispettivamente, dove  $\alpha$  indica la soglia della misura di evidenza di Pereira-Stern presa come riferimento.

	$n = 10$	$n = 17$	$n = 34$
$\pi_m(\psi \mathbf{y})$	0.941	0.949	0.951
$\pi_m^{fo}(\psi \mathbf{y})$	0.847	0.895	0.928

Tabella 3.1: Probabilità di corretta classificazione rispettivamente per  $\pi_m(\psi|\mathbf{y})$  e  $\pi_m^{fo}(\psi|\mathbf{y})$  con soglia  $\alpha=0.05$ .

	$n = 10$	$n = 17$	$n = 34$
$\pi_m(\psi \mathbf{y})$	0.897	0.901	0.903
$\pi_m^{fo}(\psi \mathbf{y})$	0.813	0.831	0.859

Tabella 3.2: Probabilità di corretta classificazione rispettivamente per  $\pi_m(\psi|\mathbf{y})$  e  $\pi_m^{fo}(\psi|\mathbf{y})$  con soglia  $\alpha=0.1$ .

Da entrambe le tabelle emerge che all'aumentare della numerosità campionaria, aumenta la probabilità di corretta classificazione, indipendentemente dalla distribuzione marginale. Inoltre, emerge che tale probabilità è più vicina a  $1-\alpha$  per la distribuzione  $\pi_m(\psi|\mathbf{y})$  che per la distribuzione  $\pi_m^{fo}(\psi|\mathbf{y})$ .

Di seguito, si riporta l'analisi effettuata per  $\psi = \beta_2$ .

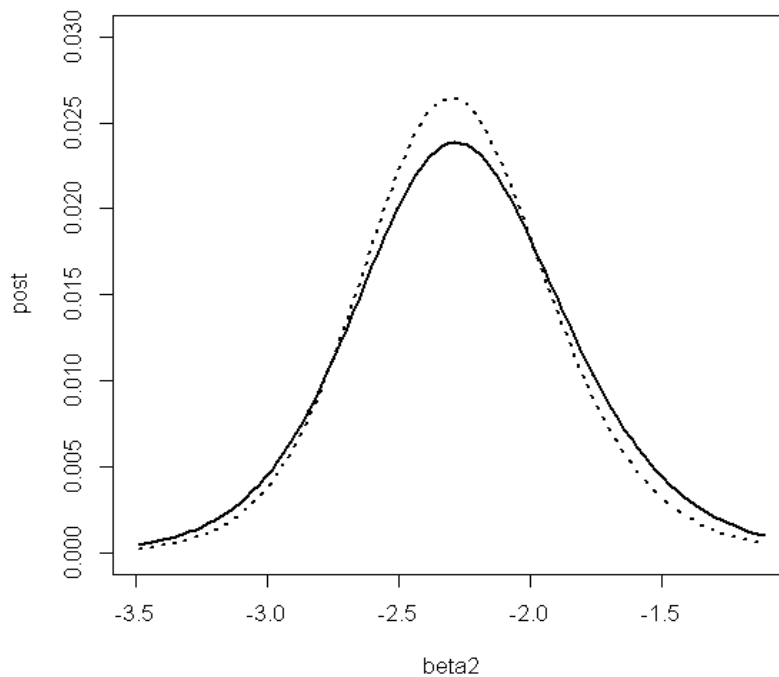


Figura 3.3: Distribuzione a posteriori marginale  $\pi_m(\psi|\mathbf{y})$  (linea continua) e distribuzione a posteriori marginale  $\pi_m^{fo}(\psi|\mathbf{y})$  (linea tratteggiata) per  $\psi = \beta_2$ .

Dalla Figura 3.3 si vede che il valore di  $\psi = \beta_2$  che massimizza la funzione a posteriori marginale  $\pi_m(\psi|\mathbf{y})$  è pari a  $-2.28$  con intervallo di credibilità asintotico HPD di livello 95% dato da  $\{-2.967; -1.635\}$ . Il valore di  $\psi = \beta_2$  che massimizza la funzione a posteriori  $\pi_m^{fo}(\psi|\mathbf{y})$  è pari a  $-2.30$  con intervallo di credibilità asintotico HPD dato da  $\{-3.063; -1.433\}$ .

Volendo verificare l'ipotesi nulla  $H_0: \psi = \beta_2 = 0$ , si ottiene

$$EV_*(\psi_0, \mathbf{y}) = 2 \cdot \Phi(r_p^*(\psi_0)) = 1.868933e-05, \quad (3.41)$$

e

$$EV_N(\psi_0, \mathbf{y}) = 2.868749e-06, \quad (3.42)$$

determinate a partire, rispettivamente, da  $\pi_m(\psi|\mathbf{y})$  e da  $\pi_m^{fo}(\psi|\mathbf{y})$ .

