Stima e Filtraggio, Tesina 2

Davide Ghion, Matr. 1225841

Padova, 19 Giugno, 2020

1 Introduzione

1.1 Descrizione del modello

La tesina è stata basata su un modello epidemiologico che descrive l'evoluzione di una malattia in una popolazione in un determinato arco temporale. Il modello in questione è detto Suscettibili-Infetti-Recuperati (SIR). Per semplificare lo studio si è supposto che tutti i pazienti, dopo essere stati infettati, guariscano e non siano più suscettibili all'infezione, oppure muoiano. Dunque questo tipo di modello può essere applicato solo per le malattie che, determinando un'immunità permanente, non possono recidivare, come per esempio morbillo e varicella.
Il modello SIR è rappresentabile tramite tre funzioni non lineari che mettono in relazione il numero di persone sane o suscettibili S, infette I e guarite, o recuperate, R. Inoltre sono presenti due coefficienti b e k che descrivono rispettivamente il tasso di infezione e di guarigione

$$\begin{cases} \dot{S} = -b \cdot S \cdot I \\ \dot{I} = b \cdot S \cdot I - k \cdot I \\ \dot{R} = k \cdot I \end{cases} \tag{1}$$

1.2 Generazione delle misure

Prima quindi di poter applicare qualsiasi tipo di algoritmo di tracking, il sistema è stato inizializzato con delle condizioni iniziali: si è considerata una popolazione di 100 individui, ci cui inizialmente almeno il 99% è sano e si è studiato l'andamento dell'infezione in un anno.

$$
\begin{cases}
S_0 = 99 \\
I_0 = 1 \\
R_0 = 0 \\
\theta = [\, b \quad k \,]^T = [\, 0.005 \quad 0.01 \,]^T
\end{cases}
\tag{2}
$$

Una volta quindi scelti i valori di θ si è fatto evolvere il modello:

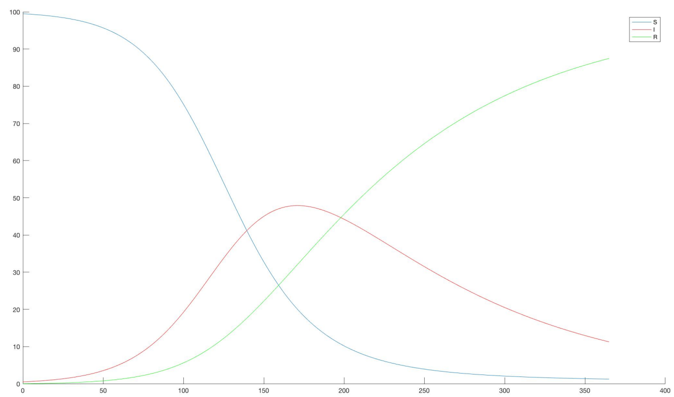

Figura 1: Evoluzione modello (1), con condizioni iniziali (2)

Conforme alle richieste della consegna le misure sono state sporcate da un rumore di misura ω, per semplicità si è scelto che il rumore si distribuisca secondo una normale di media nulla e deviazione standard $\sigma_\omega = 5$, di seguito viene mostrato il confronto tra le misure rumorose e quelle ”pulite”:

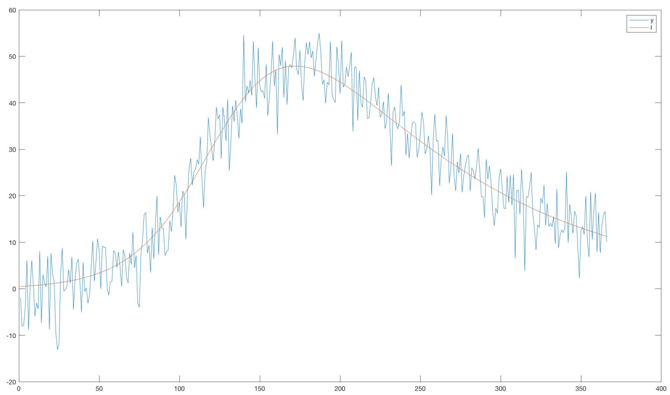

Figura 2: $y(t) = I(t) + \omega$

2 Prima parte

Lo scopo della prima parte è quello di stimare il valore del parametro $\theta = [\,b \;\; k\,]^T$ avendo a disposizione l'andamento degli infetti, il quale è affetto da rumore di misura ω, cioè $y(t) = I(t) + \omega$, in altre parole si è interessati a calcolare

$$E_\pi(g) = \int g(x)\pi(x)dx \tag{3}$$

che grazie alla approssimazione Monte Carlo dell'integrale si può esprimere come $E_\pi(g) \approx \frac{1}{N}\sum_{i=1}^{N} g(x_i)$. Si è scelto di usare l'algoritmo **Metropolis-Hastings** che permette di implementare una catena di Markov di *m* elementi con densità invariante pari alla posterior fornendo campioni correlati ma non indipendenti.
L'algoritmo è stato strutturato come segue:

