

# Le tecniche statistiche bayesiane e la loro applicazione in modelli di simulazione probabilistica

Sergio Iannazzo <sup>(1)</sup>



## ABSTRACT

Bayesian statistical methods are facing a rapidly growing level of interest and acceptance in the field of health economics. The reasons of this success are probably to be found on the theoretical fundaments of the discipline that make these techniques more appealing to decision analysis. To this point should be added the modern IT progress that has developed different flexible and powerful statistical software framework. Among them probably one of the most noticeably is the BUGS language project and its standalone application for MS Windows WinBUGS.

Scope of this paper is to introduce the subject and to show some interesting applications of WinBUGS in developing complex economical models based on Markov chains.

The advantages of this approach reside on the elegance of the code produced and in its capability to easily develop probabilistic simulations. Moreover an example of the integration of bayesian inference models in a Markov model is shown. This last feature let the analyst conduce statistical analyses on the available sources of evidence and exploit them directly as inputs in the economic model.

**Keywords:** bayesian statistic methods, probabilistic simulation, Markov model, WinBUGS  
*Farmeconomia e percorsi terapeutici 2007; 8(1): 5-13*

## INTRODUZIONE

Negli ultimi tempi le tecniche statistiche bayesiane, definite in contrapposizione alle tecniche tradizionali basate sull'approccio anche definito frequentista, hanno registrato un costante incremento di interesse e accettazione da parte della comunità scientifica internazionale coinvolta nel campo dell'economia sanitaria. Ad esempio il *National Institute for Clinical Excellence* (NICE) in Gran Bretagna ha esplicitamente inserito le tecniche basate sull'approccio bayesiano nelle linee guida per le analisi richieste agli sponsor. Ugualmente la *Food and Drug Administration* (FDA) in USA ha dimostrato una certa apertura. Le ragioni di questo recente entusiasmo sono da ricercarsi nei benefici intrinseci dell'approccio bayesiano, che Luce e O'Hagan [1] hanno ben sintetizzato e che possiamo qui riassumere sommariamente:

1. i metodi bayesiani sono in grado di utilizzare tutte le informazioni disponibili;
2. l'inferenza statistica basata sull'approccio bayesiano è più semplice e intuitiva di quella basata sull'approccio tradizionale;
3. per le due ragioni precedenti i metodi bayesiani risultano particolarmente adatti ai problemi decisionali.

A fronte di questi vantaggi esistono dei difetti che possono essere riassunti in:

1. i metodi bayesiani fanno uso di un elemento di soggettività che non è (almeno esplicitamente) presente nelle tecniche tradizionali;
2. i metodi bayesiani fanno uso di un elevato livello di sofisticazione matematica e, conseguentemente, richiedono una notevole potenza di calcolo.

Questo ultimo punto, in particolare, ha avuto un impatto decisivo nella storia dell'approccio statistico bayesiano. Il dibattito tra statistici frequentisti e bayesiani che infuria da lungo tempo (la prima pubblicazione del teorema di Bayes risale al 1763) è di fatto rimasto confinato alla teoria fino ai tardi anni '90. Fino a quel momento la teoria bayesiana è, infatti, rimasta priva di applicazione pratica a causa della sua complessità matematica, per cui risulta che solo i problemi più semplici possono essere risolti in forma analitica. Di converso la statistica frequentista e i suoi metodi hanno costituito la base e il motore con cui la scienza medica ha compiuto i suoi progressi attraverso la filosofia dell'*Evidence Based Medicine*. Verso la fine degli anni '90 gli sviluppi enormi dell'informatica moderna hanno portato alla disponibilità comune le elevate potenze di calcolo dei personal

<sup>(1)</sup>Centro di Ricerche Farmacoeconomiche, Advanced Research srl, Torino

computer, accelerando lo sviluppo di tutte le tecniche basate sul calcolo numerico. Di qui la nascita di vari software di calcolo statistico, di cui sicuramente il più importante è il linguaggio BUGS e la sua derivazione (per ambiente Microsoft) WinBUGS, ambiente sviluppato negli anni da una comunità di ricercatori, statistici ed informatici in Inghilterra. Obiettivo di questo testo è di fornire una sommaria introduzione all'argomento e mostrare alcune interessanti applicazioni di WinBUGS nello sviluppo di modelli farmaco-economici complessi basati su catene di Markov.

### APPROCCIO BAYESIANO E APPROCCIO FREQUENTISTA

Per comprendere a fondo i vantaggi e i difetti dell'approccio statistico bayesiano può essere utile fare un passo indietro e risalire all'oggetto del contendere, ossia alla definizione del concetto di probabilità.

Secondo la teoria frequentista il concetto di probabilità è associabile unicamente a fenomeni ripetibili il cui esito incerto è legato all'incertezza intrinseca del processo che li determina. Nella prospettiva bayesiana la probabilità invece descrive semplicemente l'incertezza con cui siamo in grado di prevedere l'esito del fenomeno in questione. Questa incertezza può derivare dall'intrinseca variabilità del processo, ma anche dalla nostra imperfetta conoscenza di esso. Luce e O'Hagan forniscono un esempio chiarificatore in merito [1?]. Consideriamo l'esempio classico del lancio della moneta. Secondo l'ottica frequentista la probabilità (evidentemente del 50%) che l'esito del lancio sia testa è da interpretarsi come la frequenza di esiti testa su un numero elevato, tendente ad infinito, di lanci. Secondo l'ottica bayesiana la probabilità del 50% sul lancio della moneta rappresenta l'incertezza con cui sono in grado di prevedere l'esito di un lancio. Consideriamo dunque due persone: A lancia la moneta e ne guarda l'esito, ma non lo comunica a B. Per B la probabilità che la moneta sia testa è del 50%. Questa affermazione, del tutto logica secondo il comune buon senso, è vera solo in ottica bayesiana, in quanto il processo non è intrinsecamente incerto (dato che il lancio è già avvenuto), ma quello che determina l'incertezza è la non conoscenza di B. In ottica frequentista, a rigore, non si può quindi parlare di probabilità legata all'esito di quel lancio, dato che è stato già effettuato. Il valore dell'esperimento è certo perché non si distingue tra il punto di vista di A e B e il rispettivo livello di conoscenza del fenomeno.

