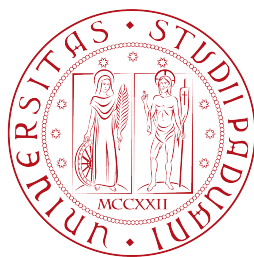


Università degli studi di Padova
Dipartimento di Scienze Statistiche
Corso di Laurea Triennale in
Statistica, Gestione delle Imprese



RELAZIONE FINALE

IL METODO ANALITICO PER LE DISTRIBUZIONI MARGINALI A
POSTERIORI BAYESIANE

Relatore Prof. Nicola Sartori
Dipartimento di Scienze Statistiche

Laureando: Davide Piatteletti
Matricola N. 1011255

Anno Accademico 2013/2014

Indice

Introduzione	5
1 Statistica bayesiana	7
1.1 Impostazione classica e bayesiana	7
1.2 Il modello statistico parametrico	8
1.3 Teorema di Bayes e distribuzione a posteriori	9
1.4 La distribuzione a priori	10
1.4.1 Distribuzioni coniugate	11
1.4.2 Distribuzioni non informative	11
1.5 Inferenza bayesiana	12
1.5.1 Stima puntuale	12
1.5.2 Stima intervallare	12
1.5.3 Verifica d'ipotesi	13
1.6 Simulazione dalla distribuzione a posteriori	15
2 Il metodo analitico	17
2.1 Passi dell'approssimazione	17
2.2 Metodo analitico e metodo HOTA	20
3 Accuratezza del metodo analitico: esempi	21
3.1 Esempio Beta-Binomiale	21
3.2 Esempio NormaleGamma-Normale	27
4 Applicazione del metodo analitico: stima di $P(X < Y)$	35
4.1 La quantità $P(X < Y)$	35
4.2 Variabili X e Y esponenziali	36
4.3 Variabili X e Y Weibull	40
Conclusioni	45
A Codici R per il capitolo 3	47

4

INDICE

B Codici R per il capitolo 4

63

Introduzione

Nell'approccio bayesiano alla statistica, l'inferenza su quantità ignote ruota attorno alla loro distribuzione a posteriori. Tranne alcuni semplici modelli accademici, nello studio comune, non sono sempre presenti modelli di cui si può reperire la distribuzione a posteriori in forma esatta.

Negli anni sono stati sviluppati diversi metodi per approssimare le distribuzioni quando non sono disponibili in forma esatta, ciascuno con i propri pregi e difetti.

In questa relazione verrà illustrato il confronto tra il metodo analitico, basato su approssimazioni asintotiche, e il metodo HOTA, che unisce alle approssimazioni asintotiche un passo di simulazione. Si vuole, infatti, stabilire se non applicando la simulazione si riescono ad ottenere gli stessi risultati ma con un'accuratezza superiore, in quanto si dovrebbe eliminare l'errore dovuto alla medesima simulazione.

Per meglio contestualizzare l'ambito di applicazione dei due metodi, nel primo capitolo vengono presentati i concetti fondamentali della statistica bayesiana, con particolare attenzione a quelli direttamente collegati a questi metodi.

Nel secondo capitolo si spiegherà la differenza teorica tra il metodo analitico e il metodo HOTA, spiegando prima lo schema di simulazione e poi lo schema del metodo analitico.

Nel terzo capitolo si valuterà l'accuratezza dell'approssimazione di tali metodi applicandoli a due esempi in cui si conoscono le distribuzioni a posteriori esatte.

Infine, nel quarto capitolo, si applicheranno questi due metodi in un'applicazione reale in cui non si riesce ad ottenere una forma chiusa per la distribuzione a posteriori marginale e l'unica soluzione risulta quello di approssimarle per poter fare inferenza.

Capitolo 1

Statistica bayesiana

In questo capitolo verranno presentati alcuni concetti di base inerenti alla statistica bayesiana. Per maggiori approfondimenti si rimanda, ad esempio, a Liseo (2008).

1.1 Impostazione classica e bayesiana

Scopo primario dell'inferenza statistica, almeno nella tradizione classica, è quello di acquisire ulteriore conoscenza su quantità incerte, sulle quali spesso si hanno informazioni parziali, non sufficienti ad eliminare del tutto l'incertezza: questo processo induttivo è spesso la necessaria premessa a un processo decisionale, dove le informazioni in precedenza acquisite, adeguatamente filtrate, vengono utilizzate per scegliere quale decisione prendere, fra diverse possibili.

Ciò che non è stato considerato finora sono le informazioni a priori che avevamo sui nostri parametri d'interesse. Lo si può fare attraverso il teorema di Bayes. La statistica bayesiana ha come obiettivo il medesimo della statistica classica, tuttavia introduce all'interno del modello statistico la distribuzione di probabilità *iniziale*. Questo e altro sulle informazioni a priori verrà discusso nel paragrafo dedicato.

Da notare il cambiamento di status dei nostri parametri d'interesse che, mentre nell'impostazione classica sono considerati delle quantità fissate ma incognite, essi diventano, nell'impostazione bayesiana, delle variabili aleatorie la cui distribuzione iniziale dipende dalle informazioni in nostro possesso, in quel dato contesto. Tale distribuzione è forzatamente soggettiva ovvero varia da individuo a individuo perchè rappresenta la sintesi delle informazioni che il singolo individuo possiede sul problema specifico.

La prima conclusione che si trae è che le informazioni a priori hanno un ruolo importante nell'inferenza e che tali informazioni introducono nell'analisi una componente soggettiva. Questo aspetto è tuttora fonte di dibattito nell'ambiente scientifico. I sostenitori dell'impostazione bayesiana affermano che il fare uso di informazioni particolari, contingenti, è il modo in cui ogni essere razionale opera in situazioni d'incertezza: del resto persone diverse possono benissimo prendere decisioni differenti anche sulla base di informazioni condivise uguali. Nell'impostazione bayesiana il risultato dell'esperimento statistico viene spesso inteso come non replicabile esattamente e quindi ci si condiziona ad esso. Chi invece critica l'approccio bayesiano sostiene che la statistica, per conservare dignità scientifica e per poter essere usata proficuamente nella pratica, deve garantire una oggettività delle conclusioni che si possono trarre da un esperimento e perciò queste non possono dipendere dalle informazioni di chi conduce l'esperimento. Chi sostiene l'approccio classico, infatti, sostiene il metodo frequentista cioè la possibilità di replicare infinite volte lo stesso esperimento statistico nelle medesime condizioni; l'inferenza si basa quindi sul confronto del risultato osservato con i potenziali risultati ottenibili in queste ipotetiche replicazioni.

Nel seguito del capitolo, si introdurranno i vari elementi formali dell'inferenza bayesiana.

1.2 Il modello statistico parametrico

Sia $y = (y_1, \dots, y_n)$ realizzazione della variabile casuale Y con legge di probabilità $p(y; \theta)$ e sia θ il parametro di interesse. Si definisce modello statistico parametrico e si indica con il simbolo Λ , la terna $\Lambda = (\mathcal{Y}, \mathcal{P}, \Theta)$, dove:

- lo spazio campionario viene indicato con \mathcal{Y} e rappresenta l'insieme di tutti i possibili risultati osservabili in un esperimento,
- con \mathcal{P} si indica la famiglia di tutte le possibili leggi di probabilità $p(y; \theta)$ che si possono associare alla variabile casuale Y ,
- lo spazio parametrico per θ viene indicato con Θ e vale $\Theta \subseteq \mathbb{R}^P$.

Scelta la famiglia di appartenenza della legge $p(y; \theta)$, il fare inferenza su θ sulla base delle osservazioni y significa identificare con maggiore precisione possibile quale sia il vero meccanismo generatore dei dati, all'interno del modello Λ .

1.3 Teorema di Bayes e distribuzione a posteriori

Strumento fondamentale per unire le informazioni a priori sulla quantità di interesse e quelle derivanti dai dati in nostro possesso è il teorema di Bayes. Esso rappresenta il modo in cui gli esseri razionali aggiornano le proprie conoscenze sulla base di nuove informazioni disponibili.

Teorema 1 (di Bayes). *Sia E un evento contenuto in $F_1 \cup \dots \cup F_k$, dove gli eventi F_j , $j=1, \dots, k$ sono a due a due incompatibili e necessari. Allora per ognuno dei suddetti eventi F_j vale la seguente formula.*

$$P(F_j|E) = \frac{P(F_j)P(E|F_j)}{\sum_{j=1}^k P(F_j)P(E|F_j)}. \quad (1.1)$$

Tale teorema mette in luce come la probabilità di un evento F_j dipenda anche dalle informazioni di cui si è in possesso, in questo caso rappresentate dall'evento E , che sarà il risultato dell'esperimento statistico. Inoltre tale teorema permette l'inversione di probabilità, ossia valutare le probabilità condizionate di varie ipotesi a partire dalle probabilità condizionate delle osservazioni.

Nella formula precedente il membro di sinistra prende il nome di probabilità finale (o a posteriori) dell'evento F_j : il termine finale sta a significare dopo che è noto che si è verificato l'evento E . Il denominatore del membro di destra è un semplice fattore di normalizzazione; nel numeratore, invece, compaiono due quantità: $P(F_j)$ è la probabilità a priori dell'evento F_j e $P(E|F_j)$ rappresenta invece la verosimiglianza di F_j . Perciò l'evento F_j sarà tanto più plausibile (verosimile) sulla base delle informazioni in nostro possesso tanto maggiore sarà la probabilità che si verifichi l'evento E condizionatamente al fatto che si sia verificato F_j . Nel contesto di un modello statistico parametrico la formula di Bayes permette di giungere alla distribuzione a posteriori $\pi(\theta|y)$ per il parametro di interesse θ . Indicato con $\pi(\cdot)$ la distribuzione a priori di θ si ottiene che la distribuzione a posteriori ha la seguente espressione generale

$$\pi(\theta|y) = \frac{\pi(\theta)p(y; \theta)}{\int_{\Theta} \pi(\theta)p(y; \theta)d\theta}. \quad (1.2)$$

Generalmente, per y fissato, si indica con $L(\theta; y) \propto p(y; \theta)$ la verosimiglianza di θ . La formula (1.2) rende evidente che in ottica bayesiana la quantità di

interesse non è più fissata come nell'impostazione classica ma è una variabile aleatoria la cui distribuzione di probabilità è influenzata sia dalle informazioni a priori sia dai dati a disposizione. La distribuzione a posteriori $\pi(\theta|y)$, contenendo tutta l'informazione riguardante il parametro θ , sarà utilizzata per produrre indicatori sintetici, per la determinazione di stime puntuali o intervallari, e per la verifica d'ipotesi, reperendo da essa, se necessario, le eventuali distribuzioni marginali delle singole componenti di θ .

1.4 La distribuzione a priori

La scelta della distribuzione iniziale per i parametri presenti nel modello statistico è l'aspetto cruciale dell'impostazione bayesiana. Di fatto, la distribuzione iniziale è il vettore attraverso cui le informazioni extra-sperimentali vengono inserite nel procedimento induttivo, e l'adozione di una distribuzione rende l'analisi statistica, almeno sul piano formale, soggettiva. L'uso di distribuzioni iniziali comporta inoltre problemi di tipo computazionale. A meno di non utilizzare forme analitiche particolari, le cosiddette distribuzioni coniugate al modello statistico utilizzato, risulta spesso difficile ottenere espressioni analitiche in forma esplicita per le distribuzioni finali. Attualmente convivono due diverse filosofie all'interno dell'impostazione bayesiana: quella soggettiva e quella oggettiva.

La filosofia soggettiva implica la scelta della distribuzione a priori che meglio rappresenta le proprie convinzioni iniziali: soggetti diversi potranno specificare sia le stesse distribuzioni di probabilità ma con valori dei parametri che la caratterizzano differenti, sia distribuzioni di probabilità completamente diverse. Da ciò deriva una forte componente soggettiva. Inoltre, nei risultati vi sarà una sorta di mancanza di precisione dovuta al fatto che lo specificare una distribuzione a priori pone dei vincoli e non permette quindi di poter esprimere in modo preciso e completo le proprie conoscenze.

La specificazione di una distribuzione a priori secondo un approccio oggettivo parte dalla debolezza applicativa del metodo soggettivo, pur riconoscendone la validità teorica, propone la determinazione e l'adozione di distribuzioni a priori di tipo convenzionale, derivabili sulla base delle proprietà matematiche del modello statistico utilizzato. Così, da un lato si cerca sia di conservare la pretesa oggettività delle procedure inferenziali che dipendono in questo modo esclusivamente dal modello statistico prescelto e dal risultato campionario, dall'altro, l'uso di una legge iniziale consente ancora di fare uso del teorema di Bayes per produrre conclusioni inferenziali nel linguaggio probabilistico.

Nel seguito verranno presentate alcune classi di distribuzioni a priori e verranno illustrati brevemente i metodi per l'individuazione di una distribuzione a priori non informativa.

1.4.1 Distribuzioni coniugate

Sia $Y = (Y_1, \dots, Y_n)$ un vettore aleatorio a componenti somiglianti e indipendenti condizionatamente a un vettore di parametri $\theta \in \Theta$, con funzione di verosimiglianza

$$L(\theta; y) \propto \prod_{j=1}^n p(y_j; \theta). \quad (1.3)$$

Una distribuzione di probabilità iniziale $\pi(\theta)$ si dice *coniugata* al modello utilizzato se la forma funzionale della distribuzione iniziale e della distribuzione finale sono uguali. Il risultato sperimentale entrerà in tale meccanismo andando ad aggiornare gli iper-parametri della distribuzione a priori individuando così i parametri della distribuzione a posteriori. Tali parametri conterranno anche l'informazione disponibile a priori. In generale, l'esistenza di una distribuzione a priori coniugata è garantita quando $p(y; \theta)$ appartiene alla famiglia esponenziale.

1.4.2 Distribuzioni non informative

Le distribuzioni non informative nascono con l'obiettivo di ottenere delle conclusioni oggettive, pur operando in ambito bayesiano. Segue che tali distribuzioni vengano individuate sulla base del modello statistico scelto per l'esperimento e non sulla base delle informazioni a priori di cui dispone chi conduce l'esperimento. Tali distribuzioni saranno spesso delle distribuzioni improprie, cioè che non rispettano tutte le caratteristiche di una legge di probabilità. Ciò non costituisce un problema fintanto che la distribuzione a posteriori risulta essere una distribuzione propria.

La famiglia delle distribuzioni non informative include, ad esempio, le distribuzioni a priori di Jeffreys, le distribuzioni individuate col metodo delle *reference prior*. La distribuzione non informativa di Jeffreys è

$$\pi^J(\theta) \propto \sqrt{\det(I(\theta))} \quad (1.4)$$

dove $I(\theta)$ rappresenta la matrice d'informazione attesa di Fisher relativa ad una singola osservazione. Tali distribuzioni sono invarianti rispetto a trasformazioni sui singoli parametri di interesse.

Il metodo delle *reference prior* si pone come il metodo di *default* per l'individuazione di una distribuzione a priori non informativa per un modello statistico. Tale metodo consiste nel massimizzare la distanza di Kullback-Leibler tra la distribuzione a priori e quella a posteriori indotta da essa rispetto alla distribuzione a priori. Si otterrà quindi la distribuzione a priori con minor apporto informativo, cioè quella che massimizza il valore atteso della distanza di Kullback-Leibler.

1.5 Inferenza bayesiana

L'approccio bayesiano ci permette di giungere ad una distribuzione finale sintesi della distribuzione a priori arricchita dalla funzione di verosimiglianza dei dati. Come nella statistica classica, anche nell'ambito bayesiano, si riesce a giungere, dalla distribuzione a posteriori, alle stime puntuali, intervallari e alle verifiche d'ipotesi.

1.5.1 Stima puntuale

In ambito bayesiano, una volta determinata la distribuzione a posteriori per θ , la sintesi più naturale è fornita da un qualche indicatore sintetico di posizione. La stima puntuale permette di trovare un valore $\tilde{\theta} = \tilde{\theta}(y) \in \Theta$ che individui quale legge all'interno di \mathcal{P} abbia agito e generato il risultato sperimentale $y \in \mathcal{Y}$. Per semplicità di calcolo e per il suo ruolo anche al di fuori della statistica, il valore atteso $E(\theta|y)$ è il più usato. A volte il calcolo del valore atteso è complicato, altre volte non esiste. In questi casi è auspicabile ricorrere a stime puntuali alternative come la mediana, la moda o altri quantili della distribuzione a posteriori.