1. la prima cosa di cui si ha bisogno è identificare una proposal density $q(\cdot|\cdot)$ che permette di "pescare" dei nuovi candidati, si è scelto di usare il metodo *random walk* cioè $q(\cdot|\cdot) = N(0, \Sigma^2)$, con $\Sigma = [\sigma_{prior_b} \;\; \sigma_{prior_k}]^T = [10^{-5} \;\; 5 \cdot 10^{-4}]^T$

2. si può dimostrare che la posterior puo? essere nota a meno di una costante di normalizzazione, cioè $\pi(\cdot) \propto likelihood \cdot prior$, per l'informazione a priori (la prior) di θ si è considerata che si comporti come una normale di media nulla e varianza $\sigma^2 = 3$.
   Rimane allora di calcolare la likelihood, si è scelto di usare la log-likelihood

l_x per evitare problemi dovuti all'aritmetica finita e per efficienza numerica, cioè si è implementata la formula

$$l_x = -\frac{N}{2}ln(2\pi) - \frac{N}{2}ln(\sigma_w^2) - \frac{1}{2}\sum_{i=1}^{T=365}(\frac{\hat{x}-y}{\sigma_w})^2 \tag{4}$$

Dunque il calcolo della posterior diventa

$$\pi(x) \propto l_x + log(prior)$$

Questo viene usato sia per il candidato *c* che per la misura di partenza *x*.

3. grazie alla scelta della proposal tramite *random walk*, la probabilità di accettare il nuovo candidato si semplifica diventando

$$\alpha(c,x) = min(1, \frac{\pi(c)}{\pi(x)}) = min(1, \exp^{\pi(c)-\pi(x)}) \tag{5}$$

In questo modo, considerata una variabile aleatoria uniforme *u* in [0,1], decido di accettare il nuovo candidato se

$$u \leqslant \alpha(c,x)$$

quindi aggiorna la catena con il nuovo candidato o rimanendo alla posizione precedente.

Si propone ora un esempio di come funziona l'algoritmo, sono state scelte $m = 1000$ iterazioni, il target è stato settato $\theta = [\,5 \cdot 10^{-4} \;\; 10^{-2}\,]^T$ e si è partiti con un guess iniziale nullo $\hat{\theta} = [\,0 \;\; 0\,]^T$