### IL TEOREMA DI BAYES

Il punto chiave dei metodi di inferenza statistica basati sull'approccio bayesiano è

il cosiddetto teorema di Bayes, dal nome del reverendo Thomas Bayes che nel 1763 introdusse il problema. L'idea base è quella di fare inferenza statistica, ossia sviluppare delle previsioni, su quantità non note in base all'evidenza disponibile e a un'ipotesi iniziale. Quest'ultima è la cosiddetta probabilità a priori, che costituisce l'elemento di soggettività delle analisi bayesiane criticato dagli statistici frequentisti.

Il teorema di Bayes permette di calcolare la probabilità di un evento E (detto evento condizionato) dato un evento A (detto evento condizionante). In termini matematici questo è espresso con la (1).

$$P(E|A) \tag{1}$$

Indicato  $P(E)$  la stima della probabilità dell'evento E sulla base delle informazioni iniziali, ossia la probabilità a priori, il teorema di Bayes ci dice che la probabilità a posteriori  $P(E|A)$  è data dalla (2).

$$P(E|A) = P(E) \cdot \frac{P(A|E)}{P(A)} \tag{2}$$

#### Esempio: eventi discreti

Un semplice esempio può aiutare a chiarire meglio il concetto. Supponiamo di avere una scatola chiusa in cui sono stati posti due gioielli. Questi gioielli possono essere d'oro o d'argento. Le possibili combinazioni possono essere dunque:

- oro-oro;
- oro-argento;
- argento-argento.

A priori senza aprire la scatola potremmo dire che queste tre combinazioni sono equiprobabili. Ossia possiamo scrivere la nostra semplice distribuzione di probabilità a priori (3).

$$\begin{aligned} P(o-o) &= \frac{1}{3} \\ P(o-a) &= \frac{1}{3} \\ P(a-a) &= \frac{1}{3} \end{aligned} \tag{3}$$

Supponiamo poi di estrarre un gioiello e che questo sia d'oro. A questo punto quale sarà la probabilità di ciascuna delle tre combinazioni data la nuova informazione ossia quale sarà la probabilità a posteriori dato l'evento condizionante? Il risultato è dato dall'applicazione della (2).

$$\begin{aligned}
 P(o-o|o) &= \frac{1}{3} \cdot \frac{1}{0,5} = \frac{2}{3} = 67\% \\
 P(o-a|o) &= \frac{1}{3} \cdot \frac{0,5}{0,5} = \frac{1}{3} = 33\% \\
 P(a-a|o) &= \frac{1}{3} \cdot \frac{0}{0,5} = 0
 \end{aligned}
 \tag{4}$$

In questo semplice esempio gli eventi considerati hanno valori discreti per cui anche le distribuzioni di probabilità sono discrete. Le cose si complicano quando si considerano eventi con valori continui.

#### Esempio: analisi “coniugata”

Supponiamo di considerare un trattamento farmacologico su pazienti affetti da una determinata patologia e di voler fare una previsione sul numero di pazienti che risponderà al trattamento. Immaginiamo di poter disporre di precedenti esperienze che ci suggeriscono che la frequenza di risposta avrà in media il valore 0,4 e che comunque i valori plausibili saranno in gran parte compresi nell'intervallo 0,17 – 0,67. L'esperienza ci suggerisce che la distribuzione beta ha le caratteristiche adatte per rappresentare questa informazione e, con un po' di tecnica statistica, si possono calcolare i parametri. La distribuzione  $beta[5,2, 7,5]$  rappresenta bene l'informazione a priori.

Immaginiamo ora di avere a disposizione un dato sperimentale in cui si è verificato che il trattamento ha dato risposta positiva in 22 casi su 30. Questa informazione può essere rappresentata da una distribuzione binomiale. La combinazione, mediante il teorema di Bayes, della distribuzione a priori con l'informazione sperimentale produce la distribuzione a posteriori.

È, in generale, possibile dimostrare che se la distribuzione a priori è di tipo  $beta[a,b]$  e se osserviamo  $r$  successi in  $n$  prove, la distribuzione a posteriori sarà anch'essa di tipo beta con parametri dati dalla (5).

$$\text{posteriori} = \text{beta}[a+r, b+n-r]
 \tag{5}$$

Nel nostro caso quindi la distribuzione a posteriori sarà di tipo  $beta[27,2, 15,5]$ .