1.5.2 Stima intervallare

La stima intervallare permette di individuare un insieme di valori $\tilde{\Theta}_{1-\alpha} = \tilde{\Theta}_{1-\alpha} \subseteq \Theta$ per il parametro di interesse θ , in modo tale che $P(\theta \in \tilde{\Theta}_{1-\alpha}|y) = 1-\alpha$, dove $1-\alpha$ è il livello specificato di credibilità. La regione può essere costruita sfruttando i quantili della distribuzione a posteriori di θ oppure inserendo in esso i valori di θ più plausibili. Nel primo modo, gli estremi dell'intervallo verranno individuati prendendo il quantile, o i quantili, tali per cui la probabilità della regione così individuata sia pari a $1-\alpha$. possibili intervalli potrebbero essere:

- $\{\theta \in \Theta : \tilde{\theta}_{\alpha/2} \leq \theta \leq \tilde{\theta}_{1-\alpha/2}\}$

- $\{\theta \in \Theta : \theta \leq \tilde{\theta}_{1-\alpha}\}$
- $\{\theta \in \Theta : \theta \geq \tilde{\theta}_\alpha\}$

dove $\tilde{\theta}_\alpha$ rappresenta il quantile tale per cui $P(\theta \leq \tilde{\theta}_\alpha | y) = \alpha$.

Nel secondo modo, l'insieme di $\tilde{\Theta}_{1-\alpha}$ verrà individuato inserendo in esso i valori θ ritenuti più plausibili, cioè quelli che presentano densità a posteriori più elevata, finché non si raggiungerà una probabilità complessiva non inferiore al livello $1 - \alpha$ stabilito. Tale intervallo è anche detto HPD (*High Posterior Density*). La determinazione di un insieme HPD non è sempre possibile per via analitica; può accadere che la distribuzione finale risulti multimodale cosicché l'insieme HPD potrebbe essere addirittura formato da intervalli disgiunti di valori. Nel caso piuttosto generale in cui la distribuzione finale $\pi(\theta|y)$ sia unimodale, occorre ricorrere a metodi numerici. A grandi linee la procedura consiste in tre passi:

1. Per k fissato, minimo del massimo valore di $\pi(\theta|y)$, determinare le due radici $\tilde{\theta}_1(k)$ e $\tilde{\theta}_2(k)$ dell'equazione $\pi(\theta|y) = k$.
2. Si calcola $P(k) = \Pr(\theta \in (\tilde{\theta}_1(k), \tilde{\theta}_2(k)) | y)$.
3. Si determina k^* che risolve l'equazione $P(k) = 1 - \alpha$.

L'intervallo HPD sarà quindi

$$(\tilde{\theta}_1(k^*), \tilde{\theta}_2(k^*)).$$

1.5.3 Verifica d'ipotesi

In un problema di verifica delle ipotesi, si confrontano due ipotesi alternative:

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

e

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

con $\Theta_0 \cup \Theta_1 = \Theta$ e $\Theta_0 \cap \Theta_1 = \emptyset$, generalmente H_0 viene chiamata ipotesi nulla, invece, H_1 viene chiamata ipotesi alternativa.

L'approccio bayesiano al problema di verifica d'ipotesi è molto diverso rispetto a quello classico per via della presenza di leggi di probabilità per i parametri di interesse. Il confronto tra due ipotesi alternative infatti avviene calcolando la probabilità $P(H_i|y)$, $i=0,1$:

$$P(H_i|y) = P(\Theta_i|y) = \int_{\Theta_i} \pi(\theta|y) d\theta.$$

Ipotesi alternativa semplice

I ruoli svolti dalle informazioni a priori e dalla funzione di verosimiglianza nella verifica di ipotesi bayesiana sono più chiaramente espressi nel caso artificialmente semplice di due ipotesi puntuali $H_0: \theta = \theta_0$ e $H_1: \theta = \theta_1$.

Per verificare quale sia più plausibile sulla base delle informazioni in nostro possesso si può utilizzare il rapporto:

$$\frac{P(H_1|y)}{P(H_0|y)} = \frac{\pi(\theta_1|y)}{\pi(\theta_0|y)} = \frac{\pi(\theta_1)L(\theta_1; y)}{\pi(\theta_0)L(\theta_0; y)} = \frac{1 - \pi_0}{\pi_0} \frac{L(\theta_1; y)}{L(\theta_0; y)} \quad (1.5)$$

Data la natura delle ipotesi è facile notare come $\pi_1 = P(H_1) = 1 - P(H_0) = 1 - \pi_0$. Il rapporto (1.5) è il prodotto di due quantità: $(1 - \pi_0)/\pi_0$ e $L(\theta_1; y)/L(\theta_0; y)$. La prima quantità rappresenta il peso relativo delle due ipotesi prima di osservare i dati, la seconda quantità rappresenta invece il rapporto di verosimiglianza.

Definizione 1. Si definisce *fattore di Bayes per le ipotesi alternative* H_0 e H_1 e si indica con B_{10} il rapporto:

$$B_{10} = \frac{Pr(H_1|y) Pr(H_0)}{Pr(H_0|y) Pr(H_1)} \quad (1.6)$$

Esso rappresenta il fattore moltiplicativo che trasforma il rapporto a priori in quello a posteriori. Nel caso di due ipotesi entrambe semplici il fattore di Bayes coincide con il rapporto di verosimiglianza e non contiene alcuna componente soggettiva a parte la scelta del modello statistico. Il fatto che il rapporto (1.5) sia maggiore di 1 farà propendere verso il rifiuto dell'ipotesi nulla, viceversa se il rapporto sarà minore di 1 si accetterà l'ipotesi nulla.

Ipotesi alternativa composta

Consideriamo adesso un caso più generale in cui l'ipotesi nulla puntuale $H_0: \theta = \theta_0$ viene posta a confronto con un'alternativa composta $H_1: \theta \neq \theta_0$. Poichè H_0 è un'ipotesi semplice, la probabilità θ_0 risulterà concentrata sul punto θ_0 ; di contro, la probabilità H_1 deve essere distribuita su tutti i valori di θ diversi da θ_0 .

Il rapporto (1.5) diventa:

$$\frac{P(H_1|y)}{P(H_0|y)} = \frac{1 - \pi_0}{\pi_0} \frac{\int_{\theta \neq \theta_0} L(\theta; y) g(\theta) d\theta}{L(\theta_0; y)} = \frac{m_1(y)}{m_0(y)} \quad (1.7)$$

dove $g(\cdot)$ è la distribuzione a priori per il parametro θ sotto H_1 . La distribuzione a priori per θ viene specificata utilizzando la funzione delta di Dirac nel seguente modo:

$$\pi(\theta) = \pi_0 \delta_{\theta_0}(\theta) + (1 - \pi_0) g(\theta) I_{\theta=\theta_0}(\theta) \quad (1.8)$$

Dalla (1.7) è facile trovare il fattore di Bayes. Indicando con $m_i(y)$, $i=0,1$ le distribuzioni marginali del vettore y sotto l'ipotesi nulla ($i=0$) e alternativa ($i=1$) si ottiene che:

$$B_{10} = \frac{\frac{P(H_1|y)}{P(H_0|y)}}{\frac{1-\pi_0}{\pi_0}} = \frac{\int_{\theta \neq \theta_0} L(\theta; y) g(\theta) d\theta}{L(\theta_0; y)} = \frac{m_1(y)}{m_0(y)}. \quad (1.9)$$

Analogamente per quanto detto con l'ipotesi semplice, il fatto che il fattore di Bayes sia maggiore o minore di 1 farà propendere verso il rifiuto o l'accettazione dell'ipotesi nulla.

1.6 Simulazione dalla distribuzione a posteriori

Generalmente, le distribuzioni a posteriori, tranne nel caso in cui la distribuzione a priori sia coniugata con il modello statistico, non risulta essere sempre scritta in maniera esplicita. Questo comporta difficoltà nel fare inferenza all'atto pratico, essendo necessario risolvere numerosi integrali quando si ha a che fare con distribuzioni di probabilità.

Negli anni sono stati proposti vari metodi per studiare la distribuzione a posteriori nel caso in cui essa non esista in forma chiusa. Inizialmente la strada seguita per conoscere la distribuzione a posteriori era quello di approssimarla analiticamente ricorrendo soprattutto al teorema del limite centrale o simili, al fine di ricondursi ad una distribuzione nota. Con l'avvento dei personal computer, disponendo di maggiori potenzialità di calcolo, si sono sviluppati metodi di calcolo numerico ed in particolare di integrazione numerica che permettono di giungere a misure di sintesi per la distribuzione a posteriori, quando la dimensione del parametro non è eccessiva. Si sono sviluppati inoltre metodi basati sulle tecniche Monte Carlo, come il campionamento accettazione/rifiuto, il campionamento per importanza e i metodi Monte Carlo basati sulle catene di Markov.

Il generare valori casuali dalla distribuzione a posteriori è senza dubbio il metodo maggiormente utilizzato per poterne calcolare misure di sintesi e poter fare inferenza.

Capitolo 2

Il metodo analitico

In questo capitolo si vuole confrontare il metodo HOTA (*Higher Order Tail area Approximation*), che è un metodo di simulazione dalla distribuzione a posteriori ottenuto combinando il metodo di simulazione per inversione e alcune approssimazioni asintotiche di ordine elevato della funzione di ripartizione a posteriori, e il metodo analitico, che si basa sempre sull'approssimazione analitica ma non utilizza la simulazione. Lo scopo è verificare se procedendo per via analitica l'errore è inferiore rispetto a quello ottenuto con il metodo HOTA, in cui all'errore dovuto alle approssimazioni si somma l'errore dovuto alla simulazione.

Gli argomenti trattati di seguito possono essere approfonditi in Ruli, Sartori e Ventura (2014).

2.1 Passi dell'approssimazione

Consideriamo un modello statistico parametrico con funzione di densità $p(y; \theta)$, con $\theta \in \Theta \subseteq \mathbb{R}^p$, con $p \geq 1$. Sia $\theta = (\psi, \lambda)$, con ψ parametro di interesse scalare e λ parametro di disturbo $(p - 1)$ -dimensionale. Denotiamo inoltre con $l(\theta) = l(\psi, \lambda) = l(\psi, \lambda; y)$ la log-verosimiglianza basata sui dati $y = (y_1, \dots, y_n)$ e con $\pi(\theta)$ la distribuzione a priori di (ψ, λ) . L'oggetto di interesse risulta essere la distribuzione a posteriori $\pi(\theta|y)$. Da essa infatti si può reperire la distribuzione marginale per il parametro di interesse ψ :

$$\pi(\psi|y) = \int \pi(\psi, \lambda|y) d\lambda. \quad (2.1)$$

Indichiamo inoltre con $\tilde{l}(\theta) = l(\theta) + \log \pi(\theta)$ la log-distribuzione a posteriori non normalizzata, con $\tilde{\theta} = (\tilde{\psi}, \tilde{\lambda})$ la moda a posteriori e con $\tilde{\lambda}_\psi$ la moda

a posteriori di λ per ψ fissato. Requisito fondamentale affinché valgano le approssimazioni illustrate successivamente è che la moda esista e sia unica e che la matrice Hessiana di $\tilde{l}(\theta)$ valutata in $\tilde{\theta}$ e in $(\psi, \tilde{\lambda}_\psi)$, sia definita negativa, per ogni ψ .

La distribuzione marginale a posteriori (2.1) può essere approssimata (Tierney and Kadane 1986) tramite la formula di Laplace, con un errore relativo di ordine $O_p(n^{-3/2})$

$$\pi(\psi|y) \doteq \frac{1}{\sqrt{2\pi}} \exp\{\tilde{l}(\psi, \tilde{\lambda}_\psi) - \tilde{l}(\tilde{\psi}, \tilde{\lambda})\} \sqrt{\frac{|\tilde{h}(\tilde{\psi}, \tilde{\lambda})|}{|\tilde{h}_{\lambda\lambda}(\psi, \tilde{\lambda}_\psi)|}} \quad (2.2)$$

dove $\tilde{h}(\psi, \lambda) = -\partial^2 \tilde{l}(\theta) / (\partial\theta\partial\theta^T)$ è la matrice Hessiana negativa $p \times p$ di $\tilde{l}(\theta)$ e $\tilde{h}_{\lambda\lambda}(\psi, \lambda)$ è il blocco (λ, λ) di $\tilde{h}(\psi, \lambda)$.

Un'approssimazione del terzo ordine della funzione di ripartizione marginale si può ottenere nel seguente modo. Per approfondimenti si rimanda a Davison (2003, Paragrafo 11.3.1). Partendo da

$$\int_{-\infty}^{\psi_0} \pi(\psi|y) d\psi \doteq \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\psi_0} \exp\{\tilde{l}(\psi, \tilde{\lambda}_\psi) - \tilde{l}(\tilde{\psi}, \tilde{\lambda})\} \sqrt{\frac{|\tilde{h}(\tilde{\psi}, \tilde{\lambda})|}{|\tilde{h}_{\lambda\lambda}(\psi, \tilde{\lambda}_\psi)|}} d\psi. \quad (2.3)$$

si può cambiare la variabile di integrazione da ψ a $\tilde{r}(\psi)$, che ha la seguente espressione

$$\tilde{r}(\psi) = \text{sign}(\tilde{\psi} - \psi) [2(\tilde{l}(\tilde{\psi}, \tilde{\lambda}) - \tilde{l}(\psi, \tilde{\lambda}_\psi))]^{1/2},$$

e ha $-\tilde{l}_\psi(\psi) / \tilde{r}(\psi)$ come Jacobiano dove $\tilde{l}_\psi = \partial \tilde{l}(\psi, \tilde{\lambda}_\psi) / (\partial \psi)$. Di conseguenza si ottiene

$$\int_{-\infty}^{\psi_0} \pi(\psi|y) \doteq \frac{1}{2\pi} \int_{-\infty}^{\tilde{r}(\psi_0)} \exp\left\{-\frac{1}{2}\tilde{r}^2\right\} \frac{\tilde{r}}{\tilde{l}_\psi(\psi)} \sqrt{\frac{|\tilde{h}(\tilde{\psi}, \tilde{\lambda})|}{|\tilde{h}_{\lambda\lambda}(\psi, \tilde{\lambda}_\psi)|}} d\tilde{r}. \quad (2.4)$$

Il passo successivo è un nuovo cambio di variabile da $\tilde{r}(\psi)$ a

$$\tilde{r}^*(\psi) = \tilde{r}(\psi) + \frac{1}{\tilde{r}(\psi)} \log \frac{\tilde{q}_B(\psi)}{\tilde{r}(\psi)}$$

con

$$\tilde{q}_B = \tilde{l}_\psi(\psi) \sqrt{\frac{|\tilde{h}_{\lambda\lambda}(\psi, \tilde{\lambda}_\psi)|}{|\tilde{h}(\tilde{\psi}, \tilde{\lambda})|}}.$$

Dato che lo Jacobiano di quest'ultima trasformazione contribuisce solamente all'errore della (2.4), è possibile dimostrare che l'approssimazione della funzione di ripartizione è data da

$$\int_{-\infty}^{\psi_0} \pi(\psi|y) d\psi \doteq \frac{1}{2\pi} \int_{-\infty}^{\tilde{r}^*(\psi_0)} \exp\left\{-\frac{1}{2}t^2\right\} dt = \Phi\{\tilde{r}^*(\psi)\}, \quad (2.5)$$

dove $\Phi(\cdot)$ è la funzione di ripartizione di una normale standard.

Si può notare come l'espressione (2.5) dipenda solamente da quantità a posteriori di facile calcolo con qualsiasi software in grado di ottimizzare e derivare numericamente funzioni.

Attraverso lo stesso procedimento sopra spiegato, possiamo scrivere la (2.1) come

$$\int_{\psi_0}^{\infty} \pi(\psi|y) d\psi \doteq \Phi\{r_p^*(\psi_0)\}, \quad (2.6)$$

dove

$$r_p^*(\psi) = r_p(\psi) + r_p(\psi)^{-1} \log \frac{q_B(\psi)}{r_p(\psi)}$$

e

$$r_p(\psi) = \text{sgn}(\hat{\psi} - \psi) [2(l_p(\hat{\psi}) - l_p(\psi))]^{1/2}$$

e

$$q_B(\psi) = l'_p(\psi) j_p(\psi)^{-1/2} \sqrt{\frac{|j_{\lambda\lambda}(\psi, \hat{\lambda}_\psi)|}{|j_{\lambda\lambda}(\hat{\psi}, \hat{\lambda})|}} \frac{\pi(\psi, \lambda)}{\pi(\psi, \hat{\lambda}_\psi)}$$

con $l'_p(\psi) = \partial l_p(\psi) / (\partial \psi)$, dove $l_p(\psi) = l(\psi, \hat{\lambda}_\psi)$ è verosimiglianza profilo, e $\hat{\lambda}_\psi$ è la stima di massima verosimiglianza di λ per ψ fissato e $(\hat{\psi}, \hat{\lambda})$ è la stima di massima verosimiglianza di (ψ, λ) .

