

Simulazione di una pandemia con un modello SMRV

Giulio Pastorello, Federico Gnudi

Novembre 2023

1 Il Modello

Per descrivere l'andamento della pandemia si è scelto di usare un modello SIR al quale è stata aggiunta una quarta equazione per rappresentare i vaccinati.

Le equazioni che descrivono l'evoluzione della pandemia sono quindi:

- $S(t)$: "Susceptible". Descrive il numero delle persone non malate che non hanno mai contratto l'infezione.
- $I(t)$: "Infected". Il numero di persone che ad un dato tempo t sono malate.
- $R(t)$: "Removed". Una volta che una persona è stata contagiata, dopo un certo intervallo di tempo viene *rimossa*, ovvero guarisce o muore. Si assume che, una volta guariti, i soggetti diventino immuni. L'unica possibilità di transizione da uno stato all'altro è quindi: $S \rightarrow I \rightarrow R$.
- $V(t)$: "Vaccinated". Questa equazione è un'aggiunta al modello SIR "standard", che prevede solo le prime tre equazioni. Anche in questo caso si fa l'assunzione che una persona vaccinata non possa più ammalarsi.

Il vincolo fondamentale del modello è che la popolazione totale sia costante:

$$S(t) + I(t) + R(t) + V(t) = N \quad (1)$$

Questa è un'approssimazione ragionevole se si fa l'assunzione che il numero totale di persone che muoiono a causa del virus sia trascurabile rispetto alla popolazione totale.

1.1 Parametri

Il modello descrive l'andamento della pandemia attraverso alcuni parametri:

- $\beta \in [0, 1]$: questo è interpretabile come la *contagiosità* del virus, ovvero la probabilità di trasmissione a seguito di un contatto tra un suscettibile e un infetto.
- $\gamma \in [0, 1]$: è l'inverso della durata media di un'infezione, che corrisponde alla probabilità di guarigione o morte.
- $\eta \in \mathbb{N}, \eta < N$: il numero di persone non vaccinabili, per motivi medici o personali.
- $\mu \in [0, 1]$: rappresenta la velocità delle vaccinazioni.
- $\xi \in [0, 1]$: rappresenta l'efficacia delle vaccinazioni. Si usa come fattore moltiplicativo.

1.2 Condizioni Iniziali

Per studiare la diffusione del virus vengono anche usati come condizioni iniziali i seguenti termini:

- $I_0 \in \mathbb{N}, I_0 < N$: malati al tempo $t = 0$.
- $V_0 \in \mathbb{N}, V_0 < N$: vaccinati al tempo $t = 0$.
- $R_0 \in \mathbb{N}, R_0 < N$: rimossi al tempo $t = 0$.

Ovviamente si avrà che il numero di suscettibili dall'inizio della simulazione ($t = 0$) S_0 sarà:

$$S_0 = N - I_0 - V_0 - R_0$$

1.3 Equazioni

Usando i parametri descritti sopra, le equazioni differenziali che descrivono la diffusione della pandemia sono quindi quattro:

$$\frac{dS}{dt}(t) = -\beta \frac{S(t)}{N} I(t) - \frac{dV}{dt}(t) \quad (2)$$

$$\frac{dI}{dt}(t) = \beta \frac{S(t)}{N} I(t) - \gamma I(t) \quad (3)$$

$$\frac{dR}{dt}(t) = \gamma I(t) \quad (4)$$

$$\frac{dV}{dt}(t) = \mu \frac{V(t)}{\xi} \left(1 - \frac{V(t)}{\xi(N - \eta)} \right) \quad (5)$$

Una simulazione basata su dati rappresentativi dell'Emilia-Romagna è mostrata in figura (1). Si è usato: popolazione totale $N = 4459000$, $\beta = 0.05$, $\gamma = 0.02$, $I_0 = 530502$, $R_0 = 320000$, $V_0 = 600437$, non vaccinabili $\eta = 251042$, velocità vaccinazioni $\mu = 0.05$, efficacia vaccino $\xi = 0.83$.

1.4 Curva Logistica

Per rappresentare la progressione della campagna vaccinale, nell'equazione (5) si è scelto di usare un'equazione logistica generalizzata: $f'(x) = f(x)(1 - f(x))$, che ha come soluzione un sigmoide. Quest'approssimazione è ragionevole perché ci si aspettava un inizio più lento dovuto alla scarsità di vaccini e all'organizzazione non ancora roduta, seguito da una crescita veloce della quantità di persone vaccinate, seguito inevitabilmente da una diminuzione dovuta al fatto che le persone da vaccinare sono sempre meno. Il limite dell'equazione logistica è che usando come numero di vaccinati iniziali $V_0 = 0$ la campagna vaccinale resta a zero, ma partendo da un qualsiasi altro numero si arriverà sicuramente a vaccinare tutti i vaccinabili, in un tempo che dipende dalla velocità delle vaccinazioni μ . In figura (2) un andamento possibile della curva dei vaccinati.