Osservando le curve delle distribuzioni (Fig. 1 e 2) possiamo notare i risultati prodotti dall'analisi. L'aggiunta dell'evidenza sperimentale ha modificato in due modi il nostro stato di informazione:

- in primo luogo ritenevamo che la frequenza di risposta media fosse intorno a 0,4. L'evidenza sperimentale ha, invece, dato un valore più alto (22 risposte su 30 pazienti) per cui intuitivamente ci aspettiamo di rivedere al rialzo la stima. Questo accade nella distri-

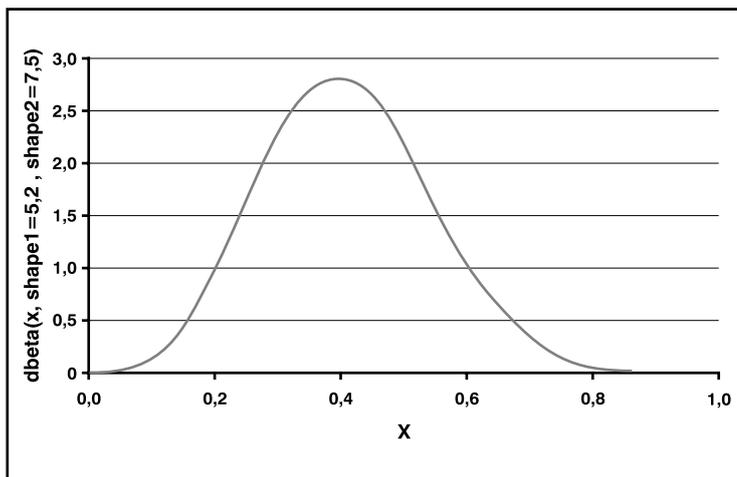


Figura 1  
Distribuzione a priori, beta [5,2, 7,5]

buzione a posteriori evidenziata in Figura 2, in cui la media è spostata oltre 0,6;

- in seconda battuta, ci aspettiamo che l'evidenza sperimentale abbia migliorato il nostro stato di informazione. Anche questo accade nella distribuzione a posteriori che risulta più stretta intorno al suo valore medio.

Utilizzando la distribuzione a posteriori possiamo ora fare delle previsioni. Se ad esempio vogliamo stimare la probabilità con cui il trattamento si dimostrerà efficace almeno nel 50% dei casi, dobbiamo valutare l'area sotto la curva di Figura 2 compresa tra 0,5 e 1, ossia dobbiamo integrare la funzione di densità di probabilità a posteriori tra detti limiti.

Purtroppo l'esempio visto costituisce un caso molto particolare, detto di analisi “coniugata”, in cui la distribuzione a priori e a posteriori appartengono alla stessa famiglia. In questi casi è infatti possibile ricavare la distribuzione a posteriori in forma analitica. In tutti gli altri casi è necessario procedere con metodi numerici.

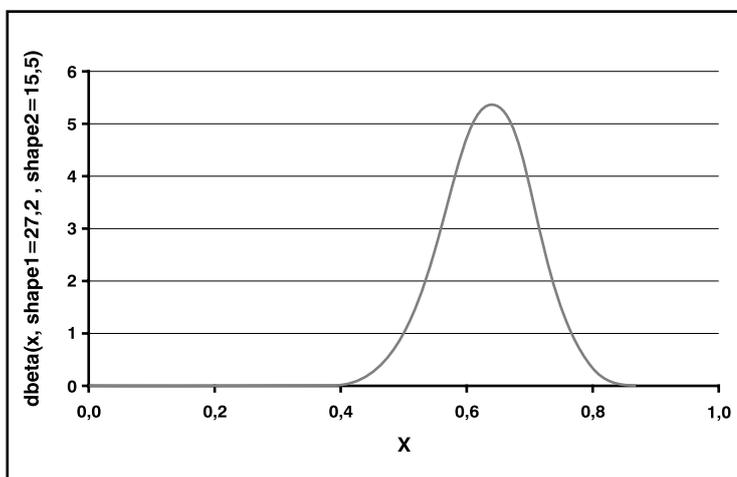


Figura 2  
Distribuzione a posteriori, beta [27,2, 15,5]

### IL PROGETTO BUGS E WINBUGS

Come già accennato il progetto BUGS rappresenta forse il più importante ambiente software disponibile per l'analisi e calcolo di modelli statistici bayesiani. Il progetto, iniziato nel 1989 al *MRC Biostatistics Unit*, Cambridge, UK, ha portato dapprima allo sviluppo di un linguaggio multi-piattaforma (oggi chiamato Classic BUGS) e successivamente, in collaborazione con l'*Imperial College School of Medicine*, Londra, UK, ad un applicativo indipendente per sistemi operativi Microsoft Windows (WinBUGS). WinBUGS è stato sviluppato fino alla versione 1.4.1 rilasciata nel settembre del 2004. Successivamente il programma è proseguito nella forma di sviluppo open source ed oggi il programma OpenBUGS è condotto dall'Università di Helsinki, Finlandia.

WinBUGS è un programma liberamente scaricabile dal sito <http://www.mrc-bsu.cam.ac.uk/bugs/winbugs/contents.shtml> e utilizzabile gratuitamente senza limitazione, previa registrazione.

### ANALISI BAYESIANA CON WINBUGS

WinBUGS è un programma di calcolo che implementa un suo specifico linguaggio. In linea di massima i passi attraverso cui si può eseguire un'analisi con WinBUGS sono sintetizzabili in:

- definizione del modello;
- verifica della correttezza formale del modello;
- compilazione e inizializzazione del modello;
- calcolo di un numero elevato di iterazioni;
- analisi dei risultati.