In genere la (2.5) è più accurata della (2.6).

2.2 Metodo analitico e metodo HOTA

In questo paragrafo si cerca di confrontare il metodo HOTA che si basa sulla simulazione con l'approccio analitico.

Lo schema di simulazione del metodo HOTA si basa sul seguente algoritmo:

1. Si ripetono N volte i seguenti passi
2. Generare casualmente $z_i \sim N(0, 1)$
3. Risolvere $\tilde{r}^*(\psi_i) = z_i$ rispetto a ψ_i
4. Salvare ψ_i come un elemento di un campione approssimato da $\pi(\psi|y)$

Attraverso l'approccio analitico si vuole evitare di introdurre l'errore di simulazione. Ad esempio, se si vuole ottenere $E(X^r) = \int x^r f(x) dx = \int x^r dF(x)$ si può fare un cambio di variabile $u = F(x)$, con $x = F^{-1}(u)$ e $dx = \frac{1}{f(F^{-1}(u))} du$. Quindi, $E(X^r) = \int_0^1 \{F^{-1}(u)\}^r du$.

Per approssimare questo integrale fissiamo una griglia di n valori equispaziati in $[0, 1]$, ad esempio $\frac{k}{n+1}$, $k = 1, \dots, n$ e calcolando la media $\frac{1}{n} \sum_{i=1}^n \{F^{-1}(\frac{k}{n+1})\}^r$.

In particolare, si utilizzerà l'approssimazione asintotica $F(x) \doteq \Phi\{r^*(x)\}$.

Quindi, possiamo calcolare i valori necessari per l'approssimazione attraverso le relazioni

$$\begin{aligned} F(F^{-1}(\frac{k}{n+1})) &= \frac{k}{n+1} \Rightarrow \Phi(r^*(F^{-1}(\frac{k}{n+1}))) \doteq \frac{k}{n+1} \\ &\Rightarrow r^*(F^{-1}(\frac{k}{n+1})) \doteq \Phi^{-1}(\frac{k}{n+1}) = z_{\frac{k}{n+1}}. \end{aligned}$$

dove z_α è il quantile di livello α della normale standard.

Una volta calcolati i momenti si possono calcolare media e varianza che confronteremo con quelle ottenute con il metodo HOTA.

Capitolo 3

Accuratezza del metodo analitico: esempi

Obiettivo di questo capitolo è quello di mostrare l'accuratezza del metodo analitico confrontando una serie di indicatori sintetici della distribuzione a posteriori ottenuti con il metodo analitico con quelli ottenuti dalla distribuzione esatta e con quelli ottenuti con il metodo HOTA. In particolare, si considererà un modello binomiale e un modello normale. Per giungere alla distribuzione a posteriori esatta, nei due esempi si è utilizzata una distribuzione a priori coniugata specificata in modo che sia anche non informativa, invece gli indici sintetici ottenuti con il metodo HOTA si basano sulla tesi di Enrico Gasparin.

3.1 Esempio Beta-Binomiale

Si supponga di voler ottenere una stima della percentuale di pezzi difettosi prodotti da una macchina. Si sa che tale macchina produce quindici pezzi al giorno e si supponga che il numero di pezzi difettosi registrati in cinque giorni sia pari a 2, 1, 3, 1, 1.

Sia θ la proporzione di pezzi difettosi. I dati in nostro possesso possono essere considerati delle realizzazioni indipendenti di una variabile casuale *Binomiale* per cui la distribuzione a priori coniugata è di tipo *Beta*. La distribuzione a posteriori esatta sarà ancora di tipo *Beta* ma con parametri aggiornati. Sia $\theta \sim Beta(\alpha, \beta)$ in cui i parametri α e β sono pari a valori prossimi a zero affinché tale distribuzione sia non informativa. Indicando con $n = 75$ il numero di prove effettuate e con $y = 8$ il numero di pezzi difettosi totali si ottiene che la distribuzione a posteriori esatta è (Liseo 2008):

$$\theta|y \sim \text{Beta}(\alpha + y, \beta + n - y).$$

Per l'implementazione del metodo analitico, al fine di agevolare la risoluzione numerica, si è applicata la trasformazione logit alla proporzione di pezzi difettosi. Così facendo si è potuto lavorare nell'intero campo dei numeri reali senza dover porre vincoli alle quantità coinvolte. Al momento della restituzione dei valori simulati si è tornati alla parametrizzazione originale. Per generare l'output simile a quello riportato nel seguente paragrafo si possono utilizzare questi comandi, il codice è riportato nell'Appendice A:

```
#dati
>dati=cbind(rep(15,5),c(2,1,3,1,1))
#parametri della distribuzione a priori
>a=1e-7
>b=1e-7
#numero punti in cui si suddivide l'intervallo [0,1]
>N=5000
>s=model.beta.binom(N,a,b,dati)
>summary.beta.binom(s)
```

Risultato

Come visibile nelle Figure 3.1 e 3.2, l'approssimazione della distribuzione a posteriori tramite il metodo analitico segue in modo molto soddisfacente la distribuzione esatta sia in termini di densità che in termini di funzione di ripartizione. Tali conclusioni trovano conferma anche nei risultati riportati nelle Tabelle 3.1 e 3.2.

Sempre in Tabella 3.1 è possibile notare come, aumentando il numero di punti della griglia da $N=500$ a $N=5000$ nell'intervallo $[0,1]$ l'approssimazione dei primi due momenti sia vicina a quella della distribuzione esatta. Nel metodo HOTA, N indica che sono stati generati casualmente N valori da una normale standard.

Nella Tabella 3.2 si vuole mostrare invece l'accuratezza del metodo analitico su piccoli campioni. Supponiamo, ad esempio, di considerare solamente i dati del primo giorno.

	Metodo analitico N=500	Metodo analitico N=5000	Valori esatti
Media	0.10661	0.10667	0.10667
Deviazione satandard	0.03504	0.03536	0.03541
I quantile	0.08119	0.08110	0.08108
Mediana	0.10319	0.10319	0.10317
III quantile	0.12837	0.12849	0.12848
Estremo inferiore HPD	0.04292	0.04248	0.04242
Estremo superiore HPD	0.17646	0.17699	0.17703

(a) Statistiche ottenute con il metodo analitico.

	Metodo HOTA N=500	Metodo HOTA N=5000	Valori esatti
Media	0.10602	0.10709	0.10667
Deviazione standard	0.03520	0.03556	0.03541
I quantile	0.08089	0.08167	0.08108
Mediana	0.10309	0.10362	0.10317
III quantile	0.12598	0.12814	0.12848
Estremo inferiore HPD	0.04335	0.04135	0.04242
Estremo superiore HPD	0.18173	0.17788	0.17703

(b) Statistiche ottenute con il metodo HOTA.

Tabella 3.1: Statistiche relative alla proporzione di pezzi difettosi.

	Metodo analitico N=500	Metodo analitico N=5000	Valori esatti
Media	0.13326	0.13350	0.13333
Deviazione standard	0.08421	0.08499	0.08498
I quantile	0.06906	0.06889	0.06879
Mediana	0.11724	0.11724	0.11702
III quantile	0.18101	0.18131	0.18104
Estremo inferiore HPD	0.00475	0.00451	0.00446
Estremo superiore HPD	0.29733	0.29927	0.29898

(a) Statistiche ottenute con il metodo analitico

	Metodo HOTA N=500	Metodo HOTA N=5000	Valori esatti
Media	0.13198	0.13448	0.13333
Deviazione standard	0.08440	0.08579	0.08498
I quantile	0.06847	0.07001	0.06879
Mediana	0.11702	0.11830	0.11702
III quantile	0.17467	0.18039	0.18104
Estremo inferiore HPD	0.00523	0.00506	0.00446
Estremo superiore HPD	0.30373	0.30400	0.29898

(b) Statistiche ottenute con il metodo HOTA

Tabella 3.2: Statistiche relative alla proporzione dei pezzi difettosi del primo giorno.

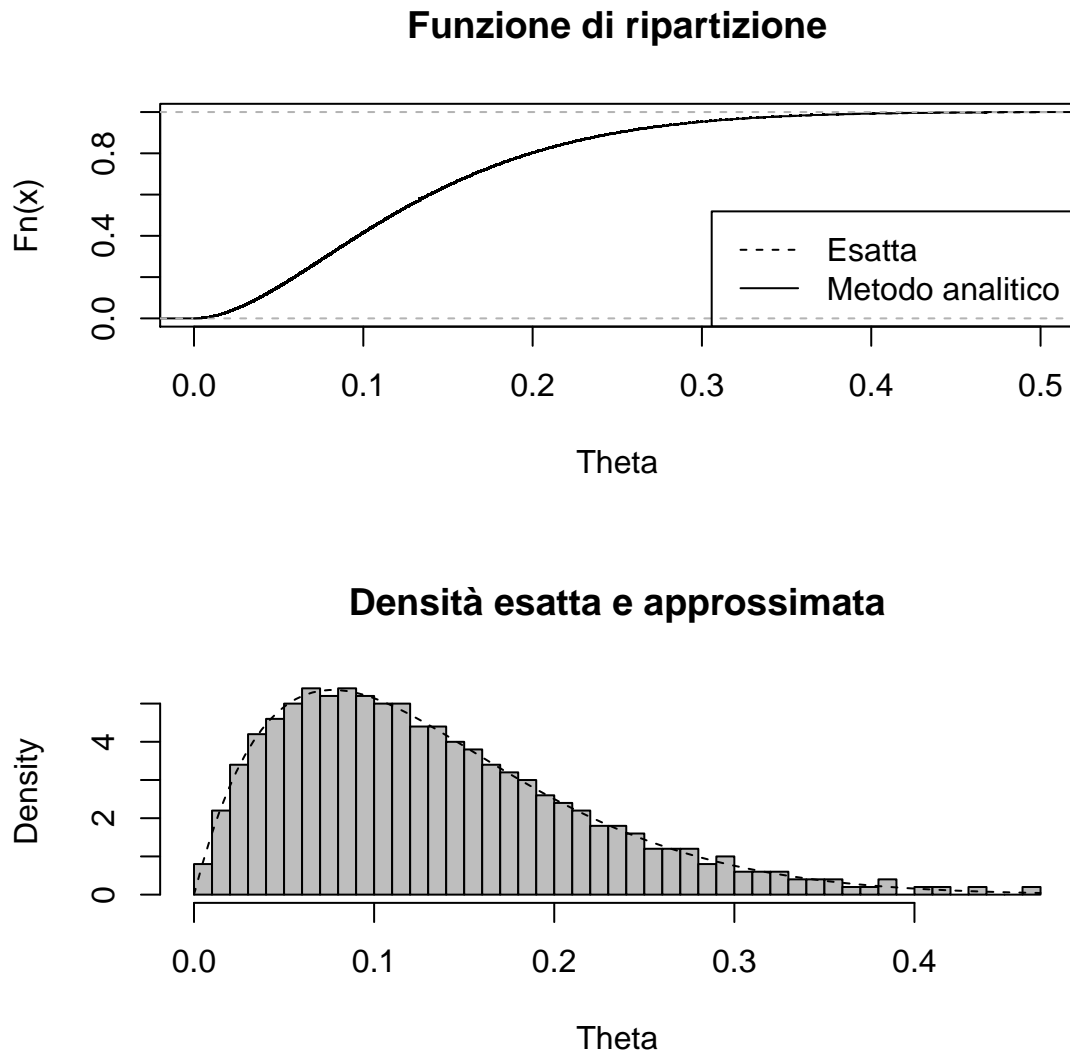


Figura 3.1: Modello Beta-Binomiale: confronto tra distribuzioni empiriche ottenute con il metodo analitico e distribuzioni teoriche esatte: funzione di ripartizione e densità $N=500$.

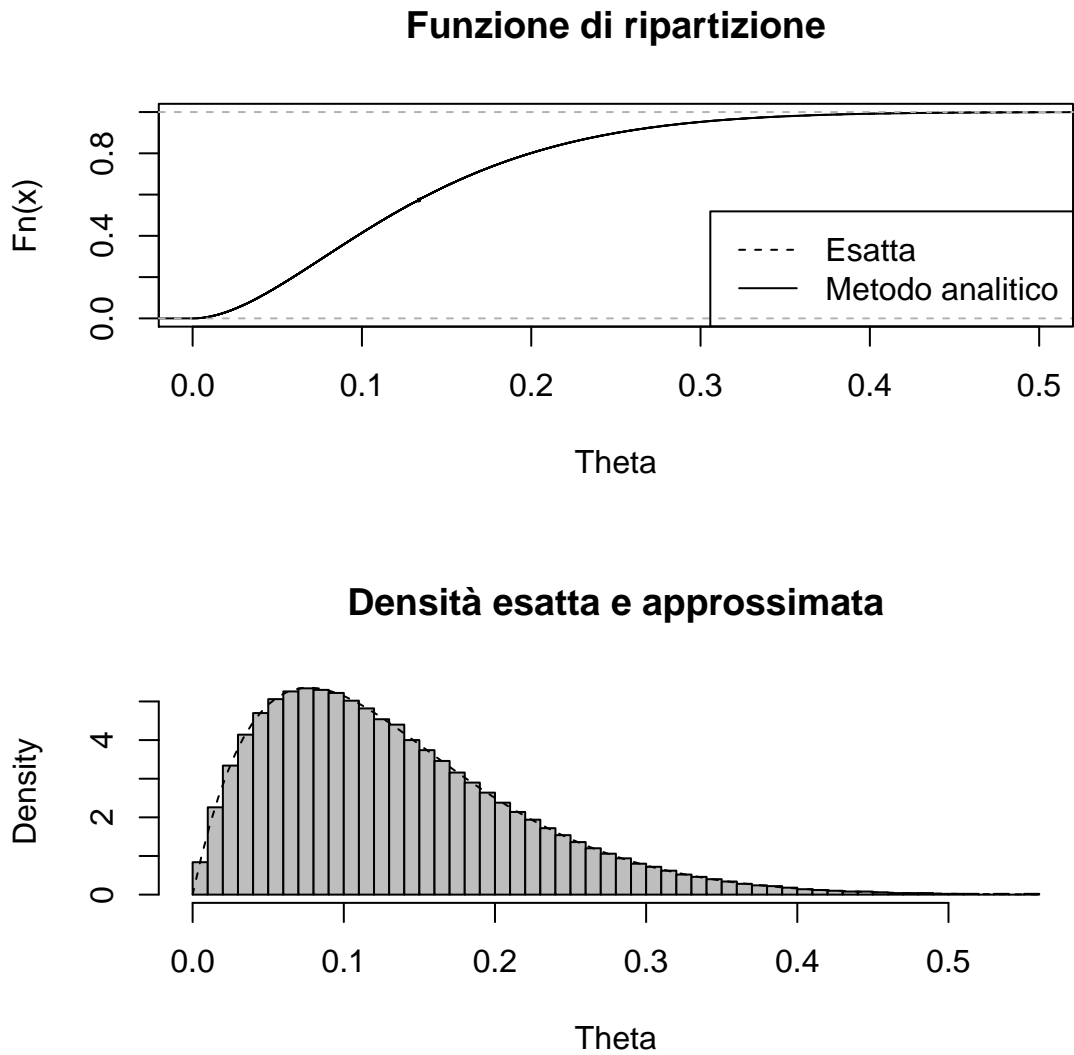


Figura 3.2: Modello Beta-Binomiale: confronto tra distribuzioni empiriche ottenute con il metodo analitico e distribuzioni teoriche esatte: funzione di ripartizione e densità $N=5000$.