Entrambi i risultati mostrano evidenza contro l'ipotesi nulla a favore dell'ipotesi alternativa: c'è evidenza a favore della presenza di un qualche effetto esercitato dal numero di globuli bianchi, rilevato per ciascun paziente al suo ingresso nello studio, sulla sopravvivenza dei pazienti.

I risultati della simulazione, effettuata per 5000 replicazioni, sono riportati in Tabella 3.3 e in Tabella 3.4 considerando, rispettivamente, le soglie della misura di evidenza  $\alpha=0.05$  e  $\alpha=0.1$ .

	$n = 10$	$n = 17$	$n = 34$
$\pi_m(\psi \mathbf{y})$	0.949	0.951	0.952
$\pi_m^{fo}(\psi \mathbf{y})$	0.867	0.908	0.930

Tabella 3.3: Probabilità di corretta classificazione rispettivamente per  $\pi_m(\psi|\mathbf{y})$  e  $\pi_m^{fo}(\psi|\mathbf{y})$  con soglia  $\alpha=0.05$ .

	$n = 10$	$n = 17$	$n = 34$
$\pi_m(\psi \mathbf{y})$	0.896	0.898	0.903
$\pi_m^{fo}(\psi \mathbf{y})$	0.818	0.847	0.878

Tabella 3.4: Probabilità di corretta classificazione rispettivamente per  $\pi_m(\psi|\mathbf{y})$  e  $\pi_m^{fo}(\psi|\mathbf{y})$  con soglia  $\alpha=0.1$ .

Anche in questo caso, i risultati dello studio di simulazione confermano che la probabilità di corretta classificazione, è maggiore per l'approssimazione della distribuzione marginale  $\pi_m(\psi|\mathbf{y})$ , rispetto a quella ottenuta per la distribuzione marginale  $\pi_m^{fo}(\psi|\mathbf{y})$ , e che aumenta all'aumentare della numerosità campionaria. Questa conclusione vale indipendentemente dal valore assunto da  $n$  e dal valore della soglia  $\alpha$  fissato.

## CONCLUSIONI

Scopo di questa tesi era presentare e discutere l'accuratezza delle approssimazioni asintotiche di ordine elevato per il calcolo della *misura di evidenza di Pereira-Stern*, proposta come alternativa alla procedura di verifica d'ipotesi di Jeffreys, per studiare l'effetto di un singolo regressore sulla variabile risposta. La procedura proposta è stata illustrata nel caso particolare di una *matching prior*, che garantisce l'equivalenza tra un  $p$ -value frequentista e la probabilità individuata dalla *tail area*.

A questo scopo è stato svolto uno studio di simulazione per un modello di Weibull con dati di sopravvivenza. L'utilizzo di tali classi di distribuzioni a priori e l'uso delle corrispondenti approssimazioni asintotiche, hanno permesso di ricorrere al pacchetto `margin` della libreria `hoa`.

La verifica di nullità dei singoli coefficienti di regressione ha mostrato evidenza a favore dell'ipotesi alternativa, confermando il loro effetto sulla sopravvivenza.

È stato poi effettuato uno studio di simulazione con 5000 replicazioni Monte Carlo, per diverse numerosità campionarie, fissando come soglie della misura di Pereira-Stern  $\alpha=0.05$  e  $\alpha=0.1$ . È risultato che la probabilità di corretta classificazione è maggiore per l'approssimazione della distribuzione marginale  $\pi_m(\psi|\mathbf{y})$ , rispetto a quella ottenuta per la distribuzione marginale  $\pi_m^{fo}(\psi|\mathbf{y})$ , e che aumenta all'aumentare della numerosità campionaria.





## BIBLIOGRAFIA

- Barndorff – Nielsen, O. E., Chamberlin, S. R. (1994). Stable and invariant adjusted directed likelihoods. *Biometrika*, **81**, 485-499.
- Borra, S., Di Ciaccio, A. (2004). *Statistica: metodologie per le scienze economiche e sociali*. McGraw-Hill, Milano.
- Brazzale, A. R., Davison, A. C., Reid, N. (2007). *Applied Asymptotics*. Cambridge University Press, Cambridge.
- Brazzale, A.R., (2005). Hoa: an R package bundle for higher order likelihood inference. *R News*, **5**, 20-27.
- Breslow, N. E., Day, N. E. (1980). *Statistical Methods in Cancer Research. Vol.1: The Analysis of Case-control Studies*. International Agency for Research on Cancer, London.
- Cifarelli, D. M., Muliere, P. (1989). *Statistica bayesiana: appunti ad uso degli studenti*. Gianni Iuculano Editore, Pavia.
- Cipollini, F., Stefanini, F. M. (1995). *Appunti di inferenza statistica bayesiana*. [www.ds.unifi.it](http://www.ds.unifi.it).
- Cox, D.R. (1972). Regression models and life-tables (with discussion). *Journal of the Royal Statistical Society B*, **34**, 187-220.
- Cox, D.R. (1975). Partial likelihood. *Biometrika*, **62**, 269-276.
- Cox, D.R., Snell, E.J. (1981). *Applied statistics: Principles and examples*. Chapman & Hall, Londra.
- Cutler, S.J., Ederer, F. (1958). Maximum utilization of the life table method in analyzing survival. *Journal of Chronic Disease*, **8**, 699-713.
- Datta, G. S., Mukerjee, R. (2004). *Probability Matching Priors: Higher Order Asymptotics*. Springer-Verlag, New York.