Una vera e propria introduzione all'utilizzo di WinBUGS esula dagli scopi di questo articolo, per chi fosse interessato a muovere i primi passi in questo campo o ad approfondire le proprie conoscenze, si suggerisce di sfruttare la documentazione e i tutorial pubblicati sul sito precedentemente riportato.

Per mostrare la potenza di questo ambiente di calcolo possiamo considerare l'esempio di un modello per prevedere la frequenza di risposta di un trattamento farmacologico.

Il modello statistico è scritto in poche righe.

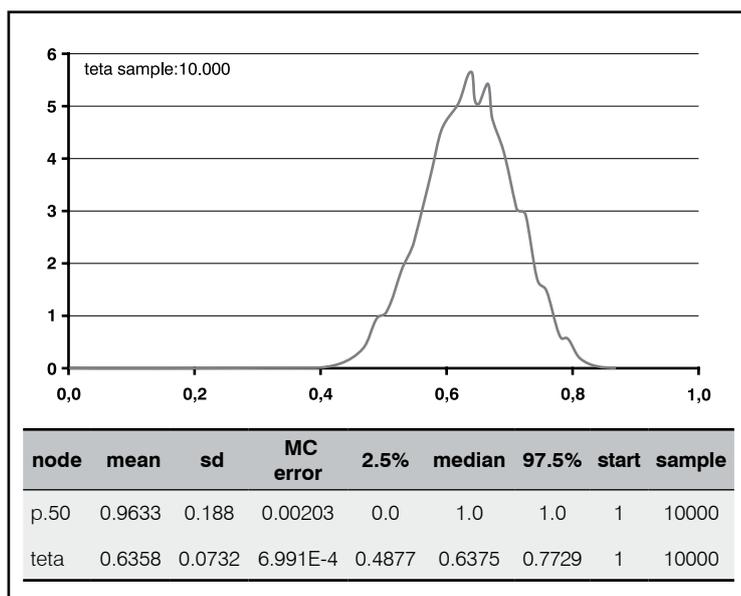
```
model {
  teta ~ dbeta(5.2,7.5)
  # distribuzione a priori
  y ~ dbin(teta,30)
  # successi su 30 pazienti
  y <- 22
  # evidenza sperimentale di 22 risposte su 30
  # pazienti
  p.50 <- step(teta-0.5)
  # probabilità trattamento efficace in almeno 50%
  # dei casi
}
```

Sommariamente il codice può essere descritto in questi termini:

- *teta ~ dbeta (5.2,7.5)*  
la prima riga stabilisce la distribuzione a priori, ossia definisce la grandezza stocastica *teta* che rappresenta la probabilità di successo del trattamento su un paziente, secondo lo stato di conoscenza iniziale (il simbolo  $\sim$  caratterizza un'assegnazione probabilistica);
- *y ~ dbin (teta,30)*  
si definisce una seconda variabile stocastica che rappresenta il numero di successi che si possono ottenere con il trattamento su 30 pazienti;
- *y <- 22*  
si imposta il dato emerso dall'evidenza sperimentale di 22 successi;
- *p.50 <- step (teta-0.5)*  
si definisce la variabile numerica che misura la probabilità che il trattamento sia efficace almeno nel 50% dei casi. La funzione *step()* restituisce 1 se il suo argomento è  $\geq 0$ , restituisce 0 altrimenti. Ad ogni iterazione del modello dunque la variabile *p.50* vale 1 o 0 se *teta* è maggiore o minore di 0,5.

Elaborando un numero elevato di iterazioni (10.000 nel nostro caso) si ottengono la curva di densità di *teta* e le statistiche di *teta* e *p.50*.

Si può notare come la distribuzione della probabilità che il trattamento dia risposta (*teta*) calcolata da WinBUGS sia quella a posteriori. Avendo inserito il dato sperimentale il programma ha, infatti, applicato il teorema di Bayes e calcolato la distribuzione a posteriori. La media



**Figura 3**  
Distribuzione a posteriori e statistiche sulle 10.000 iterazioni calcolate con WinBUGS

sulle 10.000 iterazioni della variabile  $p.50$  dà la probabilità che il trattamento possa risultare efficace almeno nel 50% dei casi.

Da questo, ancorché semplice, esempio riteniamo possano emergere tre grandi valenze del programma WinBUGS:

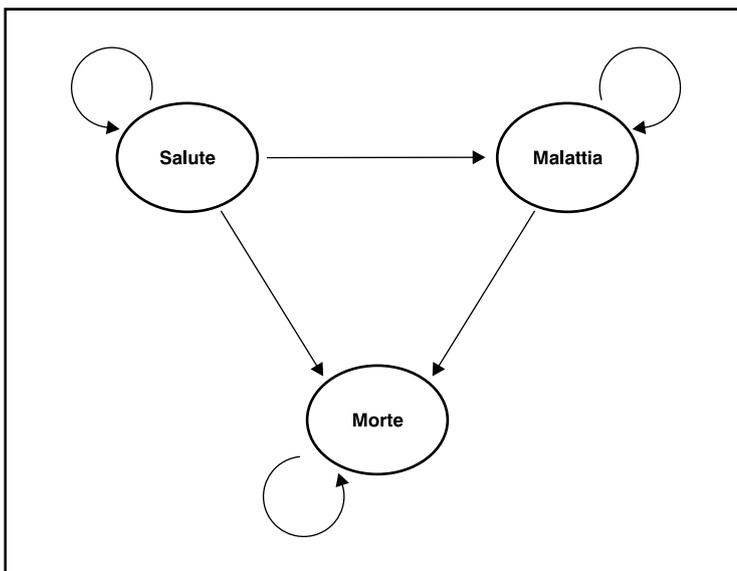
- calcolo bayesiano – l'incorporazione dell'evidenza sperimentale nel modello statistico alla luce del teorema di Bayes avviene in modo istantaneo e del tutto naturale;
- code di distribuzioni – abbiamo visto che la stima della probabilità che il trattamento dia risposta positiva in più del 50% dei casi, se ricavata analiticamente, deve essere fatta integrando la curva di densità di probabilità a posteriori da 0,5 a 1. WinBUGS permette di risparmiare questo calcolo;
- previsioni – WinBUGS permette di fare delle previsioni probabilistiche. Ad esempio la probabilità di risposta in più del 50% dei casi è del 96,33% con deviazione standard di 0,188.