3.2 Esempio NormaleGamma-Normale

I dati utilizzati in questo esempio si riferiscono a sette misurazioni craniche che possono essere considerate come realizzazioni di variabili casuali normali indipendenti di media e varianza incognite (Liseo, 2008). Le misurazioni in centimetri cubici sono 925, 892, 900, 875, 910, 906, 899. Si vuole fare inferenza sui parametri μ , media della distribuzione normale, e $\psi = 1/\sigma^2$, parametro di precisione pari all'inverso della varianza. Disponendo di realizzazioni di variabili casuali normali, la distribuzione a priori coniugata per il parametro di interesse $\theta = (\mu, \psi)$ è una *NormaleGamma*, ossia $\theta \sim Noga(\alpha, g, \delta, \lambda)$. I parametri α e g riassumono le informazioni disponibili a priori per quanto riguarda la media. I parametri δ e λ caratterizzano con le informazioni a priori la distribuzione di ψ . Affinchè tale distribuzione sia non informativa tutti e quattro i parametri vengono posti pari a zero. La distribuzione finale sarà ancora una *NormaleGamma* ma con parametri aggiornati secondo le espressioni:

$$\alpha^* = \frac{n\bar{y} + g\alpha}{n+g}, \quad g^* = g + n,$$

$$\lambda^* = \lambda + \frac{ns^2}{2} + \frac{ng}{2(n+g)}(\bar{y} - \alpha)^2, \quad \delta^* = \delta + \frac{n}{2}$$

dove \bar{y} rappresenta la media campionaria e s^2 rappresenta la varianza campionaria non corretta dei dati. L'uso di una distribuzione *NormaleGamma* per (μ, ψ) implica che marginalmente i due parametri abbiano rispettivamente una distribuzione *t di Student* e una distribuzione *Gamma* (Liseo, 2008). Tale risultato permette di ottenere le distribuzioni marginali esatte utilizzate per valutare l'accuratezza del metodo analitico. Per l'implementazione del metodo analitico, al fine di agevolare la risoluzione numerica, si è lavorato in scala logaritmica per il parametro ψ . Così facendo si è evitata l'introduzione di vincoli, ad esempio quello di positività, e si è potuto lavorare nell'intero campo dei numeri reali. Al momento di restituire il campione simulato si sono applicate le opportune trasformazioni inverse per ritornare alla parametrizzazione originale. Per generare l'output simile a quello riportato nel seguente paragrafo si possono utilizzare i seguenti comandi:

```
#dati
>dati=c(925,892,900,875,910,906,899)
#parametri della priori
>param=rep(1e-7,4)
#numero di punti nell'intervallo [0,1]
>n=5000
#output per mu
```

	Metodo analitico N=500	Metodo analitico N=5000	Valori esatti
Media	900.99979	900.99998	900.99999
Deviazione standard	8.43545	8.69394	6.42335
I quantile	896.16819	896.14685	897.13939
Mediana	900.97918	901.01401	900.99999
III quantile	905.83177	905.85311	904.86058
Estremo inferiore HPD	883.75141	883.58239	888.16139
Estremo superiore HPD	918.24855	918.46341	913.83858

(a) Media con il metodo analitico.

	Metodo HOTA N=500	Metodo HOTA N=5000	Valori esatti
Media	901.00677	900.87167	900.99999
Deviazione standard	6.19916	6.28187	6.42335
I quantile	897.23026	897.06383	897.13939
Mediana	901.08945	900.83630	900.99999
III quantile	904.87431	904.76560	904.86058
Estremo inferiore HPD	889.61745	888.17818	888.16139
Estremo superiore HPD	913.69001	913.06733	913.83858

(b) Media con il metodo HOTA.

Tabella 3.3: Statistiche per la media.

```

>media=model.mu(n,param,dati)
>summary.mu(media)
#output per psi
>var=model.psi(n,param,dati)
>summary.psi(var)

```

Risultati

Le tabelle Tabelle 3.3 e 3.4 mettono in evidenza che il metodo analitico non sia molto vicino alla distribuzione esatta mentre i valori ottenuti con il metodo HOTA sono molto più soddisfacenti. Questo si evince anche dalle figure sotto riportate. Probabilmente questa discrepanza tra i valori analitici e quelli esatti è dovuta ad un errore presente nel codice, che non si è riusciti a risolvere.

	Metodo analitico N=500	Metodo analitico N=5000	Valori esatti
Media	0.00323	0.00324	0.00485
Deviazione standard	0.00209	0.00211	0.00259
I quantile	0.00169	0.00168	0.00295
Mediana	0.00280	0.00280	0.00439
III quantile	0.00431	0.00432	0.00626
Estremo inferiore HPD	0.00016	0.00015	0.00069
Estremo superiore HPD	0.00733	0.00738	0.00996

(a) Parametro di precisione con il metodo analitico.

	Metodo HOTA N=500	Metodo HOTA N=5000	Valori esatti
Media	0.00488	0.00483	0.00485
Deviazione standard	0.00254	0.00256	0.00259
I quantile	0.00300	0.00295	0.00295
Mediana	0.00447	0.00436	0.00439
III quantile	0.00632	0.00627	0.00626
Estremo inferiore HPD	0.00086	0.00067	0.00069
Estremo superiore HPD	0.00984	0.00981	0.00996

(b) Parametro di precisione con il metodo HOTA.

Tabella 3.4: Statistiche per il parametro di precisione

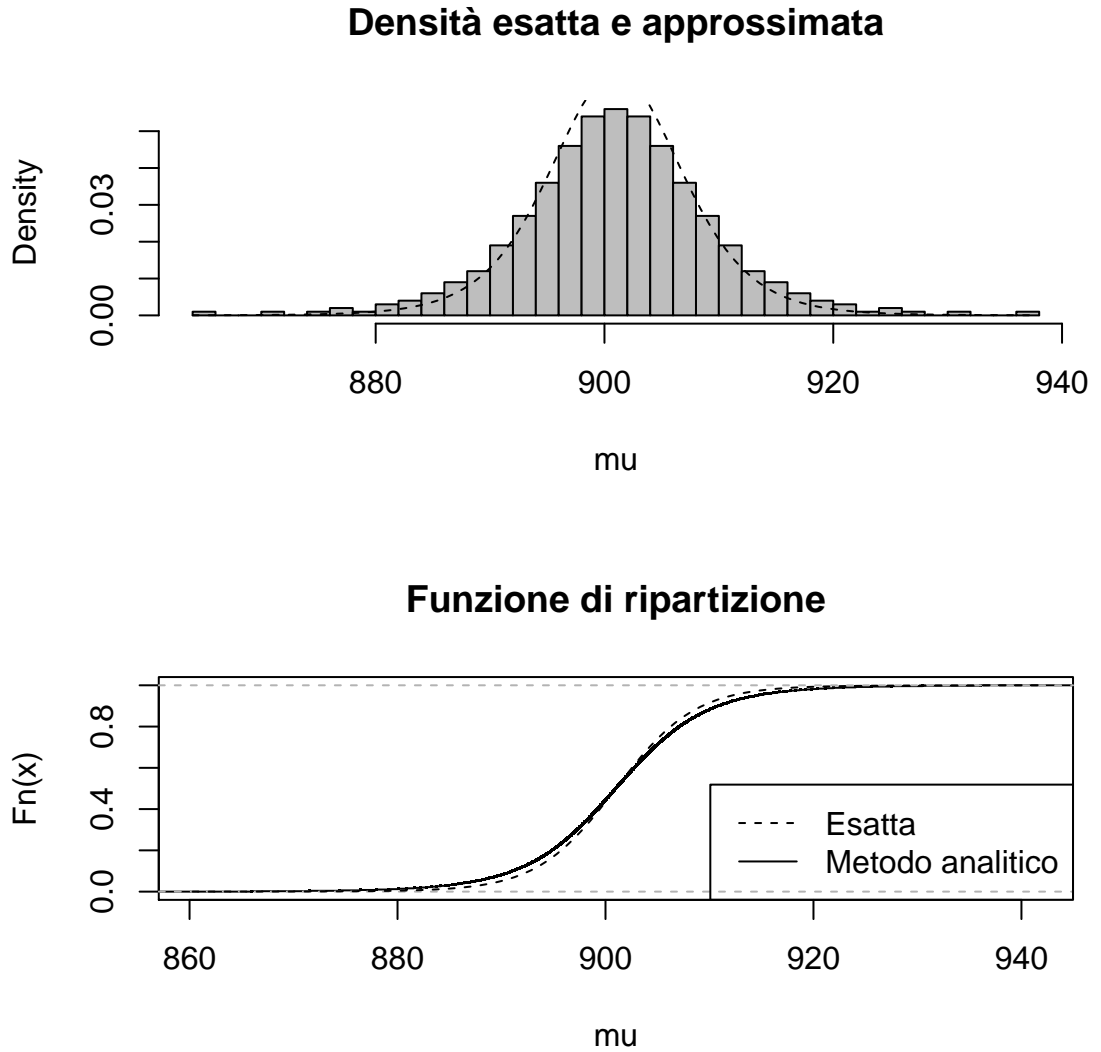


Figura 3.3: Modello NormaleGamma: confronto tra distribuzioni empiriche ottenute con il metodo analitico e ditribuzioni teoriche esatte: funzione di densità e di ripartizione per la media ($N=500$).

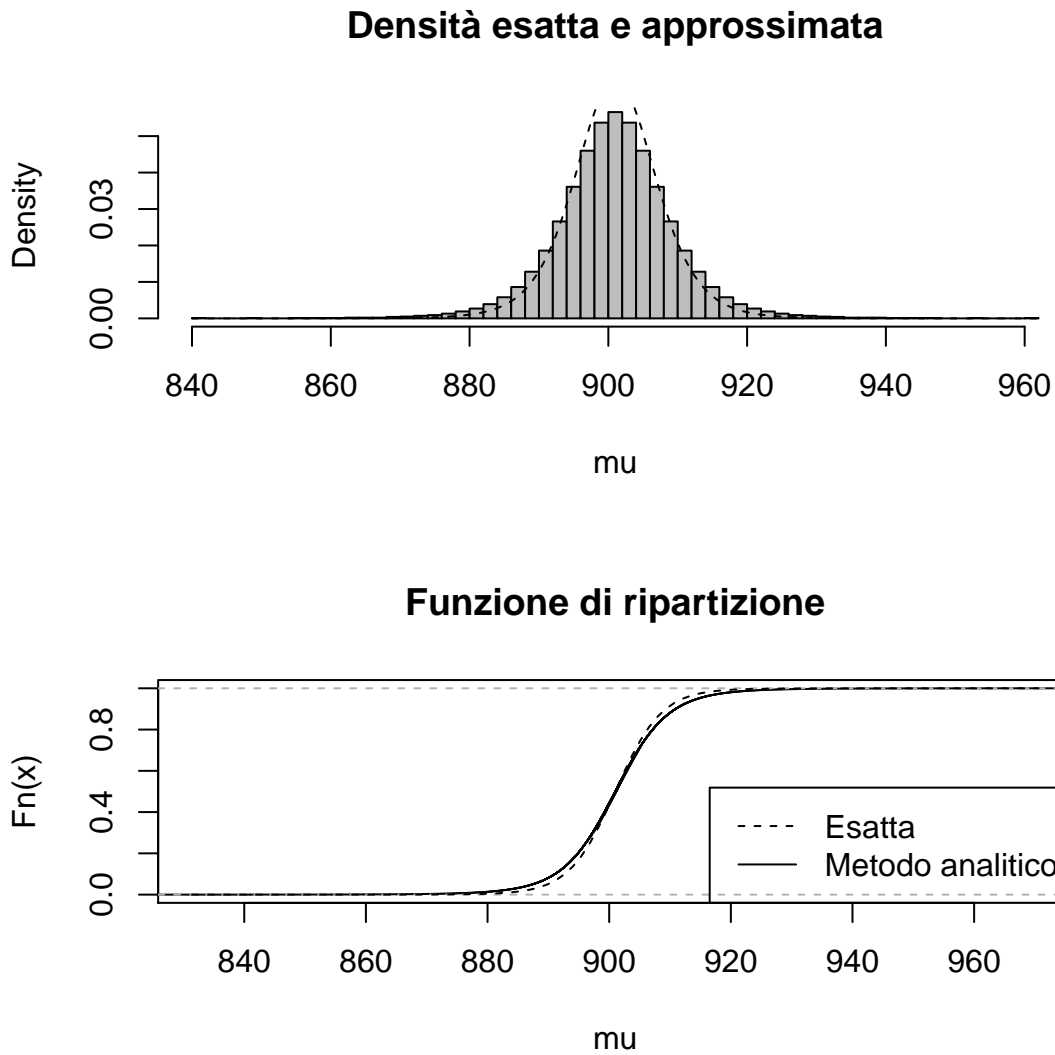


Figura 3.4: Modello NormaleGamma: confronto tra distribuzioni empiriche ottenute con il metodo analitico e distribuzioni teoriche esatte: funzione di densità e di ripartizione per la media ($N=5000$).

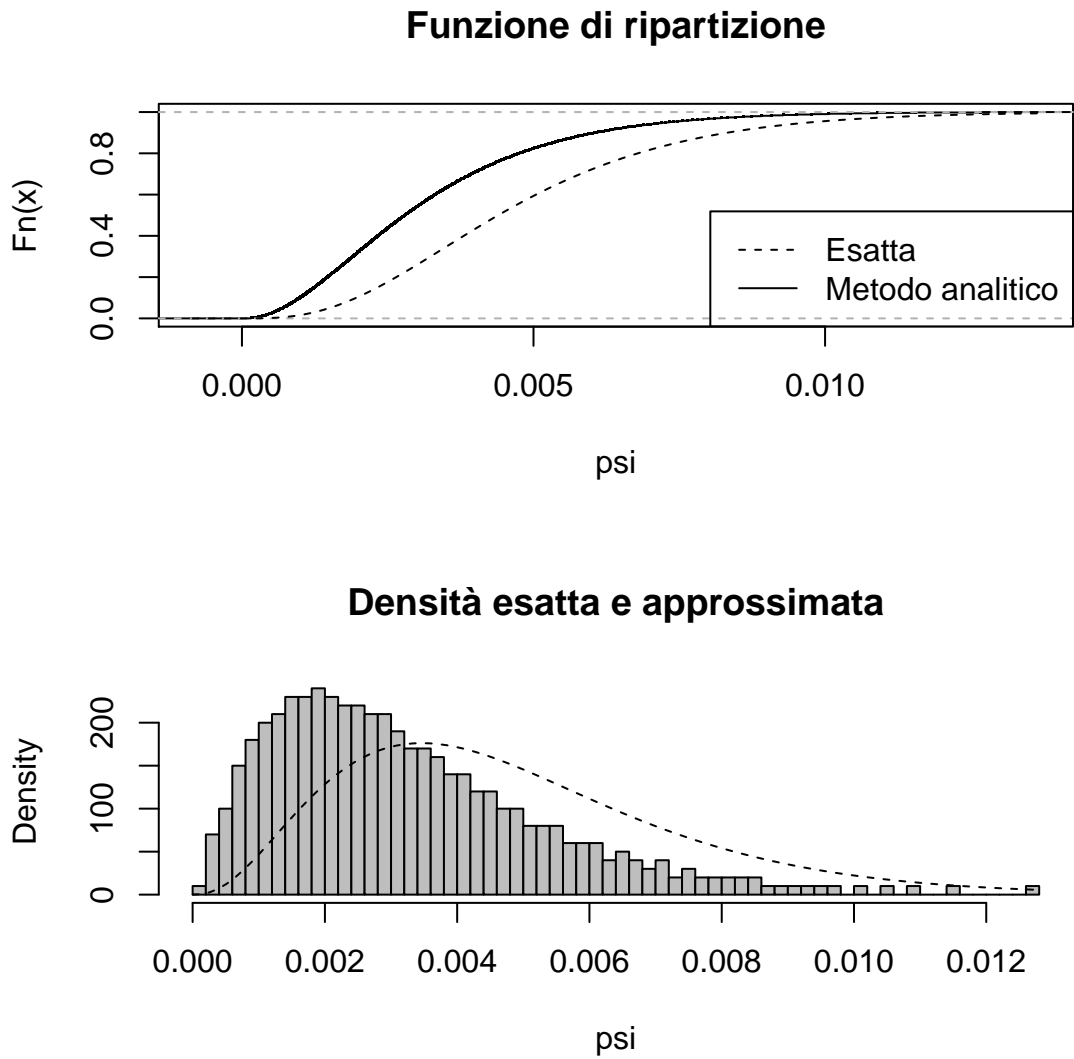


Figura 3.5: Modello NormaleGamma: confronto tra distribuzioni empiriche ottenute con il metodo analitico e distribuzioni teoriche esatte: funzione di ripartizione e di densità per il parametro di precisione ($N=500$).