- Di Orio, F. (1994). *Elementi di Metodologia Epidemiologica Clinica*. Piccin Nuova Libreria S.p.A., Padova.
- Glasser, M. (1967). Exponential survival with covariance. *Journal of the American Statistical Association*, **62**, 561-568.
- Ibrahim J.G., Chen M.-H., Sinha D. (2001). *Bayesian Survival Analysis*. Springer-Verlag, New York.
- Kaplan, E. L., Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, **53**, 457-481.
- Kass, E. R., Raftery, A. E. (1995). Bayes Factor. *Journal of the American Statistical Association*, **90**, 773-795.
- Kimball, A. W. (1960). Estimation of mortality intensities in animal experiments. *Biometrics*, **16**, 505-521.
- Liseo, B. (2008). *Introduzione alla statistica bayesiana*. Roma.  
[www.geostasto.eco.uniroma1.it/utenti/liseo/didattica/bayes.pdf](http://www.geostasto.eco.uniroma1.it/utenti/liseo/didattica/bayes.pdf)
- Madrugá, M. R., Esteves, L. G., Wechsler, S. (2001). On the Bayesianity of Pereira-Stern tests. *Test*, **10**, 1-9.
- Mantel, N., Haenszel, W. (1959). Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of the National Cancer Institute*, **22**, 719-748.
- Marubini, E., Valsecchi, M. G. (1995). *Analysing Survival Data from Clinical Trials and Observational Studies*. John Wiley & Sons, Chichester, England.
- Pace, L., Salvan, A. (2001). *Introduzione alla Statistica II. Inferenza, verosimiglianza, modelli*. Cedam, Padova.
- Pace, L., Salvan, A. (1996). *Teoria della statistica. Metodi, modelli, approssimazioni asintotiche*. Cedam, Padova.

- Parmar, M. K. B., Machin, D. (1995). *Survival analysis: A Practical Approach*. John Wiley & Sons, Chichester, England.
- Pereira, C. A., Stern, J. M. (1999). Evidence and Credibility: Full Bayesian Significance Test for Precise Hypothesis. *Entropy*, **1**, 99-110.
- Piccinato, L. (1996). *Metodi per le decisioni statistiche*. Springer-Verlag Italia, Milano.
- Piccinato, L. (1999). Il concetto statistico di evidenza. *Pubblicazioni del Dipartimento di Statistica, Probabilità e Statistiche applicate, Università degli studi "La Sapienza"*, Roma.
- Reid, N. (2003). The 2000 Wald memorial lectures: Asymptotics and the theory of inference. *Ann. Statist.*, **31**, 1695–1731.
- Robert, C. P. (1993). A note on Jeffreys-Lindley paradox. *Statistica Sinica*, **3**, 601-608.
- Therneau, T. M., Grambsch, P. M. (2000). *Modeling Survival Data: Extending the Cox model*. Springer-Verlag, New York.
- Tripepi, G., Catalano, F. (2004). L'analisi di sopravvivenza con il metodo di Kaplan Meier. *Giornale italiano di nefrologia*, **6**, 540-546.
- Ventura, L., Cabras, S. , Racugno, W. (2009). Prior distributions from pseudo-likelihoods in the presence of nuisance parameters. *J. Amer. Statist. Assoc.* **104**, 768–774.
- Ventura, L., Racugno, W. (2011). Recent advances on Bayesian inference for  $P(X < Y)$ . *Bayesian Analysis*, **6**, 1-18.
- Ventura, L., Sartori, N., Racugno, W. (2011). Objective Bayesian higher-order asymptotics in models with nuisance parameters. *Working Paper Series*, **11**, Dipartimento di Scienze Statistiche, Università degli Studi di Padova.



## **RINGRAZIAMENTI**

Ringrazio in primo luogo la mia Relatrice, la Prof.ssa Laura Ventura, per la disponibilità, la pazienza e il tempo dedicatomi.

Ringrazio i miei genitori per esserci sempre, per sostenermi in ogni momento e per aver sempre fiducia in me e nelle mie decisioni.

Ringrazio le mie sorelle Paola, Michela e Giulia, le migliori consulenti, confidenti e sorelle che possa avere.