## I MODELLI DI MARKOV IN WINBUGS

WinBUGS, al di là della sua indiscussa potenza di calcolo, dispone di un linguaggio flessibile che permette di definire modelli economici complessi come possono essere i modelli di Markov [5,6]. Un modello di Markov è caratterizzato da stati discreti e mutuamente esclusivi che descrivono la dinamica delle situazioni del problema in esame. Allo stesso modo il tempo è discretizzato in cicli di pari durata. Consideriamo ad esempio di voler modellizzare la dinamica di una patologia cronica. Per semplicità, consideriamo solo tre stati possibili: salute, malattia e morte. I soggetti simulati inizialmente sono nello stato di salute, successivamente una parte di essi sviluppa la patologia, che non è reversibile. Sia i soggetti nello stato di salute che quelli nello stato di malattia fronteggiano una determinata probabilità di morte nel tempo. Sempre per semplicità supponiamo che l'incidenza della patologia sia valutata al 10% ad ogni ciclo (di un anno) e che questa sia costante nel tempo. Inoltre supponiamo che la probabilità di morte per i soggetti sani sia costante all'1% e che la patologia causi un rischio relativo di morte di 2,5.

Codifichiamo il modello descritto in WinBUGS al fine di calcolare l'aspettativa di vita media di un soggetto. Questo non è altro che la trascodifica della notazione matriciale del calcolo di una catena di Markov nel linguaggio di WinBUGS.

```
model{
  # probabilità
  p.malattia <- 0.1
  p.morte.salute <- 0.01
  rr.malattia <- 2.5
  p.morte.malattia <- p.morte.salute * rr.malattia
```



**Figura 4**

Semplice schema di un modello di Markov per una patologia cronica irreversibile

```
#traccia del modello di Markov
for (t in 1:T){
  prob.trans[t,1,1] <- 1 - p.malattia - p.morte.salute
  prob.trans[t,1,2] <- p.malattia
  prob.trans[t,1,3] <- p.morte.salute
  prob.trans[t,2,1] <- 0
  prob.trans[t,2,2] <- 1 - p.morte.malattia
  prob.trans[t,2,3] <- p.morte.malattia
  prob.trans[t,3,1] <- 0
  prob.trans[t,3,2] <- 0
  prob.trans[t,3,3] <- 1
}
```

```
#distribuzione iniziale
pop[1,1] <- 1
pop[1,2] <- 0
pop[1,3] <- 0
```

```
#calcolo della popolazione
for (t in 2:T){
  for (s in 1:S){
    pop[t,s] <- inprod(pop[t-1,1:S], prob.trans[t-1,1:S, s])
  }
}
```

```
#anni di vita guadagnati per stato
anni.ciclo[1] <- 1
anni.ciclo[2] <- 1
anni.ciclo[3] <- 0
```

```
#calcolo degli anni vissuti ad ogni ciclo
anni[1] <- 0
for (t in 2:T){
  anni[t] <- inprod(pop[t-1,1:S], anni.ciclo[1:S])
}
```

```
#aspettativa di vita
anni.tot <- sum(anni[1:T])
}
```

```
#dati
list(
  T = 100,
  #numero di cicli
  S = 3,
  # numero di stati
)
```

Per chiarezza di lettura esplicitiamo qui alcune notazioni in uso nel linguaggio di WinBUGS; la struttura generale del codice dovrebbe comunque risultare chiara al lettore con un minimo di esperienza di programmazione:

- la parentesi quadra racchiude uno o più indici di un vettore (a tante dimensioni quanti sono gli indici). Ad esempio  $anni[t]$  è un vettore ad una dimensione mentre  $pop[t,s]$  è una matrice (vettore a due dimensioni);
- gli stati del modello, indicati dall'indice  $s$ , sono codificati nella seguente maniera: 1-salute, 2-malattia, 3-morte;
- la funzione  $sum()$  esegue la somma degli elementi di un vettore. Ad esempio  $sum(anni[1:T])$  fa la somma  $anni[1]+anni[2]+...+anni[T]$ ;
- la funzione  $inprod()$  esegue una somma di prodotti. Ad esempio  $inprod(pop[t-1,1:S],anni.ciclo[1:S])$  corrisponde a  $pop[t-1,1]*anni.ciclo[1]+pop[t-1,2]*anni.ciclo[2]+...+pop[t-1,S]*anni.ciclo[S]$ .

WinBUGS calcola istantaneamente il modello sopra descritto e fornisce il risultato desiderato: l'aspettativa di vita media di un soggetto è 41,62 anni.