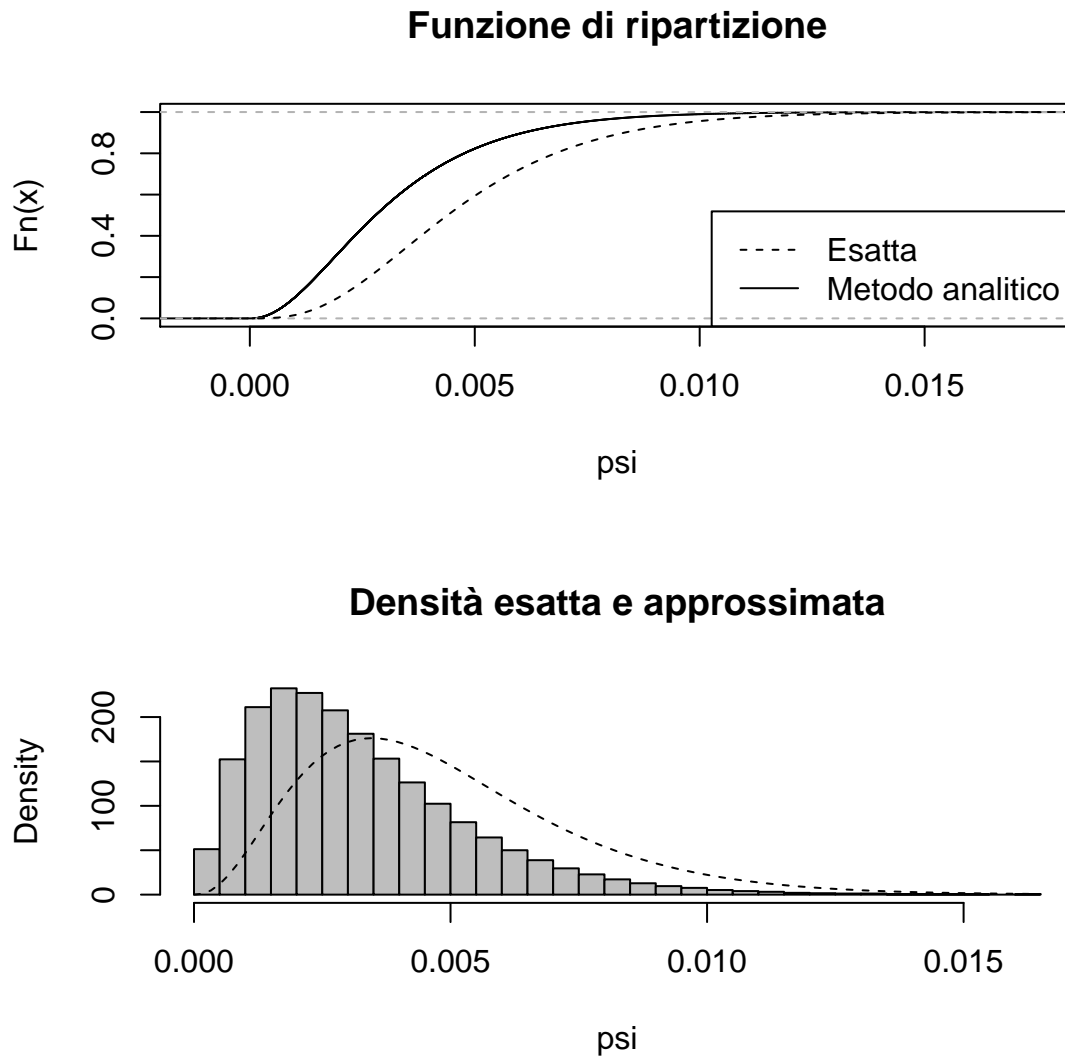


Figura 3.6: Modello NormaleGamma: confronto tra distribuzioni empiriche ottenute con il metodo analitico e distribuzioni teoriche esatte: funzione di ripartizione e di densità per il parametro di precisione ($N=5000$).

Capitolo 4

Applicazione del metodo analitico: stima di $P(X < Y)$

Nel precedente capitolo si è illustrata l'accuratezza del metodo analitico nell'approssimare le distribuzioni a posteriori marginali confrontandole con quelle esatte. Nella maggior parte delle applicazioni pratiche dell'inferenza bayesiana non si è in grado di giungere ad una distribuzione a posteriori disponibile in forma chiusa. In tutti questi casi si ricorre quindi a delle approssimazioni delle quantità di interesse, come quella ottenibile con il metodo analitico. Obiettivo di questo capitolo è di applicare il metodo analitico in due contesti in cui non si conosce la forma esatta della distribuzione a posteriori.

4.1 La quantità $P(X < Y)$

Ogni volta che un sistema viene sottoposto ad una qualche forma di azione *stressante* al fine di stimare la resistenza entra in gioco la quantità $P(X < Y)$. In termini più semplici si può interpretare la variabile casuale X come stress vissuto dalla componente e Y come forza della componente disponibile per superare lo stress. Secondo questo scenario semplificato se lo stress supera la forza ($X > Y$) la componente fallirebbe e viceversa. L'affidabilità viene quindi definita come la probabilità di non fallire: $P(X < Y)$. Per quanto possa apparire semplice, essa riveste un ruolo fondamentale in molti settori che spaziano dall'ingegneria alla psicologia fino alla medicina. Risultano quindi numerose le applicazioni in diverse discipline dei cosiddetti *stress-strength test*. Si pensi, ad esempio, al dover valutare l'affidabilità di una qualche componente elettronica o la sopravvivenza di un paziente sottoposto ad una particolare terapia. In quest'ultimo caso, X rappresenterebbe le aspettative

di vita di un paziente non trattato ed Y quelle di un paziente trattato. L'interesse sta quindi nel valutare se le aspettative di vita di un paziente non trattato siano inferiori rispetto a quelle di uno trattato, al fine di giudicare l'efficacia della terapia, cioè valutare proprio $P(X < Y)$.

Per una trattazione più esauriente dell'argomento si rimanda a Kotz, Lumelskii e Pensky (2003). Nel seguito verranno illustrati due esempi in cui le variabili X e Y si ipotizzeranno distribuite entrambe come due esponenziali o come due Weibull.

4.2 Variabili X e Y esponenziali

Si considera un insieme di dati relativi a delle misurazioni di uno studio sul Linfoma Anaplastico a Grandi Cellule eseguito dal Centro Oncoematologico Pediatrico di Padova, già analizzati in Ventura e Racugno, (2011). L'obiettivo era quello di meglio comprendere il legame della proteina Hsp70 con tale malattia. A tale scopo si è rilevato il livello della proteina Hsp70 su dieci pazienti malati, vettore x , e su quattro sani, vettore y , ottenendo:

$$\begin{aligned}x &= (2.8, 1.4, 0.13, 0.2, 0.8, 0.56, 0.44, 5.2, 1.7, 1.14) \\y &= (0.23, 0.44, 0.19, 0.08).\end{aligned}$$

Da informazioni extra sperimentali si può supporre che x e y siano realizzazioni delle variabili casuali indipendenti $X \sim Exp(\alpha)$ e $Y \sim Exp(\beta)$ con α e β parametri incogniti e a loro volta indipendenti. A priori si assume che, marginalmente, $\alpha \sim Gamma(\mu, \gamma)$ e che $\beta \sim Gamma(\nu, \delta)$, per cui la distribuzione di $\theta = (\alpha, \beta)$ è il prodotto delle due *Gamma* data l'indipendenza dei parametri. La quantità di interesse è $\psi = P(X < Y)$ e per poter fare inferenza è necessario ricavare la sua distribuzione marginale a posteriori. Per fare ciò si effettua la seguente riparametrizzazione (Kotz, Lumelskii e Pensky, 2003, Paragrafo 2.3.3):

$$\tau : \psi = \frac{\alpha}{\alpha + \beta}, \lambda = \alpha + \beta \quad (4.1)$$

che ha come inversa:

$$\theta : \alpha = \psi\lambda, \beta = \lambda(1 - \psi). \quad (4.2)$$

Il determinante dello Jacobiano per quest'ultima parametrizzazione è pari a:

$$|J_{\theta}(\psi, \lambda)| = \lambda. \quad (4.3)$$

Si giunge così alla distribuzione a posteriori nella nuova parametrizzazione tramite l'espressione:

$$\pi_{\tau}(\psi, \lambda|x, y) = \pi(\theta(\psi, \lambda)|x, y)|J_{\theta}(\psi, \lambda)|. \quad (4.4)$$

L'espressione (4.4), non essendo disponibile in forma chiusa, è stata approssimata tramite il metodo analitico, descritto nel Capitolo 2. Per facilitare la risoluzione numerica, si è utilizzata una riparametrizzazione aggiuntiva:

$$\omega^* : \omega = \log\left(\frac{\psi}{1-\psi}\right), \lambda^* = \lambda, \quad (4.5)$$

che ha come inversa:

$$\tau^* : \psi = \frac{\exp(\omega)}{1 + \exp(\omega)}, \lambda = \lambda^*. \quad (4.6)$$

Il determinante dello Jacobiano per quest'ultima parametrizzazione è pari a:

$$|J_{\tau^*}(\omega, \lambda^*)| = \frac{\exp(\omega)}{(1 + \exp(\omega))^2}. \quad (4.7)$$

Si è quindi applicato il metodo analitico a

$$\pi_{\omega^*}(\omega, \lambda^* | x, y) = \pi_{\tau}(\tau^*(\omega, \lambda^*) | x, y) |J_{\tau^*}(\omega, \lambda^*)|. \quad (4.8)$$

Al momento di restituire i valori si è tornati alla parametrizzazioni τ .

Per generare l'output simile a quello riportato nel seguente paragrafo si possono utilizzare questi comandi, le funzioni sono riportate nell'Appendice B:

```
#dati
>x=c(2.8,1.4,0.13,0.2,0.8,0.56,0.44,5.2,1.7,1.14)
>y=c(0.23,0.44,0.19,0.08)
#parametri della priori
>param=rep(1e-7,4)
#numero di punti in cui suddividere l'intervallo [0,1]
>N=5000
>s=model.psi(N,param,x,y)
>summary.psi(s)
```

Risultati

Nella tabella vengono riportati alcuni indicatori sintetici che riassumono la distribuzione a posteriori approssimata per il parametro di interesse ψ . Nelle Figure 4.1 e 4.2 si possono vedere invece la forma della funzione di densità, ottenuta con un semplice istogramma, e della funzione di ripartizione.

Confrontando i risultati ottenuti con il metodo analitico e quelli ottenuti con il metodo HOTA, riportati nella Tabella 4.1 si nota che la varianza ottenuta con il primo metodo è inferiore rispetto a quella ottenuta con il metodo HOTA.

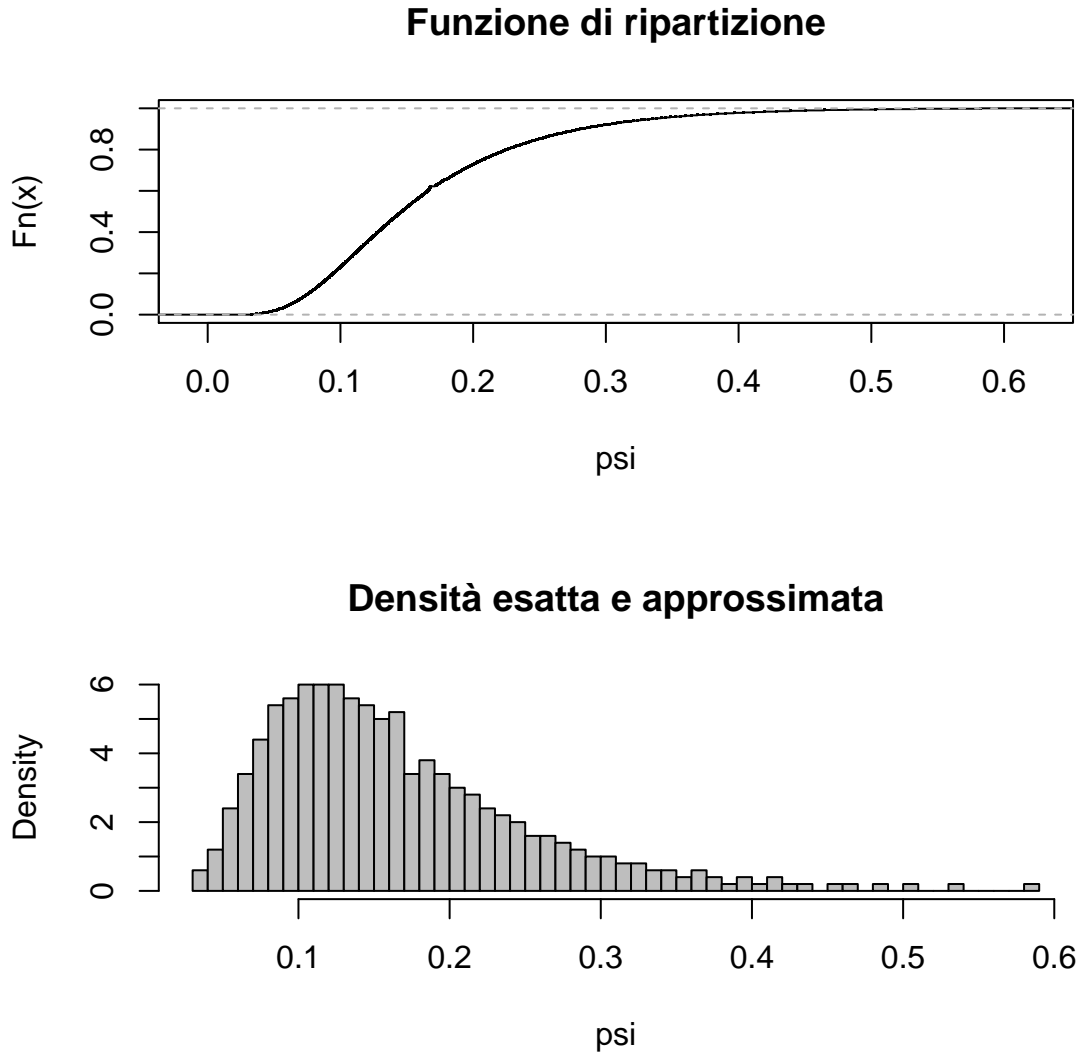


Figura 4.1: Modello Esponenziale: funzione di ripartizione e densità ottenute con il metodo analitico, $N=500$. Variabili esponenziali.

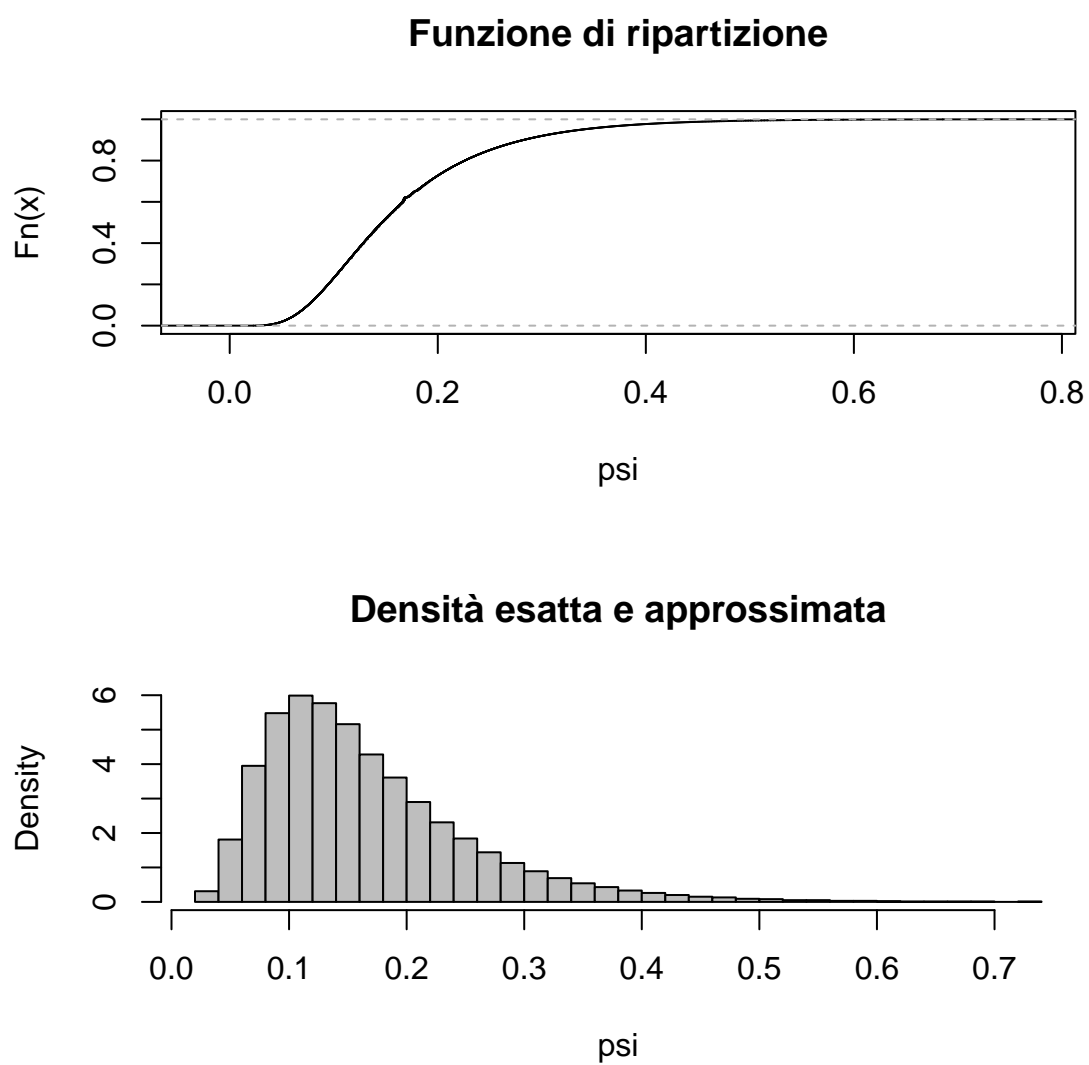


Figura 4.2: Modello Esponenziale: funzione di ripartizione e densità ottenute con il metodo analitico, $N=5000$. Variabili esponenziali.