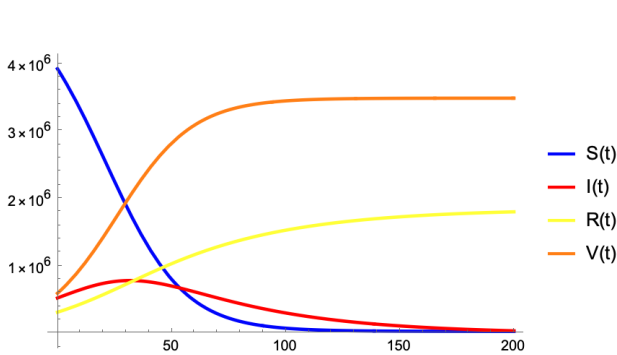


Figure 1: *Comportamento del modello SIRV con dati rappresentativi dell'Emilia-Romagna*

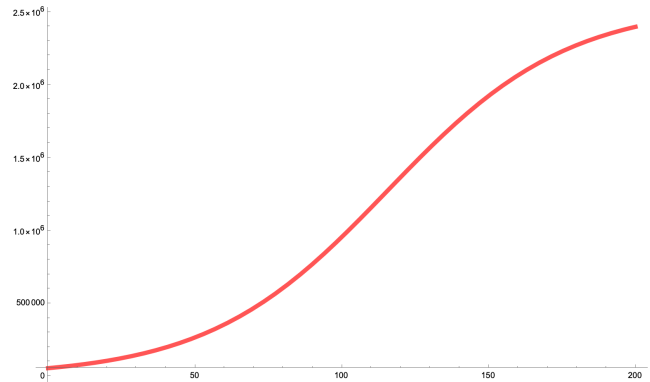


Figure 2: *Curva logistica con $\nu = 56980$, $\mu = 0.0274216$, $\eta = 1418957$, e popolazione totale della regione Emilia-Romagna.*

2 I Metodi Numerici

Per calcolare i risultati prodotti dal modello si sono scritti due diversi metodi numerici: il *metodo di Eulero*, meno preciso ma anche meno dispendioso a livello di calcoli, e il metodo *Runge-Kutta 4*, che è più complesso ma produce risultati migliori.

2.1 Il metodo di Eulero

Il metodo di Eulero è eseguito dalla funzione `evolve()` in `infection.cpp`. Dato un problema di Cauchy nella forma

$$\dot{\vec{x}} = f(t, \vec{x}), \quad \vec{x}(t_0) = \vec{x}_0 \quad (6)$$

con $f(t, \vec{x})$ campo vettoriale che descrive il sistema di equazioni, $\vec{x} \in \mathbb{R}^n$ variabile di stato, t variabile indipendente (tempo), e $\vec{x}(t_0) = \vec{x}_0$ condizione iniziale, il *metodo di Eulero* discretizza il sistema nella seguente maniera:

$$\vec{x}(t_{n+1}) = \vec{x}(t_n) + hf(t_n, \vec{x}(t_n)) \quad (7)$$

dove h è lo *step size*, un intervallo di tempo arbitrario che ovviamente determina la precisione dei risultati ottenuti. Nel caso dell'*algoritmo di Eulero*, si dimostra che l'errore totale è dell'ordine di $O(h^2)$.

Nel caso del modello SMRV, il sistema è *autonomo*, ovvero le equazioni non dipendono esplicitamente dal tempo. L'algoritmo di discretizzazione assume quindi la forma:

$$\vec{x}(t_{n+1}) = \vec{x}(t_n) + hf(\vec{x}(t_n)) \quad (8)$$

Le equazioni risultanti sono i valori che stampiamo:

$$\begin{cases} S(t_{n+1}) = S(t_n) + h \left(-\beta \frac{S(t_n)}{N} I(t_n) - \mu \frac{V(t_n)}{\xi} \left(1 - \frac{V(t_n)}{\xi(N-\eta)} \right) \right) \\ M(t_{n+1}) = M(t_n) + h \left(\beta \frac{S(t_n)}{N} I(t_n) - \gamma M(t_n) \right) \\ R(t_{n+1}) = R(t_n) + h\gamma M(t_n) \\ V(t_{n+1}) = V(t_n) + h\mu \frac{V(t_n)}{\xi} \left(1 - \frac{V(t_n)}{\xi(N-\eta)} \right) \end{cases} \quad (9)$$

Sfruttando il vincolo della popolazione totale costante espresso dall'eq. (1), si può "rimuovere" una delle quattro equazioni esprimendone il valore in funzione delle altre tre. Il risultato che otteniamo è il seguente:

$$\begin{cases} V(t_{n+1}) = V(t_n) + h\mu \frac{V(t_n)}{\xi} \left(1 - \frac{V(t_n)}{\xi(N-\eta)} \right) \\ S(t_{n+1}) = S(t_n) + h \left(-\beta \frac{S(t_n)}{N} I(t_n) - \Delta V \right) \\ R(t_{n+1}) = R(t_n) + h\gamma M(t_n) \\ M(t_{n+1}) = M(t_n) + h(-\Delta V - \Delta S - \Delta R) \end{cases} \quad (10)$$

dove si è usata la notazione $\Delta X = X(t_{n+1}) - X(t_n)$. Questo permette di ridurre il numero di valori da calcolare per ogni iterazione da 4 a 3.