Fino a questo punto il modello è puramente deterministico e, in definitiva, stiamo usando WinBUGS come un normale foglio di calcolo Excel. Vogliamo ora mostrare qualche interessante potenzialità dell'implementazione di modelli di Markov in WinBUGS, che permettono di superare alcuni limiti tradizionalmente imposti dalla tecnica "classica". Queste potenzialità possono essere riassunte nei seguenti argomenti:

- integrazione di variabili di sistema;
- simulazione probabilistica e analisi di sensibilità integrata;
- sintesi bayesiana di molteplici fonti di evidenza.

#### Variabili di sistema

Il linguaggio di programmazione di WinBUGS, completo e flessibile, ci permette di aggirare uno dei grossi limiti della tecnica di modellizzazione con catene di Markov. È noto, infatti, che nella definizione di un modello gli stati devono racchiudere tutta l'informazione rilevante per il problema. Così se, ad esempio, è importante tenere conto dell'età di un soggetto malato si dovranno definire molteplici stati di malattia, uno per ogni fascia d'età considerata. Ne consegue una proliferazione (spesso esplosiva) del numero di stati del modello. Per aggirare il problema è possibile definire alcune variabili di stato il cui valore viene aggiornato in funzione di eventi che accadono nella simulazione. Le probabilità di transizione del modello, i guadagni e costi specifici possono poi essere espressi in funzione di queste variabili di stato.

L'uso di variabili di sistema è una tecnica utile per elevare il grado di complessità di un modello di Markov senza appesantirne la logica fondamentale ed è ormai una possibilità offerta da tutti i sistemi di calcolo decisionale (come ad esempio il software TreeAge). È, inoltre, implementabile, pur con qualche complessità, anche in modelli sviluppati in Excel. Tuttavia riteniamo che l'implementazione in WinBUGS risulti particolarmente pulita ed elegante. Di seguito illustriamo come sia possibile, a partire dall'esempio di modello di Markov in WinBUGS usato in precedenza, tenere conto dell'età del soggetto simulato per la determinazione della sua probabilità di morte, in base alle tavole di mortalità della popolazione generale italiana [2].

Le tavole di mortalità esprimono, in funzione dell'età di un soggetto ad inizio anno, la sua probabilità di morte entro l'anno. Per integrare questo dato è sufficiente modificare il codice WinBUGS visto in precedenza in modo che la variabile  $p.morte.salute$  non abbia più il valore fisso di 0,01 ma sia determinata dinamicamente in funzione dell'età del soggetto simulato. Viene modificata solo la parte iniziale del codice:

```
# probabilità di morte in funzione dell'età
eta.iniziale <- 35
for (t in 1:T){
eta.corrente[t] <- eta.iniziale + t - 1
p.morte.salute[t] <- tavole.mort[eta.corrente[t]]
p.morte.malattia[t] <- p.morte.salute[t] * rr.malattia
}

# probabilità
p.malattia <- 0.1
rr.malattia <- 2.5

#traccia del modello di Markov
for (t in 1:T){
prob.trans[t,1,1] <- 1 - p.malattia - p.morte.salute[t]
prob.trans[t,1,2] <- p.malattia
prob.trans[t,1,3] <- p.morte.salute[t]
prob.trans[t,2,1] <- 0
prob.trans[t,2,2] <- 1 - p.morte.malattia[t]
prob.trans[t,2,3] <- p.morte.malattia[t]
prob.trans[t,3,1] <- 0
prob.trans[t,3,2] <- 0
prob.trans[t,3,3] <- 1
}
```

Il resto del modello rimane invariato. In questo modo ad ogni ciclo di Markov la probabilità di morte sarà estratta dalla tabella  $tavole.mort[età]$  (qui non riportata per brevità) ricavata dai dati ISTAT. Se l'età iniziale è 35 anni l'aspettativa di vita calcolata risulta di 36,99 anni.

#### Simulazione probabilistica

Anche con l'inserimento di variabili di sistema il nostro modello di Markov continua a rimanere puramente deterministico. WinBUGS ci permette, però, di sviluppare facilmente una simulazione probabilistica utilizzando una tecnica di tipo Monte Carlo. Sommarariamente, la simulazione probabilistica ci consente di tenere in considerazione due livelli di incertezza:

- l'incertezza conoscitiva relativa ai dati utilizzati;
- l'incertezza intrinseca legata all'accadere casuale di eventi.

Il primo livello di incertezza riguarda i dati che usiamo nel modello. Usare un valore fisso per un dato non ci consente, infatti, di considerare il livello di incertezza che lo caratterizza. Ad esempio supponiamo di aver ricavato da studi pubblicati precedentemente il valore 2,5 per la variabile *rr:malattia* (il rischio relativo di morte conferito dalla patologia rispetto alla popolazione generale di pari età). Se vogliamo sfruttare il dato che abbiamo a disposizione in modo completo dobbiamo considerare anche l'incertezza che lo caratterizza, ossia tipicamente la sua varianza. Per questo possiamo scegliere una distribuzione statistica che ragionevolmente rappresenti le caratteristiche del parametro (con la data media e varianza) e procedere ad una simulazione di tipo Monte Carlo. Questa viene eseguita ripetendo per un numero elevato di iterazioni il calcolo del modello ed estraendo ad ogni iterazione un valore del parametro dalla distribuzione. È evidente che questo ragionamento è esattamente identico a quello che sta alla base dell'analisi di sensibilità probabilistica (PSA) [3,4].