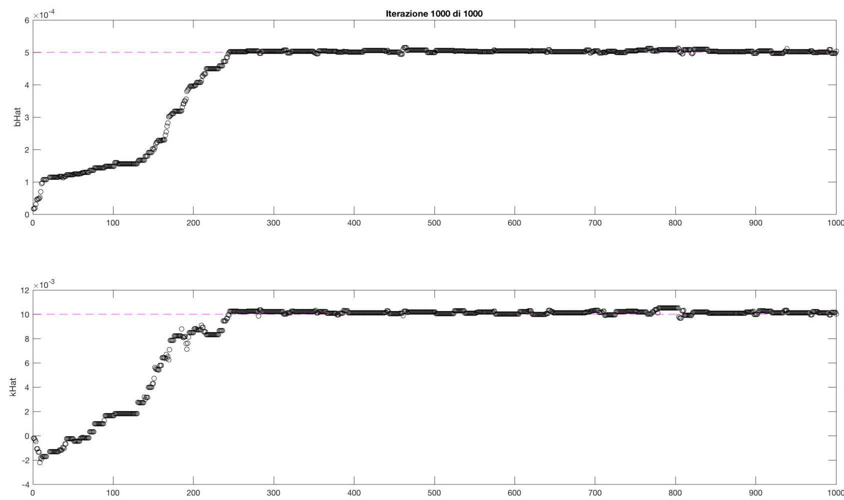

Figura 3: Convergenza di $\hat{\theta}$ tramite M-H

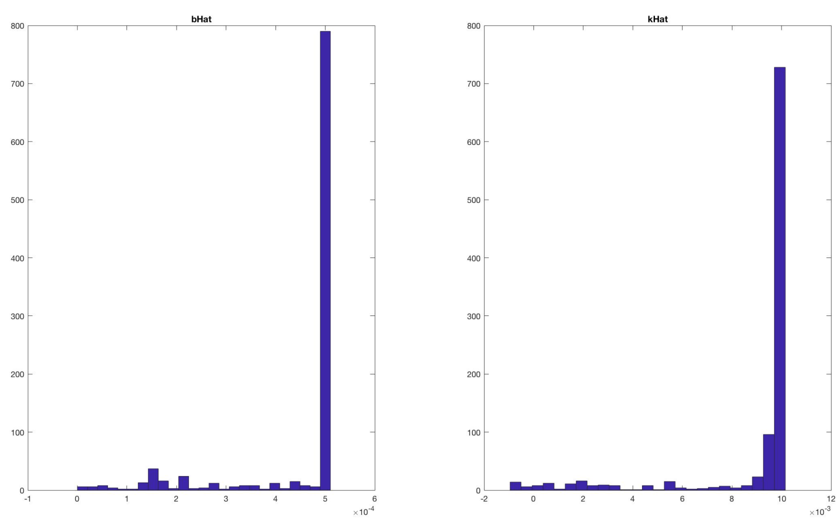

Figura 4: Istogramma delle marginali proposte da M-H

3 Seconda parte

In questa parte quello che si vuole fare è andare a stimare(da cambiare?), tramite Particle Filter, gli andamenti dei parametri S,I,R avendo a disposizione le misure degli infetti nel corso del tempo.

3.1 Discretizzazione del modello

Considerando il modello della consegna in forma di stato discreto, si è reso necessario adattare (1) discretizzandolo. Dunque come tipologia di discretizzazione è stata scelta *Eulero in avanti* anche se avrebbe potuto presentare problemi di instabilità

$$y_{n+1} = y_n + T \cdot h(t_n, y_n) \quad , \quad T \in \mathbb{N} \tag{6}$$

Dunque il modello (1) discretizzato diventa

$$\begin{cases} S_{t+1} = S_t + T \cdot (-b \cdot S_t I_t) \\ I_{t+1} = I_t + T \cdot (b \cdot S_t I_t - k \cdot I_t) \\ R_{t+1} = R_t + T \cdot (k \cdot I_t) \end{cases} \tag{7}$$

3.2 Setup parametri e condizioni iniziali

Come descritto nella sottosezione **(1.2)** per prima cosa si sono rese disponibili tutte le misure, da notare però che è stato usato il modello discretizzato (7), inoltre è stato scelto come passo di campionamento $T = 1$.

3.3 Particle Filter

L'obiettivo proposto è stato quello di implementare un particle filter che possa approssimare la pdf posterior ($\pi(x)$) degli stati S,I,R tramite misure casuali rappresentate da campioni random (particelle) con pesi associati:

$$p(x_k|y_{0:k}) \approx \sum_{i=1}^{N_s} w_k^i \delta(x_k - x_k^i) \tag{8}$$

È fondamentale allora capire come scegliere e aggiornare le particelle ed i loro pesi $\{x_{0:k}^i, w_k^i\}_{i=1}^{N_s}$, con N_s il numero di particelle.

Importance Sampling Dato che non è possibile estrarre le particelle x_k^i dalla $\pi(x)$ allora si identifica una proposal density $q(x)$ da cui è possibile farlo, inoltre considero un'ulteriore pdf $p(x)$ dalla quale è difficile estrarre i campioni ma per cui si può calcolare $\pi(x)$, allora la posterior vale

$$p(x) \approx \sum_{i=1}^{N_s} w^i \delta(x - x^i) \quad , \quad w^i \propto \frac{\pi(x^i)}{q(x^i)} \tag{9}$$

In particolare si è scelto di usare il metodo del *random walk* anche per questa parte e quindi si è considerata la proposal come una normale di media e varianza scelte nel momento dell'esecuzione del codice. Considerando le condizioni iniziali, durante i vari test, si è scelto di partire da delle distribuzioni abbastanza precise ed accurate.

Sequential Importance Resampling Definito il numero di particelle N_s allora si può procedere con l'algoritmo di predizione e aggiornamento degli stati della posterior:

1. si procede con l'estrazione delle particelle dalla mia proposal e vengono fatte evolvere secondo il modello (7), quindi ho l'approssimazione al tempo t degli stati S,I,R

2. assegno alle particelle i pesi secondo la formula

$$w_k^i \propto w_{k-1}^i \frac{p(y_k|x_k^i)p(x_k|x_{k-1}^i)}{q(x_k|x_{k-1}^i, y_k)} \overset{q(x_k|x_{k-1}^i, y_k) = p(x_k|x_{k-1}^i)}{=} w_{k-1}^i p(y_k|x_k^i)$$