Tabella 4.1: Statistiche per ψ ottenuti con il metodo analitico e con il metodo HOTA

	Metodo analitico N=500	Metodo analitico N=5000	Metodo HOTA N=5000
Media	0.16566	0.16599	0.16716
Deviazione standard	0.08706	0.08838	0.08976
I quantile	0.10338	0.10322	0.10423
Mediana	0.14627	0.14628	0.14715
III quantile	0.20703	0.20733	0.20642
Estremo inferiore HPD	0.03682	0.03614	0.03722
Estremo superiore HPD	0.34161	0.34383	0.34978

4.3 Variabili X e Y Weibull

I dati utilizzati sono gli stessi dell'esempio precedente ma questa volta si assume che le variabili X e Y abbiano una distribuzione *Weibull*. Di conseguenza consideriamo che x e y siano realizzazioni di due variabili casuali indipendenti e aventi lo stesso parametro di forma $X \sim Weibull(\alpha, k)$ e $Y \sim Weibull(\beta, k)$, e che α , β e k siano tre parametri incogniti e indipendenti. Il parametro di interesse è $\psi = P(X < Y)$ e per fare inferenza è necessario reperire la sua distribuzione a posteriori marginale. Per fare ciò si effettua la seguente parametrizzazione (Kotz, Lumelskii e Pensky, 2003, Paragrafo 3.3.3):

$$\tau : \psi = \frac{\beta^k}{\alpha^k + \beta^k}, \lambda_1 = \frac{\alpha^k \beta^k}{\alpha^k + \beta^k}, \lambda_2 = k, \quad (4.9)$$

che ha come inversa:

$$\theta : \alpha = \left(\frac{\lambda_1}{\psi}\right)^{\frac{1}{\lambda_2}}, \beta = \left(\frac{\lambda_1}{1-\psi}\right)^{\frac{1}{\lambda_2}}, k = \lambda_2. \quad (4.10)$$

Per il parametro $\tau = (\psi, \lambda_1, \lambda_2)$ si è specificata una distribuzione a priori non informativa di Jeffreys la cui espressione per la densità è data da (Kotz, Lumelskii e Pensky, 2003, Paragrafo 3.3.3):

$$\pi(\psi, \lambda_1, \lambda_2) = \frac{1}{(1-\psi)\psi\lambda_1\lambda_2}. \quad (4.11)$$

Si è ottenuta la distribuzione a posteriori nella parametrizzazione τ :

$$\pi_\tau(\psi, \lambda_1, \lambda_2 | x, y), \quad (4.12)$$

da cui si è approssimata con il metodo analitico la distribuzione marginale per il parametro di interesse ψ . Anche qui, per facilitare la risoluzione numerica, si è utilizzata una riparametrizzazione aggiuntiva:

$$\omega^* : \omega = \log\left(\frac{\psi}{1-\psi}\right), \lambda_1^* = \log \lambda_1, \lambda_2^* = \log \lambda_2 \quad (4.13)$$

che ha come inversa

$$\tau^* : \psi = \frac{\exp(\omega)}{1 + \exp(\omega)}, \lambda_1 = \exp(\lambda_1^*), \lambda_2 = \exp(\lambda_2^*). \quad (4.14)$$

Il determinante dello Jacobiano per quest'ultima parametrizzazione è pari a:

$$|J_{\tau^*}(\omega, \lambda_1^*, \lambda_2^*)| = \frac{\exp(\omega + \lambda_1^* + \lambda_2^*)}{(1 + \exp(\omega))^2}. \quad (4.15)$$

Si è quindi applicato il metodo analitico a

$$\pi_{\omega^*}(\omega, \lambda_1^*, \lambda_2^* | x, y) = \pi_{\tau}(\tau^*(\omega, \lambda_1^*, \lambda_2^*) | x, y) |J_{\tau^*}(\omega, \lambda_1^*, \lambda_2^*)|. \quad (4.16)$$

Al momento di restituire i valori si è tornati alla parametrizzazione τ .

Per generare l'output simile a quello riportato nel seguente paragrafo si possono utilizzare questi comandi:

```
#dati
>x=c(2.8,1.4,0.13,0.2,0.8,0.56,0.44,5.2,1.7,1.14)
>y=c(0.23,0.44,0.19,0.08)
#numero di punti nell'intervallo [0,1]
>N=5000
>s=model.psi(N,x,y)
>summary.psi(s)
```

Risultati

Alcuni indicatori sintetici sono riportati nella Tabella 4.2 e riassumono la distribuzione a posteriori approssimata per il parametro di interesse ψ . Come si vede nella Tabella 4.2 i risultati sembrerebbero soddisfacenti, tuttavia andando a vedere le Figure 4.3 e 4.4 si nota che c'è un'instabilità numerica tra 0.2 e 0.3 e intorno a 0.8. Al momento della stesura di questo elaborato non è ancora chiara quale sia la causa di questo problema, anche se è probabile che sia imputabile a qualche errore nel codice.

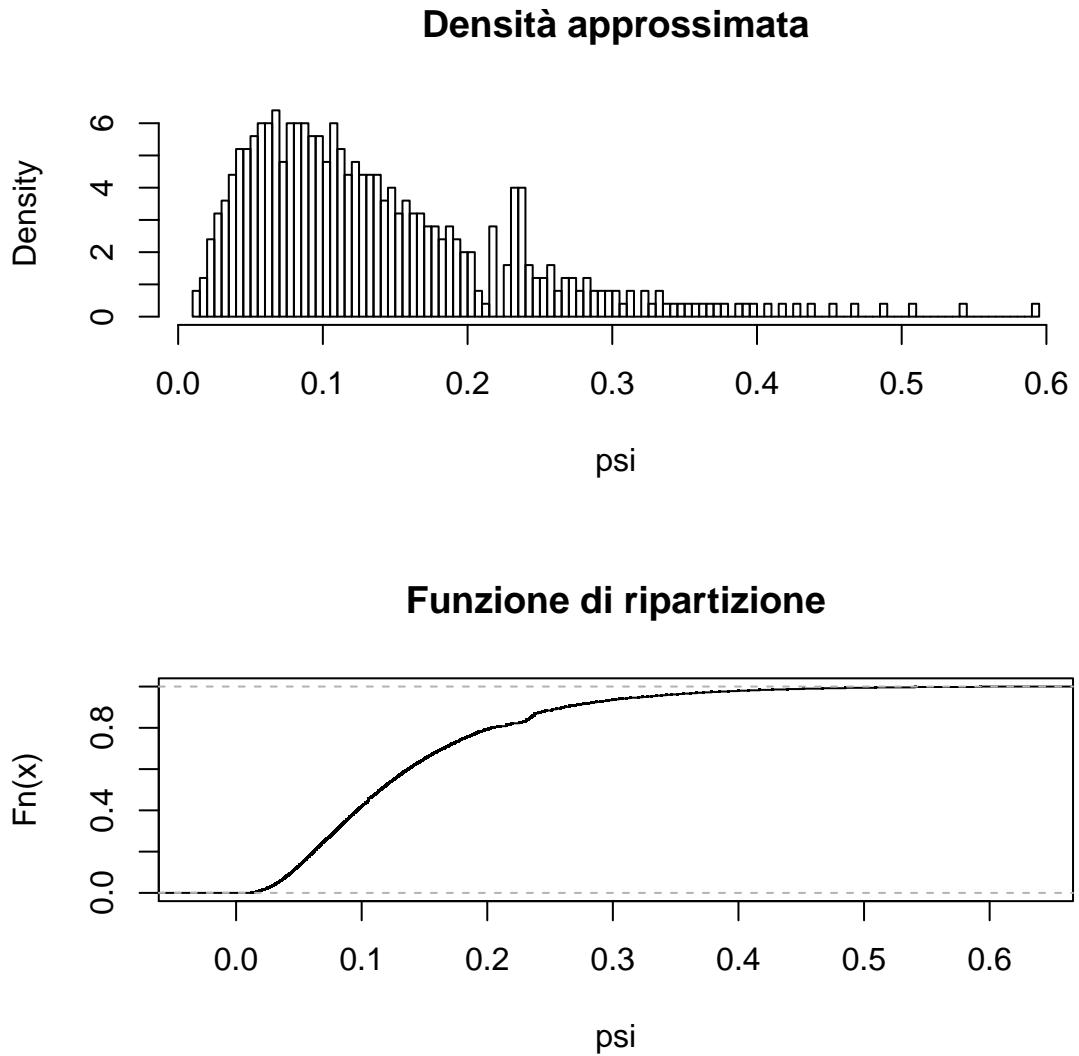


Figura 4.3: Modello Weibull: funzione di densità e di ripartizione ottenute con il metodo analitico, $N=500$. Variabili X e Y distribuite come Weibull.

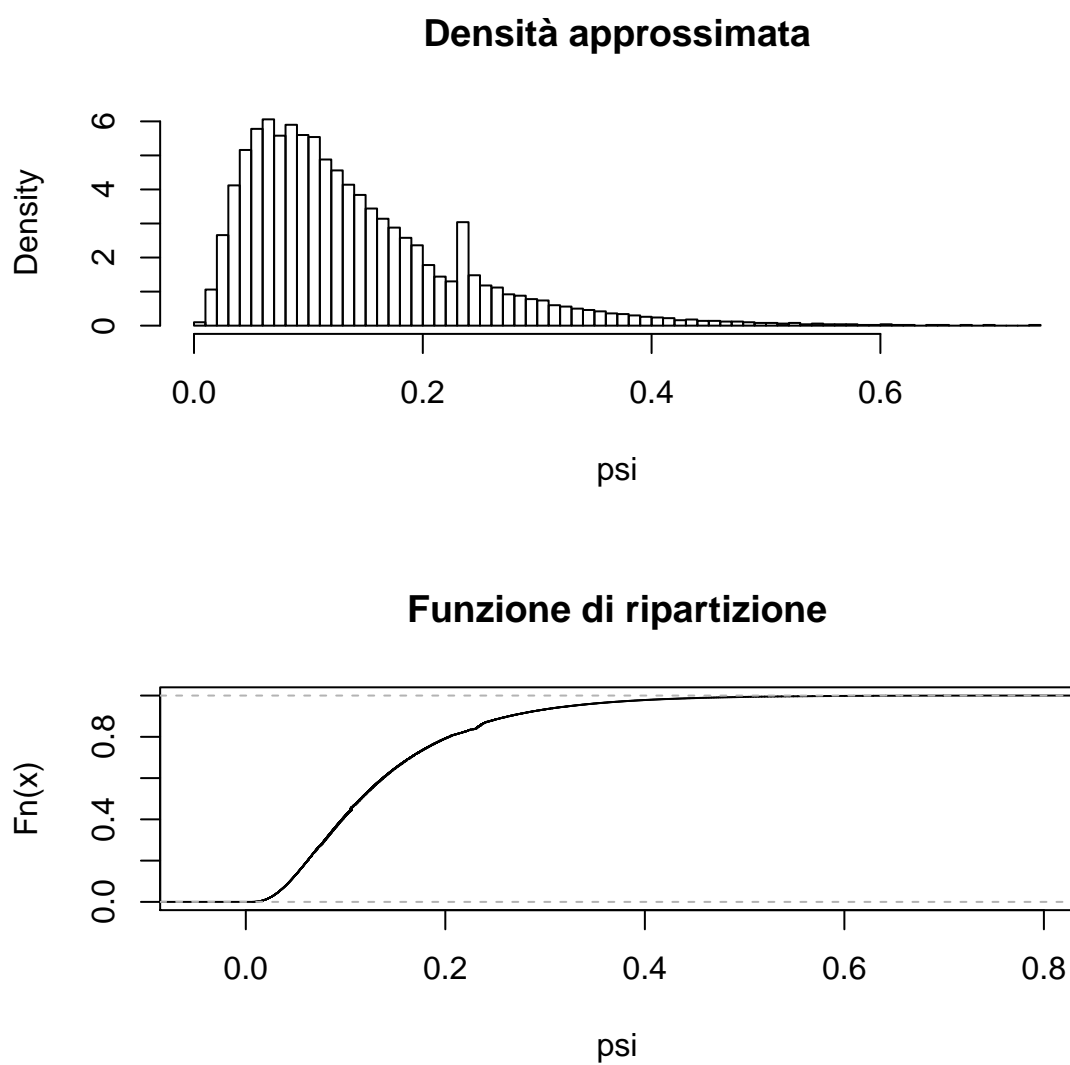


Figura 4.4: Modello Weibull: funzione di densità e di ripartizione ottenute con il metodo analitico, $N=5000$. Variabili X e Y distribuite come Weibull.

Tabella 4.2: Statistiche per ψ ottenute con il metodo analitico e con il metodo HOTA

	Metodo analitico N=500	Metodo analitico N=5000	Metodo HOTA N=5000
Media	0.13897	0.13907	0.13861
Deviazione standard	0.09399	0.09527	0.09615
I quantile	0.07024	0.07009	0.07108
Mediana	0.11517	0.11500	0.11478
III quantile	0.18222	0.18219	0.17769
Estremo inferiore HPD	0.01188	0.01155	0.01227
Estremo superiore HPD	0.32734	0.32943	0.33477

Conclusioni

Nello sviluppo della relazione si è messo a confronto un metodo analitico, basato su approssimazioni asintotiche della distribuzione marginale a posteriori, con il metodo HOTA, che unisce alle approssimazioni analitiche un passo di simulazione. Si voleva soprattutto verificare che l'accuratezza con il metodo analitico fosse superiore a quella ottenuta con il metodo HOTA. Si nota che la varianza del metodo analitico è inferiore di quella ottenuta con il metodo HOTA, tuttavia, in alcuni esempi, si è notato che il metodo di simulazione è riuscito ad approssimare in modo migliore la distribuzione a posteriori marginale. Inoltre, il tempo di elaborazione impiegato dal metodo HOTA è di molto inferiore rispetto a quello impiegato dal metodo analitico.