Il secondo livello di incertezza riguarda la casualità degli eventi. Nel modello di Markov deterministico si calcola sempre il valore atteso di eventi incerti. Se la probabilità di morte in un ciclo è 10%, si calcola che gli anni vissuti in quel ciclo da un paziente siano 0,9. Nella simulazione probabilistica (come nella realtà) invece il paziente vive oppure muore e l'evento di morte è rappresentato come un processo aleatorio con probabilità 0,1. Gli anni di vita vissuti da un paziente in quel ciclo sono quindi 1 oppure 0. Ripetendo la simulazione per un numero elevato di iterazioni la media degli anni vissuti in quel ciclo sarà (ovviamente) 0,9. La simulazione probabilistica tiene, dunque, conto della casualità degli eventi. Su un numero elevato di iterazioni i due approcci non differiscono sostanzialmente nel valore medio finale prodotto, ma piuttosto nella dispersione dei dati intorno ad esso. Tuttavia risultati differenti possono emergere se il modello non è lineare, cioè se il verificarsi di un evento ha un effetto su altri eventi e quindi può modificare il corso della simulazione.

Supponiamo, dunque, di trasformare il modello sviluppato fin qui in WinBUGS in modo da ottenere una simulazione probabilistica. Iniziamo con l'incertezza sui parametri. Supponiamo di voler rappresentare una popolazione con età  $35 \pm 3$  anni (media  $\pm$  deviazione standard) e che studi pubblicati in precedenza riportino un'incidenza della patologia di  $0,1 \pm 0,025$  e un rischio relativo di morte di  $2,5 \pm 0,95$ .

Nel codice le definizioni deterministiche dei parametri vengono sostituite da quelle probabilistiche.

```
# assegnazioni probabilistiche
eta.iniziale ~ dnorm(eta.iniziale.media, eta.iniziale.
tau)I(0,)
eta.iniziale.media <- 35
eta.iniziale.tau <- 1/pow(3,2)
p.malattia ~ dbeta(p.malattia.alfa, p.malattia.beta)
p.malattia.alfa <- 14.3
p.malattia.beta <- 128.7
rr.malattia ~ dnorm(rr.malattia.media, rr.malattia.
tau)I(0,)
rr.malattia.media <- 2.5
rr.malattia.tau <- 1/pow(0.95,2)

# probabilità di morte in funzione dell'età
for (t in 1:T){
eta.corrente[t] <- eta.iniziale + t - 1
p.morte.salute[t] <- tavole.mort[eta.corrente[t]]
x[t] <- p.morte.salute[t] * rr.malattia
p.morte.malattia[t] <- x[t] - (x[t] - 1) * step(x[t]-1)
}
```

Ora, nel modello, ad ogni iterazione l'età iniziale verrà estratta da una distribuzione normale, la probabilità di sviluppare la malattia da una distribuzione beta, mentre il rischio relativo di morte da una distribuzione normale. Come nella PSA la scelta di una distribuzione con cui rappresentare un parametro è in larga parte dettata dall'esperienza e dal buon senso dell'analista. Ad esempio la distribuzione beta costituisce quasi sempre la prima scelta per rappresentare parametri che, come la probabilità di un evento, hanno valori compresi tra 0 e 1. Qualche commento sulle notazioni probabilistiche è necessario:

- *eta.iniziale* ~ *dnorm(eta.iniziale.media, eta.iniziale.tau)I(0,)*  
l'età iniziale viene estratta da una distribuzione normale. In WinBUGS la distribuzione normale è definita con media e precisione ( $\tau=1/\text{sd}^2$ ). La funzione *I(0,)* serve per modificare la distribuzione per evitare che vengano estratti valori negativi (che non avrebbero senso);
- *p.malattia* ~ *dbeta(p.malattia.alfa, p.malattia.beta)*  
la probabilità di sviluppare la malattia è rappresentata con una distribuzione beta i cui parametri sono stati calcolati in modo che risulti media e sd pari a  $0,1 \pm 0,025$ ;
- *p.morte.malattia[t] <- x[t] - (x[t] - 1) \* step(x[t]-1)*  
questa funzione serve a garantire che in ogni caso la probabilità di morte per i soggetti malati non sia maggiore di 1.

Il secondo livello riguarda l'incertezza intrinseca sugli eventi casuali, che nel nostro caso sono l'insorgenza della malattia e la morte. Per tenere in considerazione questi aspetti il modello deve essere modificato nella parte di calcolo della popolazione ad ogni istante.

```
#calcolo della popolazione
for (t in 2:T){
for (s in 1:S){
y[t,s] <- inprod(pop[t-1,1:S], prob.trans[t-1,1:S, s])
}
pop[t,1:S] ~ dmulti(y[t,1:S],1)
}
```

node	mean	sd	MC error	2.5%	median	97.5%	start	sample
anni.tot	38.33	11.99	0.1327	12.0	39.0	59.0	1	10000
eta.iniziale	34.94	3.014	0.03299	29.12	34.92	40.89	1	10000
p.malattia	0.1001	0.02499	2.574E-4	0.05678	0.09843	0.1545	1	10000
rr.malattia	2.518	0.9282	0.007991	0.7338	2.512	4.347	1	10000

**Tabella I**

Risultati del calcolo di 10.000 iterazioni

node	mean	sd	MC error	2.5%	median	97.5%	start	sample
anni.tot	38.35	5.126	0.05489	30.69	37.7	50.3	1	10000

**Tabella II**

Risultato di aspettativa di vita del nostro modello ottenuto senza considerare il secondo livello di incertezza

Il vettore  $y[t,s]$  rappresenta il valore atteso della popolazione ad ogni istante, dati la popolazione all'istante precedente e le probabilità di transizione. La popolazione sugli stati è, però, estratta da una distribuzione multinomiale che ha come probabilità i valori di  $y[t,s]$ . La distribuzione multinomiale è una generalizzazione della distribuzione binomiale per rappresentare un processo aleatorio con più di due esiti possibili.

