

SeirModelEpidemic

Buccheri Francesco

Cdl 2019-2020 Settembre 2020

Introduzione

Il programma **SeirModelEpidemic** esegue calcoli basati sul modello matematico SEIR per lo studio dello sviluppo di un'epidemia. L'utente che si troverà alle prese con questo codice scritto in C++ potrà inserire un campione di popolazione a scelta, dal piccolo borgo in provincia di Bologna alla Slovacchia, ottenendo una tabella che mostrerà l'andamento numerico di un'epidemia; successivamente si potranno attivare espansioni del codice che mostreranno ulteriori modi per analizzare e capire meglio questo tipo di sviluppo.

1 Teoria del modello SEIR

Il modello SEIR (acronimo di Susceptible, Exposed, Infectious, Recovered) è un modello matematico costituito da più equazioni, dipendenti da diversi parametri, coinvolte nella genesi e nell'evoluzione di un fenomeno epidemico/pandemico. Il modello SEIR è da considerarsi una rielaborazione arricchita del modello SIR, più basilare e approssimativo, basato solamente sui tre parametri dei suscettibili, infetti e rimossi, senza quindi il gruppo degli esposti. I parametri sopra citati si riferiscono alla condizione di salute in cui versa un componente del campione di popolazione: per *suscettibili* si intende chi, sebbene non immune al virus, non ha ancora contratto la malattia, ovviamente supponendo che non siano ancora state adottate campagne di vaccinazione. Per *esposti* ci si riferisce a chi ha contratto la malattia ma si trova nel periodo di incubazione; di conseguenza, per *infetti* si intende chi ha contratto la malattia e ne vede la manifestazione. Infine, per *rimossi* non si intende solo chi è **guarito** dal morbo, ma per semplicità progettuale, anche chi è sfortunatamente **deceduto**; basta che una persona suscettibile incontri una esposta o addirittura infetta, dato che entrambi possono trasmettere il morbo, per entrare nel gruppo degli esposti.

1.1 Ipotesi di base e assunzioni

Il modello SEIR è da considerarsi, dunque, un modello compartimentale, ovvero un modello matematico in cui si sono assunte delle ipotesi apriori per semplificare la simulazione e soprattutto l'implementazione. Ad esempio, si è optato per la descrizione di una epidemia generica, un virus X non specificato, la cui novità prevede ovviamente l'assenza di norme di contrasto, ma i cui parametri intrinseci sono stati ispirati da quelli trovati in letteratura. Ulteriore ipotesi è stata quella di definire alcune costanti caratteristiche dell'epidemia considerata, di seguito approfondite, per semplificare ulteriormente la trattazione; oltre a ciò, è stata supposta una popolazione divisa in compartimenti di individui tutti con le stesse caratteristiche, senza ad esempio tener conto di possibili portatori sani, neonati immuni alla nascita o tratti genetici più o meno predisposti alla malattia. Inoltre, per rappresentare ancor meglio la realtà, abbiamo deciso di estendere il modello introducendo un tasso di mortalità costante e un numero di nascite relativamente molto ridotto, anch'esso costante.

1.2 Costanti utilizzate

Al fine di realizzare un'implementazione del modello matematico il più possibile fedele a un'ipotetica realtà epidemica, si sono definite alcune costanti, impropriamente chiamate tali poiché la maggior parte è strettamente correlata a parametri che vengono costruiti nella funzione `main()` del programma `SeirModelEpidemic.cpp`:

- **tau** : rappresenta il periodo medio di manifestazione della malattia per ciascun individuo che l'abbia contratta, interpretabile in giorni. Nel metodo pubblico createTable(), della classe Epidemic dello header epidemic.hpp, viene definita un'ulteriore variabile che dipende strettamente da tau: **gamma**, che indica la probabilità di un individuo di guarire, in generale di passare nel gruppo dei rimossi. Nel codice Doctest.infomodel.cpp si trova la possibilità di verificare la correttezza della relazione $\text{gamma}=1/\text{tau}$; tale risultato è frutto di una formula a successione di somme parziali
- **t** : si riferisce al periodo medio di incubazione della malattia di ciascun individuo. È correlato ad un altro parametro, **alfa**, che corrisponde matematicamente al suo reciproco
- **lambda** : si riferisce al numero di nascite giornaliere che avvengono all'interno del campione di popolazione. Probabilmente, questa costante è il dato più inverosimile e sbagliato tra tutti quelli definiti, poiché per scelta è stato inizializzato a 50, a prescindere dal numero iniziale di suscettibili
- **mi** : rappresenta il tasso di mortalità giornaliero costante all'interno della popolazione campione
- **ni** : è il numero medio di persone che un individuo, a prescindere dal gruppo di appartenenza, incontra nell'arco di una giornata
- **P** : rappresenta la probabilità che un soggetto ha di infettarsi e dipende strettamente dal numero di persone ni incontrate e dal numero di suscettibili presenti nel campione. Questo dato ha la caratteristica di mutare in base a quanti suscettibili siano ancora presenti dopo un certo periodo di tempo. Il programma "Doctest.infomodel.cpp" fornisce la possibilità di valutare la correttezza della formula approssimata utilizzata per costruire la variabile corrispondente a questo dato. La decisione di adottare una formula relativamente più semplice dipende dal fatto che essa sia in correlazione con quella più complessa tramite un fattore molto piccolo, di norma intorno a 1.33, per cui poco incidente; sarà quindi osservata una discrepanza da non considerarsi eccessivamente significativa su campioni di dati molto numerosi. Infine, da P e da ni dipende il parametro **beta** secondo la formula $\text{beta}=\text{ni}*\text{P}$