Appendice A

Codici R per il capitolo 3

Codice A.1: Esempio Beta-Binomiale

```
1 library(TeachingDemos)
2 library(numDeriv)
3
4 llik.binom=function(theta,data) {
5 #calcolo della log-verosimiglianza della funzione binomiale
6 sum.xi = sum(data[,2])
7 sum.mi.xi = sum(data[,1]-data[,2])
8 k = prod(choose(data[,1],data[,2]))
9 sum.xi * log(theta) + sum.mi.xi * log(1-theta) + log(k)
10 }
11
12 lprior.beta=function(theta,alpha,beta) {
13 #calcolo dell'apriori
14 dbeta(theta,alpha,beta,log=TRUE)
15 }
16
17 lpost.un=function(theta,alpha,beta,data) {
18 #calcolo della log-posteriori non normalizzata l(theta)+l(prior)
19 llik.binom(theta,data)+lprior.beta(theta,alpha,beta)
20 }
21
22 nlpost.un=function(theta,alpha,beta,data) {
23 #calcolo della log-posteriori per trovare il massimo
24 -(llik.binom(theta,data)+lprior.beta(theta,alpha,beta))}
25
26 lpost.un.omega=function(omega,alpha,beta,data) {
```

```

26 #riparamp(omega)/(1+exp(omega))
27 lpost.un(theta,alpha,beta,data)+
28     log(exp(omega)/(1+exp(omega))^2)
29 }
30
31 nlpost.un.omega=function(omega,alpha,beta,data) {
32 -lpost.un.omega(omega,alpha,beta,data)
33 }
34
35 r.tilde=function(theta,alpha,beta,data,mode) {
36
37 #calcolo di r
38
39 w = 2 * (lpost.un.omega(mode,alpha,beta,data) -
40         lpost.un.omega(theta,alpha,beta,data))
41
42 r = sign(mode-theta) * w^0.5
43
44 grad = grad(f=lpost.un.omega, x=theta, alpha=alpha, beta=beta,
45            data=data, method.args=list(eps=1e-6, d=0.0001, r=6))
46
47 D = (-hessian(f=lpost.un.omega, x=mode, alpha=alpha, beta=beta,
48             data=data, method.args=list(eps=1e-6, d=0.0001, r=6)))^0.5
49
50 qb = grad/D
51
52 rstar = r+1/r*log(qb/r)
53
54 return(rstar)
55 }
56
57 model.beta.binom=function(n,alpha,beta,data) {
58 oo = nlminb(start=0.5, objective=nlpost.un.omega, alpha=alpha,
59            beta=beta, data=data)
60
61 # moda della posteriori:
62 mode = oo$par
63
64 # hessiano della posteriori calcolato nella moda
65 j = hessian(f=nlpost.un.omega, x=mode, alpha=alpha, beta=beta,
66            data=data)
67 # "j^-1"
68 jinv = solve(j)

```



```

69
70 # "j^-0.5"
71 se = sqrt(jinv)
72
73 #preparato contenitore n valori
74 psi = rep(NA,n)#double(n)
75
76 #griglia di valori
77 y = (1:n) / (n+1)
78
79 t= qnorm(y, 0, 1)
80
81 # per ogni ... trova lo psi per cui vale ...
82 for(i in 1:n) {
83   psi[i] = uniroot(function(x)
84     r.tilde(x,alpha=alpha,beta=beta,data=data,mode=mode) - t[i]
85     ,lower=-10, upper=10)$root
86 }
87
88 theta=exp(psi)/(1+exp(psi))
89
90 return(list(data=data, alpha=alpha, beta=beta, psi=psi, y=y, theta=theta))}
91 hpd.beta=function(prob, alphaA, betaA)
92 {
93   #Calcolo di HPD a partire da distribuzione esatta a posteriori
94
95   mode=(alphaA -1)/(alphaA+betaA -2)
96
97   #Funzioni per trovare i theta con quella densit?
98   theta1=function(k){
99     uniroot(function(x) dbeta(x, alphaA, betaA)-k
100     ,lower=0, upper=mode)$root
101 }
102
103   theta2=function(k){
104     uniroot(function(x) dbeta(x, alphaA, betaA)-k
105     ,lower=mode, upper=1)$root
106 }
107
108   #Trovo valore della densit? per cui ho un intervallo per theta di
109   #probabilit? prob
110
111   dens=uniroot(function(x){

```

```

112         pbeta(theta2(x), alphaA, betaA) -
113         pbeta(theta1(x), alphaA, betaA) - prob}
114         , lower=0, upper=dbeta(mode, alphaA, betaA))$root
115
116 #HPD
117 HPD.R=theta1(dens)
118
119 HPD.L=theta2(dens)
120
121 return(list(HPD1=HPD.R, HPD2=HPD.L))
122
123
124 }
125 #riassunto
126 summary.beta.binom=function(obj.beta.binom)
127 {
128 #mostro le statistiche descrittive
129 theta=obj.beta.binom$theta
130
131 data=obj.beta.binom$data
132
133 alpha=obj.beta.binom$alpha
134
135 beta=obj.beta.binom$beta
136
137 alphaA=alpha+sum(data[,2])
138
139 betaA=beta+sum(data[,1]-data[,2])
140
141 hpdex=unlist(hpd.beta(0.95, alphaA, betaA))
142
143 listan=list(c(mean=round(mean(theta),5),
144              sd=round(sqrt(var(theta)),5),
145              Q1=round(quantile(theta,prob=0.25),5),
146              Median=round(median(theta),5),
147              Q3=round(quantile(theta,prob=0.75),5),
148              HPD=round(emp.hpd(theta),5)))
149
150 listex=list(c(mean=round(alphaA/(alphaA+betaA),5),
151              sd=round(sqrt(alphaA*betaA/((alphaA+betaA)^2*
152              (alphaA+betaA+1))),5),
153              Q1=round(qbeta(0.25, alphaA, betaA),5),
154              Median=round(qbeta(0.5, alphaA, betaA),5),

```

```

155         Q3=round(qbeta(0.75, alphaA, betaA),5),round(hpdex,5))
156
157 par(mfrow=c(2,1))
158
159 plot(x=unlist(a1$psi),y=unlist(a1$y),type="l")
160 plot(x=theta,y=unlist(a1$y),type="l")
161
162 plot.ecdf(theta,main="Funzione di ripartizione",xlab="Theta",
163           xlim=c(0,0.5))
164 curve(pbeta(x, alphaA, betaA),add=T,lty=2)
165
166 hist(a1$theta,nclass=50,freq=F,main="Densita' esatta e approssimata",
167      xlab="Theta",col=8)
168 curve(dbeta(x, alphaA, betaA),add=T,lty=2)
169
170 legend("topright",legend=c("Esatta","Metodo analitico"),lty=c(2,1))
171 result=list(Summary.analitico=unlist(listan),
172           Summary.esatto=unlist(listex))
173
174 tab=cbind(unlist(listan),unlist(listex))
175 result
176 }
177
178 data = cbind(rep(15,5), c(2,1,3,1,1))
179 data1 = cbind(15,2)
180 n=5000
181 a1 = model.beta.binom(n, 1e-7, 1e-7, data1)
182 summary.beta.binom(a1)

```

Codice A.2: EsempioNormaleGamma-Normale

```

1 library(TeachingSampling)
2 library(TeachingDemos)
3 library(lestat)
4 library(numDeriv)

5 llik.norm=function(theta,data)
6 {
7 #calcolo della log-verosimiglianza per mu e psi,
8 #parametro di precisione
9 n=length(data)
10 mu=theta[1]

```

```

11 psi=theta[2]
12 -0.1*n*log(2*psi)+0.5*n*log(psi)-0.5*psi*sum((data-mu)^2)
13 }
14
15 lprior.noga=function(theta,param)
16 {
17 #calcolo della log-priori:noga
18 alpha=param[1]
19 g=param[2]
20 delta=param[3]
21 lambda=param[4]
22 mu=theta[1]
23 psi=theta[2]
24
25 #obj=normalgamma(alpha,g,delta,lambda)
26 #probabilitydensity(obj,val=theta,log=T)
27 log((lambda^delta*sqrt(g))/(gamma(delta)*sqrt(2*psi))*psi^
28 (delta-0.5)*exp(-lambda*psi)*exp(-g*psi*0.5*(mu-alpha)^2))
29 }
30
31 lpost2.un=function(theta,param,data)
32 {
33 #calcolo della log-posteriori non normalizzata
34 #l(theta)+l(prior)
35 llik.norm(theta,data)+lprior.noga(theta,param)
36 }
37
38 nlpost2.un=function(theta,param,data)
39 {
40 #calcolo della log-posteriori non normalizzata
41 #per trovare il massimo
42 -lpost2.un(theta,param,data)
43 }
44
45 lpost2.un.omega=function(omega,param,data)
46 {
47 #nuova parametrizzazione: omega=(mu,log(psi)) per semplicita
48 #di calcolo
49 #calcolo della log-posteriori non normalizzata
50 #l(theta)+l(prior)
51 theta=c(omega[1],exp(omega[2]))
52 lpost2.un(theta,param,data)+omega[2]
53 #omega[2] e' il determinante dello jacobiano

```

```

54 #della parametrizzazione
55 }

56 nlpost2.un.omega=function(omega,param,data)
57 {
58 #calcolo della log-posteriori non normalizzata
59 #per trovare il massimo
60 -lpost2.un.omega(omega,param,data)
61 }
62
63 r.tilde.mu=function(mu,param,data,mode)
64 {
65
66 #calcolo di r con parametro di interesse mu
67 #calcolo della stima vincolata di psi per mu fissato
68 fix.psi=nlminb(start=1,objective=function(x) nlpost2.un.omega
69               (omega=c(mu,x)param=param,data=data))$par
70
71 r=sign(mode[1]-mu)*(2*(lpost2.un.omega(omega=mode,param=param,
72                                     data=data)-lpost2.un.omega(omega=c(mu,fix.psi),
73                                     param=param,data=data)))^0.5
74
75 #gradiente calcolato in mu e psi rispetto vincolato,
76 #prendo il primo elemento
77 grad=grad(f=lpost2.un.omega,x=c(mu,fix.psi),param=param,data=data,
78          method.args=list(eps=1e-6,d=0.0001,r=6))[1]
79
80 #calcolo la matrice hessiana negativa nella moda per arrivare
81 #al denominatore di qb
82 nhess.mode=-hessian(f=lpost2.un.omega,x=mode,param=param,data=data,
83                    method.args=list(eps=1e-6,d=0.0001,r=6))
84
85 D=sqrt(det(nhess.mode))
86
87 #calcolo la matrice hessiana negativa in mu e psi vincolato,
88 #prendo il blocco(2,2) e trovo il num di qb
89 nhess.fix=-hessian(f=lpost2.un.omega,x=c(mu,fix.psi),param=param
90                  ,data=data,method.args=list(eps=1e-6,d=0.0001,r=6))
91
92 N=sqrt(nhess.fix[2,2])
93
94 qb=grad*N/D
95
96 rstar=r+(1/r)*log(qb/r)

```

```

97
98 return(rstar)
99 }
100
101 r.tilde.psi=function(psi,param,data,mode)
102 {
103
104 #calcolo di r con parametro di interesse psi
105 #(parametro di precisione)
106 #calcolo della stima vincolata di mu con psi fissata
107 fix.mu=nlminb(start=.1,objective=function(x)
108               nlpost2.un.omega(omega=c(x,psi)
109               ,param=param,data=data))$par
110
111 r=sign(mode[2]-psi)*
112     (2*(lpost2.un.omega(omega=mode,param=param,data=data)
113     -lpost2.un.omega(omega=c(fix.mu,psi),
114     param=param,data=data)))^0.5
115
116 #gradiente calcolato in mu vincolato e psi,
117 #prendo il secondo elemento
118 grad=grad(f=lpost2.un.omega,x=c(fix.mu,psi),
119          param=param,data=data,
120          method.args=list(eps=1e-6,d=0.0001,r=6))[2]
121
122 #calcolo della la matrice hessiana negativa per arrivare
123 #al denominatore di qb
124 nhess.mode=-hessian(f=lpost2.un.omega,x=mode,param=param,
125                    data=data,method.args=list(eps=1e-6,d=0.0001,r=6))
126
127 D=sqrt(det(nhess.mode))
128
129 #calcolo la matrice hessiana negativa in mu vincolato e psi,
130 #prendo il blocco [1,1] e trovo il num di qb
131 nhess.fix=-hessian(f=lpost2.un.omega,x=c(fix.mu,psi),
132                  param=param,data=data,
133                  method.args=list(eps=1e-6,d=0.0001,r=6))
134
135 N=sqrt(nhess.fix[1,1])
136
137 qb=grad*N/D
138
139 rstar=r+(1/r)*log(qb/r)

```

```

140
141 return(rstar)
142 }

144 model.mu=function(n,param,data)
145 {
146
147 #cerco la moda vicino alla smv, dico al risolutore che psi
148 #non puo' essere negativo
149 smv=c(mean(data),log(length(data)/(length(data)-1)/var(data)))
150
151 oo=nlminb(start=smv,objective=nlpost2.un.omega,
152           param=param,data=data)
153
154 mode=oo$par
155
156 j=hessian(f=nlpost2.un.omega,x=mode,param=param,data=data,
157          method.args=list(eps=1e-6,d=0.0001,r=6))
158
159 jinv=solve(j)
160
161 se=sqrt(jinv[1,1])
162
163 #preparo contenitore n valori
164 mu=rep(NA,n)
165
166 #griglia di valori
167 y=(1:n)/(n+1)
168
169 t=qnorm(y,0,1)
170
171 #per ogni y trova lo psi per cui vale
172 for(i in 1:n) {
173 mu[i]=uniroot(function(x)
174               r.tilde.mu(x,param=param,data=data,mode=mode)-t[i],
175               lower=0,upper=1000)$root
176 }
177 return(list(data=data,param=param,mu=mu,y=y))
178
179 }

180
181 model.psi=function(n,param,data)
182 {
183 smv=c(mean(data),log(length(data)/(length(data)-1)/var(data)))

```

```
184
185 oo=nlminb(start=smv,objective=nlpost2.un.omega,
186           param=param,data=data)
187
188 mode=oo$par
189
190 j=hessian(f=nlpost2.un.omega,x=mode,param=param,data=data,
191          method.args=list(eps=1e-6,d=0.0001,r=6))
192
193 jinv=solve(j)
194
195 se=sqrt(jinv[2,2])
196
197 #preparo contenitore dei valori
198 val=rep(NA,n)
199
200 #griglia di valori
201 y=(1:n)/(n+1)
202
203 t=qnorm(y,0,1)
204
205 #per ogni y trovo lo psi corrispondente
206 for(i in 1:n){
207   val[i]=uniroot(function(x)
208                 r.tilde.psi(x,param=param,data=data,mode=mode)-t[i],
209                 lower=-10,upper=0)$root
210 }
211
212 psi=exp(val)
213
214 return(list(data=data,param=param,psi=psi,y=y))
215 }
216
217 hpd.mu=function(prob,paramA)
218 {
219 #calcolo di HPD a partire da distribuzione esatta a posteriori
220 alphaA=paramA[1]
221
222 gA=paramA[2]
223
224 deltaA=paramA[3]
225
226 lambdaA=paramA[4]
```



```

227
228 mode=alphaA
229
230 #funzioni per trovare i mu dato un valore per la densita'
231 theta1=function(k){uniroot(function(x)
232     dt((x-alphaA)*sqrt((deltaA*gA)/lambdaA),
233     df=2*deltaA)*sqrt((deltaA*gA)/lambdaA)-k,
234     lower=-1e10, upper=mode)$root
235 }
236
237 theta2=function(k){uniroot(function(x)
238     dt((x-alphaA)*sqrt((deltaA*gA)/lambdaA),
239     df=2*deltaA)*sqrt((deltaA*gA)/lambdaA)-k,
240     lower=mode, upper=1e10)$root
241 }
242
243 #trovo valore della densita' per cui ho un intervallo per theta
244 #di probabilita' prob
245 dens=uniroot(function(x)
246     {pt((theta2(x)-alphaA)*sqrt((deltaA*gA)/lambdaA),
247     df=2*deltaA)-prob}, lower=1e-10,
248     upper=dt(0, df=2*deltaA)*sqrt((deltaA*gA)/lambdaA))$root
249
250 #HPD
251 HPD.R=theta1(dens)
252 HPD.L=theta2(dens)
253 }
254
255 summary.mu=function(obj.mu)
256 {
257 #mostro statistiche descrittive per la distribuzione marginale di mu
258 #e confronto con esatta
259 mu=obj.mu$mu
260
261 param=obj.mu$param
262
263 data=obj.mu$data
264
265 #aggiorno i parametri per la marginale a posteriori
266 alpha=param[1]
267
268 g=param[2]
269