3. i pesi vengono normalizzati in modo che la loro somma sia sempre unitaria: $\sum_i w_k^i = 1$

4. viene eseguita la stima dello stato tramite media pesata delle particelle

$$P_{tot} = \sum_{j=1}^{N_s} w_j^i \quad , \quad \hat{x}^i = \frac{1}{P_{tot}} \sum_{s=1}^{N_s} x_s^i w_s^i$$

5. viene fatto il resample delle particelle per evitare effetti di degenerazione, si è optato per un algoritmo che utilizza la somma cumulativa dei pesi e verifica l'intervallo in cui si trova il campione recuperandone gli indici

Viene ora mostrato un esempio dell'applicazione dell'algoritmo appena discusso, si è scelto $N_s = 1000$, le condizioni iniziali scelte sono:

$$\begin{cases} S_0 = 99 \\ I_0 = 1 \\ R_0 = 0 \\ prior \sim N(\begin{bmatrix} 90 \\ 5 \\ 5 \end{bmatrix}, \begin{bmatrix} 0.5 \\ 1 \\ 0.5 \end{bmatrix}) \end{cases}$$

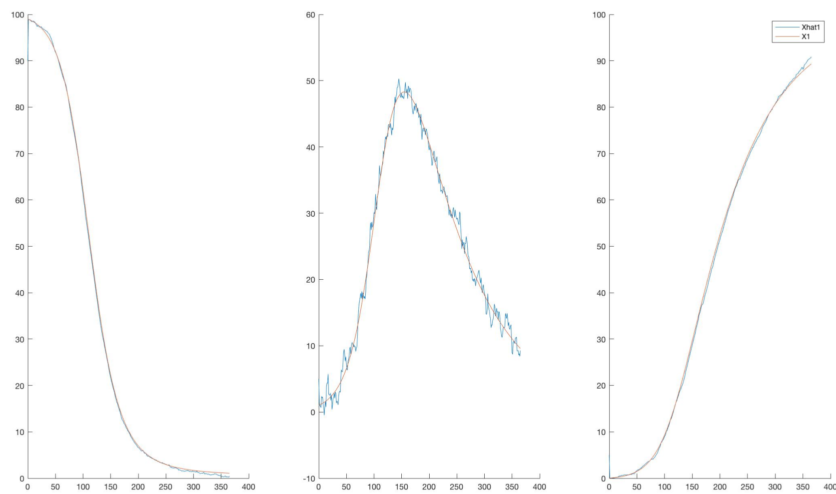

Figura 5: Stima stocastica del PF per lo stato composto dai parametri S,I,R

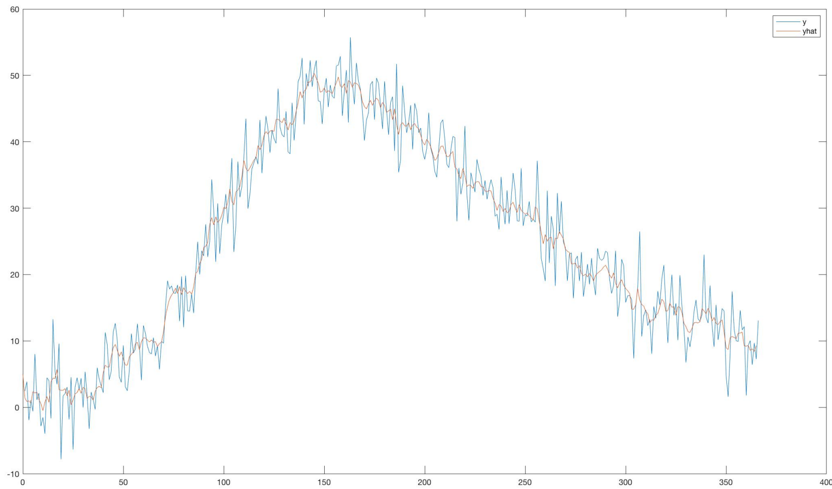

Figura 6: Focus dell'andamento degli infetti stimati dal PF

4 Riferimenti

1. A discrete SIR infectious disease model,
https://mathinsight.org/discrete_sir_infectious_disease_model#:~:text=The\%20model\%20contains\%20three\%20state,have\%20nonnegative\%20numbers\%20of\%20individuals)
.

2. Forecasting seasonal influenza with a state-space SIR model,
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5623938

3. Comments on the resampling step in particle filtering,
https://pingpong.chalmers.se/public/pp/public_noticeboard_attachment/fetch?messageId=35447&fileId=827809