Con il calcolo di 10.000 iterazioni si ottengono risultati illustrati in Tabella I.

Il risultato del modello (nel nostro semplice esempio è l'aspettativa di vita) consente di fare una previsione probabilistica: la stima è di 38,33 anni e con il 95% di confidenza sarà compreso nell'intervallo da 12 a 59 anni.

Va segnalato che l'implementazione della gestione dell'incertezza legata alla casualità degli eventi deve essere considerata con una certa cautela ed esclusivamente se ve ne è una reale necessità di analisi. È intuitivo, infatti, che la stabilità del modello ne venga influenzata. A titolo di esempio riportiamo in Tabella II il risultato di aspettativa di vita del nostro modello ottenuto senza considerare il secondo livello di incertezza.

Il valore medio risulta sostanzialmente identico, mentre la variazione intorno ad esso è dimezzata. In questo caso sappiamo che la variabilità del risultato è esclusivamente dovuta all'incertezza dei dati in ingresso. Se invece consideriamo anche il secondo livello di incertezza questa informazione risulta "coperta" dalla variazione legata all'incertezza intrinseca degli eventi.

Abbiamo già evidenziato il fatto che il primo livello di gestione dell'incertezza in un modello di Markov implementato in WinBUGS coincide con l'analisi di sensibilità probabilistica del modello. Uno dei vantaggi dell'implementazione di un modello di Markov in WinBUGS consiste proprio in questo: costruendo un modello pro-

babilistico l'analisi di sensibilità è integrata in esso e non deve essere sviluppata come analisi successiva rispetto ad un caso base.

#### Sintesi bayesiana di fonti di evidenza

L'ultimo aspetto che intendiamo introdurre è la possibilità di integrare l'analisi statistica bayesiana direttamente all'interno di un modello di Markov. Spesso i dati utilizzati per alimentare modelli di Markov derivano da molteplici fonti di letteratura o di evidenza sperimentale e, altrettanto spesso, è necessario farne una modellizzazione statistica preliminare. Come già visto, WinBUGS consente di sviluppare modelli statistici bayesiani che, attraverso la combinazione di un giudizio a priori con un'evidenza sperimentale, permettono di formulare previsioni sul comportamento del fenomeno in studio. Se volessimo utilizzare un modello statistico di questo tipo per alimentare con dati di input un modello di Markov tradizionale (ad esempio con MS Excel) saremmo costretti a considerare solo il valore medio della distribuzione a posteriori e, eventualmente, a rappresentarne la variabilità con la scelta di una distribuzione teorica per una successiva PSA. WinBUGS ci consente di scrivere il modello statistico direttamente all'interno del modello di Markov. Avendo poi l'accortezza di scartare un certo numero di iterazioni iniziali con cui il modello statistico raggiunge la convergenza, potremo ottenere una simulazione probabilistica in cui ad ogni iterazione il dato viene estratto direttamente dalla sua distribuzione a posteriori (senza che sia necessario attribuire a questa una forma analitica).

#### CONCLUSIONI

Le tecniche statistiche bayesiane stanno vivendo un momento di grande popolarità in ambito di analisi decisionali e di valutazioni economiche legate a problematiche di tipo sanitario. Le ragioni di questo successo sono da

ricercarsi nei fondamenti teorici della disciplina, che rendono queste tecniche particolarmente adatte a dette analisi. Gli sviluppi moderni dell'informatica e del calcolo numerico hanno del resto messo a disposizione strumenti di analisi e calcolo estremamente potenti e flessibili. Come esempio di ciò abbiamo visto come sia possibile utilizzare WinBUGS per sviluppare

modelli economici complessi basati su catene di Markov e come i vantaggi di questo approccio siano nell'eleganza e sintesi del codice prodotto, nella capacità di sviluppare simulazioni probabilistiche che integrano la PSA e nella possibilità di integrare in esso modelli statistici bayesiani per l'interpretazione e sintesi dei dati disponibili da varie fonti.

## BIBLIOGRAFIA

1. O'Hagan A, Luce BR. A Primer on Bayesian Statistics in Health Economics and Outcomes Research. 2003 MEDTAP International Inc.
2. ISTAT. Annuario statistico italiano 2006
3. Briggs AH, Gray AM. Methods in health care service research – handling uncertainty in economic evaluations of healthcare interventions. *Br Med J* 1999; 319: 635-638
4. Doubilet P, Begg CB, Weinstein MC, Braun P, McNeil BJ. Probabilistic sensitivity analysis using Monte Carlo simulation – a practical approach. *Med Decis Making* 1985; 5: 157-177
5. Briggs AH, Ades AE, Price MJ. Probabilistic sensitivity analysis for decision trees with multiple branches: use of the Dirichlet distribution in a Bayesian framework. *Med Decis Making* 2003; 23: 341-350
6. Spiegelhalter DJ, Best NG. Bayesian approaches to multiple sources of evidence and uncertainty in complex cost-effectiveness modelling. *Stat Med* 2003; 22: 3687-3709.