2.2 Runge-Kutta 4th Order

L'algoritmo alternativo usato per calcolare i valori delle variabili di stato giorno-per-giorno è il Runge-Kutta del quarto ordine, abbreviato spesso come RK o RK4, che è eseguito nella funzione `rk()` in `rk.cpp`. Come da nome, l'errore totale è dell'ordine di $O(h^4)$. Partendo da un'equazione generica nella forma dell'eq. (6) l'approssimazione che viene fatta è la seguente:

$$\vec{x}(t_{n+1}) = \vec{x}(t_n) + \frac{h}{6}(k_1 + k_2 + k_3 + k_4), \quad \begin{cases} k_1 = f(t_n, \vec{x}(t_n)) \\ k_2 = f(t_n + \frac{h}{2}, \vec{x}(t_n) + h\frac{k_1}{2}) \\ k_3 = f(t_n + \frac{h}{2}, \vec{x}(t_n) + h\frac{k_2}{2}) \\ k_4 = f(t_n + h, \vec{x}(t_n) + hk_3) \end{cases} \quad (11)$$

Nel caso del nostro modello, ricordando che il sistema è autonomo (ovvero il campo è nella forma $f(\vec{x})$), il calcolo dei coefficienti è dato dalle seguenti espressioni:

$$\left\{ \begin{array}{l} (V) \quad a_1 = \mu \frac{V(t)}{\xi} \left(1 - \frac{V(t)}{\xi(N-\eta)} \right) \\ (S) \quad b_1 = -\beta \frac{S(t)}{N} M(t) - a_1 \\ (R) \quad c_1 = \gamma M(t) \\ (M) \quad d_1 = -a_1 - b_1 - c_1 \end{array} \right. \quad \left\{ \begin{array}{l} a_2 = \mu \frac{V(t)+a_1 h/2}{\xi} \left(1 - \frac{V(t)+a_1 h/2}{\xi(N-\eta)} \right) \\ b_2 = -\beta \frac{S(t)+b_1 h/2}{N} (M(t) + d_1 h/2) - a_2 \\ c_2 = \gamma (M(t) + d_1 h/2) \\ d_2 = -a_2 - b_2 - c_2 \end{array} \right. \quad \dots \quad (12)$$

Scegliendo lo step size $h = 1$ (giorno), si otterranno i valori delle quattro variabili di ogni giorno. Ad ogni iterazione vengono ri-calcolati i coefficienti $a_i, b_i, c_i, d_i, i \in \{1, 2, 3, 4\}$, che sarebbero dunque 16 valori per ogni giornata. Usando lo stesso ragionamento di prima sul vincolo, si riesce a esprimere uno dei quattro coefficienti in funzione degli altri tre, e quindi i valori da calcolare si riducono a 12 per ogni giornata.

3 Il Codice

Per rendere più leggibile il codice si è deciso di dividerlo in più *translation units*: `infection.cpp`, `display.cpp` e `main.cpp`. Di seguito si dà una breve spiegazione del funzionamento e dei metodi implementati.

3.1 Infection

Per rendere il codice più leggibile si è scelto di raccogliere i metodi riguardanti l'evoluzione della pandemia all'interno del

```
namespace epidemic{.
```

Si è definita una struct `State`:

```
struct State {
    int S;
    int M;
    int R;
    int V;
};
```

che descrive i valori delle quattro variabili di stato in un dato giorno. Si è poi scritto la classe `Infection`

```
class Infection {

    int m_time_indays;
    std::vector<State> m_data;
    int const m_N;
```

che ha come membri privati la durata della simulazione, un vettore di stati, che rappresenta l'evoluzione della pandemia, e la popolazione totale.

3.2 display

In `Infection.cpp` si sono raggruppati i metodi che, come da nome, danno una rappresentazione dei risultati ottenuti.

Il primo metodo è

```
void print(Infection const &infection){
```

che stampa sul terminale i risultati della simulazione e il valore `int` del giorno, avvalendosi della funzione

```
int count_digit(int n){
```

dichiarata sempre all'interno di `display.cpp` che serve ad impaginare in modo leggibile i valori stampati. Si è usata inoltre la libreria `termcolor` che permette di colorare quanto stampato su terminale per colorare i nomi delle variabili.