```

```
270 delta=param[3]
271
272 lambda=param[4]
273
274 n=length(data)
275
276 mean=mean(data)
277
278 s2=var(data)*(n-1)/n
279
280 alphaA=(n*mean+g*alpha)/(n+g)
281
282 gA=g+n
283
284 lambdaA=lambda+0.5*n*s2+0.5*n*g*(mean-alpha)^2/(n+g)
285
286 deltaA=delta+n*0.5
287
288 #reciproco del parametro di scala
289 invscale=sqrt((deltaA*gA)/lambdaA)
290
291 df=2*deltaA
292
293 hpdex=unlist(hpd.mu(0.95,c(alphaA,gA,deltaA,lambdaA)))
294
295 listan=list(c(mean=round(mean(mu),5),
296             sd=round(sqrt(var(mu)),5),
297             Q1=round(quantile(mu,prob=0.25),5),
298             median=round(median(mu),5),
299             Q3=round(quantile(mu,prob=0.75),5),
300             HPD=round(emp.hpd(mu),5)))
301
302 listex=list(c(mean=round(alphaA,5),
303             sd=round(sqrt(df/(df-2)/invscale^2),5),
304             Q1=round(qt(0.25,df=df)/invscale+alphaA,5),
305             median=round(qt(0.5,df=df)/invscale+alphaA,5),
306             Q3=round(qt(0.75,df=df)/invscale+alphaA,5),
307             round(hpdex,5)))
308
309 par(mfrow=c(2,1))
310
311 hist(mu,nclass=50,freq=F,main="Densita' esatta e approssimata",
312      xlab="mu",col=8)
```

```

313 curve(dt((x-alphaA)*sqrt((deltaA*gA)/lambdaA),df=2*deltaA)*
314         sqrt((deltaA*gA)/lambdaA),add=T,lty=2)
315
316 plot.ecdf(mu,main="Funzione di ripartizione",xlab="mu",pch=".")
317 curve(pt((x-alphaA)*sqrt((deltaA*gA)/lambdaA),df=2*deltaA),
318       add=T,lty=2)
319 legend("topright",legend=c("Esatta","Metodo analitico"),
320       lty=c(2,1))
321
322 result=list(summary.analitico=unlist(listan),
323            summary.esatto=unlist(listex))
324 result
325 }
326 hpd.gamma=function(prob,deltaA,lambdaA)
327 {
328 #calcolo HPD a partire da distribuzione esatta a posteriori
329 mode=(deltaA-1)/lambdaA
330
331 #funzioni per trovare i psi dato un valore per la densita'
332 theta1=function(k){
333 uniroot(function(x) dgamma(x,deltaA,lambdaA)-k,lower=1e-10,
334         upper=mode)$root
335 }
336 theta2=function(k){
337 uniroot(function(x) dgamma(x,deltaA,lambdaA)-k,lower=mode,
338         upper=1e10)$root
339 }
340
341 #trovo valore della densita' per cui ho un intervallo per theta
342 #di probabilita' prob
343 dens=uniroot(function(x){pgamma(theta2(x),deltaA,lambdaA)
344         -pgamma(theta1(x),deltaA,lambdaA)-prob},
345         lower=1e-10,upper=dgamma(mode,deltaA,lambdaA))$root
346
347 #HPD
348 HPD.R=theta1(dens)
349 HPD.L=theta2(dens)
350 return(list(HPD1=HPD.R,HPD2=HPD.L))
351 }
352
353 summary.psi=function(obj.psi)
354 {
355 #mostro statistiche descrittive per la distribuzione marginale

```

```
356 #di psi e confronto con esatta
357 psi=obj.psi$psi
358
359 param=obj.psi$param
360
361 data=obj.psi$data
362
363 #aggiorno parametri per la marginale a posteriori
364 alpha=param[1]
365
366 g=param[2]
367
368 delta=param[3]
369
370 lambda=param[4]
371
372 n=length(data)
373
374 mean=mean(data)
375
376 s2=var(data)*(n-1)/n
377
378 alphaA=(n*mean+g*alpha)/(n+g)
379
380 gA=g+n
381
382 lambdaA=lambda+0.5*n*s2+0.5*n*g*(mean-alpha)^2/(n+g)
383
384 deltaA=delta+n*0.5
385
386 hpdex=unlist(hpd.gamma(0.95,deltaA,lambdaA))
387
388 listan=list(c(mean=round(mean(psi),5),
389             sd=round(sqrt(var(psi)),5),
390             Q1=round(quantile(psi,prob=0.25),5),
391             median=round(median(psi),5),
392             Q3=round(quantile(psi,prob=0.75),5),
393             HPD=round(emp.hpd(psi),5)))
394
395 listex=list(c(mean=round(deltaA/lambdaA,5),
396             sd=round(sqrt(deltaA/lambdaA^2),5),
397             Q1=round(qgamma(0.25,deltaA,lambdaA),5),
398             median=round(qgamma(0.5,deltaA,lambdaA),5),
```

```
399     Q3=round(qgamma(0.75,deltaA,lambdaA),5),
400     round(hpdex,5))
401
402 result=list(summary.analitico=unlist(listan),
403            summary.esatto=unlist(listex))
404
405 par(mfrow=c(2,1))
406
407 #plot(x=unlist(a$theta),y=unlist(a$y),type="l")
408
409 plot.ecdf(psi,main="Funzione di ripartizione",xlab="psi",
410          pch=".")
411 curve(pgamma(x,deltaA,lambdaA),add=T,lty=2)
412
413 hist(psi,nclass=50,freq=F,main="Densita' esatta e approssimata",
414      xlab="psi",col=8)
415 curve(dgamma(x,deltaA,lambdaA),add=T,lty=2)
416
417 legend("topright", legend=c("Esatta","Metodo analitico"),
418       lty=c(2,1))
419 result
420 }
421
422 data=c(925,892,900,875,910,906,899)
423 param=rep(1e-7,4)
424 n=5000
425
426 a=model.mu(n,param,data)
427 summary.mu(a)
428 obj.mu=a
429
430 b=model.psi(n,param,data)
431 summary.psi(b)
432 ocj.psi=b
```

Appendice B

Codici R per il capitolo 4

Codice B.1: Esempio $P(X < Y)$: Esponenziali

```
1 library(numDeriv)
2 library(TeachingDemos)
3
4 llik.exp=function(theta,dataX,dataY)
5 {
6 #calcolo della log-verosimiglianza per alpha e beta,
7 #parametri delle esponenziali
8 alpha=theta[1]
9 beta=theta[2]
10 sum(dexp(dataX,alpha,log=T))+sum(dexp(dataY,beta,log=T))
11 }
12
13 lprior.gamma=function(theta,param)
14 {
15 #calcolo della log-priori per theta: prodotto di due gamma
16 #data l'indipendenza di X e Y e di alpha e beta
17 alpha=theta[1]
18 beta=theta[2]
19 sa=param[1]
20 ra=param[2]
21 sb=param[3]
22 rb=param[4]
23 log(dgamma(alpha,sa,ra)*dgamma(beta,sb,rb))
24 }
25
26 lpost3.un=function(theta,param,dataX,dataY)
```

```

26 {
27 #calcolo della log-posteriori non normalizzata
28 #l(theta)+l(prior)
29 llik.exp(theta,dataX,dataY)+lprior.gamma(theta,param)
30 }
31
32 nlpost3.un=function(theta,param,dataX,dataY)
33 {
34 #calcolo della log-posteriori non normalizzata
35 #per trovare il massimo
36 -lpost3.un(theta,param,dataX,dataY)
37 }
38
39 lpost3.un.omega=function(omega,param,dataX,dataY)
40 {
41 #omega e' una nuova parametrizzazione: considero il logit
42 #della prob x<y e la somma di alpha+beta
43 #come parametro di disturbo
44 #logit della prob x<y (psi)
45 w=omega[1]
46
47 #somma di alpha e beta
48 l=omega[2]
49 alpha=exp(w)/(1+exp(w))*l
50 beta=(1-exp(w)/(1+exp(w)))*l
51 lpost3.un(theta=c(alpha,beta),param,dataX,dataY)+log(l)+
52         log(exp(w)/((1+exp(w))^2))
53 }
54
55 nlpost3.un.omega=function(omega,param,dataX,dataY)
56 {
57 -lpost3.un.omega(omega,param,dataX,dataY)
58 }
59
60 r.tilde.psi=function(w,param,dataX,dataY,mode)
61 {
62 fix.l=nlminb(start=1,objective=function(x)
63             nlpost3.un.omega(omega=c(w,x),param=param,dataX=dataX,
64             dataY=dataY),lower=1e-7)$par
65
66 r=sign(mode[1]-w)*(2*(lpost3.un.omega
67             (omega=mode,param=param,dataX=dataX,dataY=dataY)
68             -lpost3.un.omega(omega=c(w,fix.l),param=param,

```



```

69         dataX=dataX,dataY=dataY)))^0.5
70
71 #gradiente calcolato in w e l vincolato
72 grad=grad(f=lpost3.un.omega,x=c(w,fix.l),param=param,
73         dataX=dataX,dataY=dataY,method.args=
74         list(eps=1e-6,d=0.0001,r=6))[1]
75
76 #matrice hessiana negativa nella moda per arrivare
77 #al denominatore di qb
78 nhess.mode=-hessian(f=lpost3.un.omega,x=mode,param=param,
79         dataX=dataX,dataY=dataY,
80         method.args=list(eps=1e-6,d=0.0001,r=6))
81
82 D=sqrt(det(nhess.mode))
83
84 #matrice hessiana negativa in w e l vincolato,
85 #prendo il blocco (2,2)
86 #e trovo il num di qb
87 nhess.fix=-hessian(f=lpost3.un.omega,x=c(w,fix.l),param=param,
88         dataX=dataX,dataY=dataY,
89         method.args=list(eps=1e-6,d=0.0001,r=6))
90
91 N=sqrt(nhess.fix[2,2])
92
93 qb=grad*N/D
94
95 rstar=r+(1/r)*log(qb/r)
96
97 return(rstar)
98 }
99
100 model.psi=function(n,param,dataX,dataY)
101 {
102     mode=nlminb(start=c(1,1),objective=nlpost3.un.omega,param=param,
103         dataX=dataX,dataY=dataY,lower=c(-Inf,1e-7))$par
104     j=hessian(f=nlpost3.un.omega,x=mode,param=param,dataX=dataX,
105         dataY=dataY,method.args=list(eps=1e-6,d=0.0001,r=6))
106
107     jinv=solve(j)
108
109     se=sqrt(jinv[1,1])
110
111     #preparo contenitore n valori

```

```
112 theta=rep(NA,n)
113
114 #griglia di valori
115 t=(1:n)/(n+1)
116
117 z=qnorm(t,0,1)
118
119 #per ogni z trova lo psi per cui vale
120 for(i in 1:n){
121   theta[i]=uniroot(function(x)
122     r.tilde.psi(x,param=param,dataX=dataX,dataY=dataY,mode=mode)
123     -z[i],lower=-10,upper=10)$root
124 }
125
126 psi=exp(theta)/(1+exp(theta))
127 return(list(psi=psi,t=t,theta=theta))
128 }
129
130 summary.psi=function(theta)
131 {
132   psi=unlist(a$psi)
133
134   listan=list(c(mean=round(mean(psi),5),
135               sd=round(sqrt(var(psi)),5),
136               Q1=round(quantile(psi,prob=0.25),5),
137               median=round(median(psi),5),
138               Q3=round(quantile(psi,prob=0.75),5),
139               HPD=round(emp.hpd(psi),5)))
140
141   par(mfrow=c(2,1))
142   plot.ecdf(theta,main="Funzione di ripartizione",
143             xlab="psi")
144
145   hist(theta,nclass=50,freq=F,
146         main="Densita' esatta e approssimata",
147         xlab="psi",col=8)
148
149   unlist(listan)
150 }
151
152 y=c(0.23,0.44,0.19,0.08)
153 x=c(2.8,1.4,0.13,0.2,0.8,0.56,0.44,5.2,1.7,1.14)
154 param=rep(1e-7,4)
```

```

155 n=5000
156 a=model.psi(n,param,x,y)
157 summary.psi(a)

```

Codice B.2: Esempio $P(X<Y)$: Weibull

```

1 library(numDeriv)
2 library(TeachingDemos)
3
4 llik.weib=function(theta,dataX,dataY){
5 #calcolo della log-verosimiglianza per s1,s2,k,
6 #in questa parametrizzazione
7
7 s1=theta[1]
8 s2=theta[2]
9 k=theta[3]
10 (sum(dweibull(dataX,scale=s1,shape=k,log=T))+
11 sum(dweibull(dataY,scale=s2,shape=k,log=T)))
12 }
13
14 lprior=function(tau){
15 #calcolo della log priori per theta; e' non informativa
16 #di Jeffreys; considera la parametrizzazione (psi,l1,l2)
17 #con psi parametro di interesse e l1 e l2 parametri
18 #di disturbo
19 psi=tau[1]
20 l1=tau[2]
21 l2=tau[3]
22 log(1/((1-psi)*psi*l1*l2))
23 }
24
25 lpost.un=function(omega,dataX,dataY){
26
27 w=omega[1]
28 l1s=omega[2]
29 l2s=omega[3]
30
31 #inverto parametrizzazione per il calcolo di lprior;
32 #trasformo con lo jacobiano
33 psi=exp(w)/(1+exp(w))
34 l1=exp(l1s)
35 l2=exp(l2s)

```

```

36 prior=lprior(c(psi,l1,l2))+log(exp(w+l1s+l2s)/((1+exp(w))^2))
37
38 #inverto parametrizzazione per il calcolo della log-veros;
39 #sfrutto parametro di invarianza
40 s1=(l1/psi)^(1/l2)
41 s2=(l1/(1-psi))^(1/l2)
42 k=l2
43 llik=llik.weib(c(s1,s2,k),dataX,dataY)
44 prior+llik
45 }
46
47 nlpost.un=function(omega,dataX,dataY){
48 -lpost.un(omega,dataX,dataY)
49 }
50
51 r.tilde.psi=function(w,dataX,dataY,mode){
52
53 fix.l=nlminb(start=c(1,1),objective=function(x)
54             nlpost.un(omega=c(w,x),dataX=dataX,dataY=dataY))$par
55
56 r=sign(mode[1]-w)*(2*(
57             lpost.un(omega=mode,dataX=dataX,dataY=dataY)
58             -lpost.un(omega=c(w,fix.l),dataX=dataX,dataY=dataY)))^0.5
59
60 grad=grad(f=lpost.un,x=c(w,fix.l),dataX=dataX,dataY=dataY,
61           method.args=list(eps=1e-6,d=0.0001,r=6))[1]
62
63 nhess.mode=-hessian(f=lpost.un,x=mode,dataX=dataX,dataY=dataY,
64                    method.args=list(eps=1e-6,d=0.0001,r=6))
65
66 D=sqrt(det(nhess.mode))
67
68 nhess.fix=-hessian(f=lpost.un,x=c(w,fix.l),dataX=dataX,
69                  dataY=dataY,method.args=list(eps=1e-6,d=0.0001,r=6))
70
71 N=sqrt(det(nhess.fix[-1,-1]))
72
73 qb=grad*N/D
74
75 rstar=r+(1/r)*log(qb/r)
76
77 return(rstar)
78 }

```

```

79
80 model.psi=function(n,dataX,dataY){
81
82 mode=nlminb(start=c(1,1,1),objective=nlpost.un,dataX=dataX,
83             dataY=dataY)$par
84
85 j=hessian(f=nlpost.un,x=mode,dataX=dataX,dataY=dataY,
86           method.args=list(eps=1e-6,d=0.0001,r=6))
87
88 jinv=solve(j)
89
90 se=sqrt(jinv[1,1])
91
92 #preparo contenitore per n valori
93 par=rep(NA,n)
94
95 #griglia di valori
96 t=(1:n)/(n+1)
97
98 z=qnorm(t,0,1)
99
100 #per ogni z trovo lo psi corrispondente
101 for(i in 1:n){
102   par[i]=uniroot(function(x)
103                 r.tilde.psi(x,dataX=dataX,dataY=dataY,mode=mode)-z[i],
104                 lower=-10,upper=10)$root
105 }
106
107 psi=exp(par)/(1+exp(par))
108
109 return(list(t=t,par=par,psi=psi))
110 }
111
112 summary.psi=function(theta){
113
114 psi=unlist(a1$psi)
115
116 listan=list(c(mean=round(mean(psi),5),sd=round(sqrt(var(psi)),5),
117             Q1=round(quantile(psi,prob=0.25),5),
118             median=round(median(psi),5),
119             Q3=round(quantile(psi,prob=0.75),5),
120             HPD=round(emp.hpd(psi),5)))
121
122 hist(psi,nclass=100,freq=F,main="Densita' approssimata",

```

```
122     xlab="psi")
123 plot.ecdf(psi, main="Funzione di ripartizione",
124           xlab="psi",pch=".")
125
126 unlist(listan)
127 }
128
129
130 y=c(0.23,0.44,0.19,0.08)
131 x=c(2.8,1.4,0.13,0.2,0.8,0.56,0.44,5.2,1.7,1.14)
132 n=5000
133 a1=model.psi(n,x,y)
134 summary.psi(a1)
```

Bibliografia

- [1] Davison, A.(2003). *Statistical Models*, Cambridge University Press, Cambridge.
- [2] Gasparin, Enrico (2013). *Il metodo HOTA per la simulazione da distribuzioni marginali a posteriori bayesiane*, Dipartimento di Scienze Statistiche, Università degli Studi di Padova.
- [3] Kotz, S., Lumelskii, Y. and Pensky, M. (2003). *The Stress-Strength Model and its Generalizations*, World Scientific, Singapore.
- [4] Liseo, B. (2008). *Introduzione alla Statistica Bayesiana*, dispense didattiche.
- [5] Ruli, E., Sartori, N. and Ventura, L. (2014). Marginal Posterior Simulation via Higher-order Tail Area Approximations, *Bayesian Analysis*, **9**, 129-146 .
- [6] Tierney, L. e Kadane, J.B. (1986) Accurate approximations for posterior moments and marginal densities, *Journal of the American Statistical Association*, 81, 82-86.
- [7] Ventura, L. e Racugno, W. (2011) Recent advances on Bayesian inference for $P(X<Y)$, *Bayesian Analysis*, 6, 411-428.