1.3 Equazioni differenziali e spiegazione

Il modello SEIR si basa sulla discretizzazione di alcune equazioni differenziali, riportate di seguito, dipendenti dai parametri già citati in precedenza:

$$\begin{aligned} S_i &= S_{i-1} + \text{lambda} - (\text{mi} * S_{i-1}) - (\text{beta} * I_{i-1} * S_{i-1}) \\ E_i &= E_{i-1} + (\text{beta} * I_{i-1} * S_{i-1}) - (\text{mi} + \text{alfa}) * E_{i-1} \\ I_i &= I_{i-1} + (\text{alfa} * E_{i-1}) - (\text{gamma} + \text{mi}) * I_{i-1} \\ R_{i-1} &= (\text{gamma} * I_{i-1}) - (\text{mi} * R_{i-1}) \end{aligned}$$

Si è ritenuto opportuno inserirne una quinta, riferita alla variazione periodica del cosiddetto **numero di riproduzione di base (r0)**, un indice di quante persone in media un individuo infetta. A causa di motivi concettuali è stato compiuto un errore di base dovuto sostanzialmente al fatto che le equazioni riportate siano solo un'approssimazione al caso discreto di equazioni differenziali; ciò ha consequenzialmente fissato un numero iniziale del nostro r0 quasi scorrelato dal numero iniziale di suscettibili e dal parametro beta. Questi due dati si sono come compensati una volta moltiplicati.

$$r0_i = S_{i-1} * (\text{beta} * \text{alfa}) / ((\text{gamma} + \text{mi}) * (\text{alfa} + \text{mi}))$$

Infine, come ulteriore assunzione realistica e di semplificazione, abbiamo deciso di fissare alla soglia di $r0=1$ il limite di trasmissibilità della malattia: dato che in media un individuo avrebbe infettato meno di un altro soggetto, la malattia si sarebbe presto spenta. La prima equazione mostra che il numero di suscettibili registrato ad un istante di tempo iesimo è ricavabile sottraendo al numero di suscettibili all'istante di tempo precedente il numero di suscettibili deceduti nel corso di quel frammento temporale non a causa della malattia e tutti quei suscettibili che hanno contratto la malattia nel frattempo, il tutto ovviamente sommato al numero, se pur esiguo, di nascite giornaliere. La seconda equazione mostra un'evidente correlazione tra il numero di esposti all'istante iesimo e il parametro alfa, infatti anche qui al numero di esposti dell'istante temporale precedente vanno sommati coloro che si infettano, ma va sottratta una certa percentuale di persone che o

muore, pur non a causa della malattia, o ha già terminato il periodo di incubazione ed entra nel gruppo degli infetti. La terza equazione tratta proprio di questo insieme, il cui valore ad un istante i esimo si ricava sommando all' I precedente il numero di coloro che provengono dal gruppo degli esposti, sottraendo poi chi, dopo il periodo medio di permanenza della malattia, è deceduto o è guarito, e chi è morto non a causa del morbo. Quest'ultimo insieme di soggetti entrerà a far parte del gruppo dei rimossi, a cui si sottrae sempre un certo numero di ulteriori deceduti.

1.4 Avvicinamento al codice

Il progetto si articola in molti file adeguatamente separati e correlati gli uni con gli altri in base alle necessità:

- File sorgente, codici di testing e istogrammi
 1. SeirModelEpidemic.cpp
 2. Doctest.infomodel.cpp (con incluso doctest.hpp)
 3. Derivate.test.cpp
 4. histo.cpp
- Header files inclusi nel SeirModelEpidemic
 1. epidemic.hpp
 2. seirstruct.hpp
 3. printtables.hpp
 4. development.hpp
- Header file inclusi nel Derivate.test.cpp
 1. doctest.h
 2. derivate.hpp
- Header files incluso in development.hpp
 - map.hpp (con incluso a sua volta seirconstants.hpp)

2 Struttura del programma

Il programma si sviluppa in diversi header files e codici sorgente, nei quali troviamo definizioni di strutture dati, come **struct** e **classi** (ciascuna riferita ad ogni tipo user-defined necessario). Tutte le classi sono state dotate di adeguati costruttori, il cui ruolo è stato molto prezioso per la creazione della cosiddetta *class invariant*. I codici epidemic.hpp e map.hpp contengono classi che giocano un ruolo chiave nel corpo dell'intero programma. Compaiono due funzioni main() all'interno dell'intero programma: la prima si ubica nel codice SeirModelEpidemic.cpp, è da considerarsi il principale; la seconda si pone nel codice sorgente histo.cpp, ma è comunque correlata alla prima, infatti essa è fruibile dall'utente solo dopo aver compiuto una serie di specifiche scelte e aver scelto un numero adeguato per il campione della popolazione. Un criterio cardine che è stato adottato in tutti i file è stato quello di assegnare nomi significativi a ogni oggetto creato e allocato in memoria, dalle semplici variabili alle denominazioni dei file stessi; inoltre, si è ritenuto opportuno definire le free function, ad esempio nel codice derivate.hpp, come *inline*.

2.1 Compilazione ed esecuzione

Per la compilazione del programma abbiamo optato per il classico compilatore g++, anche se tra i vari codici scritti l'utente può facilmente trovare una versione molto basilare di CMake; possiamo considerare il file di testo CMakeLists.txt come un tentativo di compilazione "alternativa" del programma, con l'aggiunta di una cartella build come agevolazione. Per la formattazione di ogni codice presente nel programma si è utilizzato il comando **clang-format**. Si consiglia di partire dal lancio del **SeriModelEpidemic**; ovviamente, si potranno anche lanciare altri sorgenti come i due tests o l'istogramma, anche se lo reputiamo poco logico dato che ci si troverebbe davanti una serie di statements di oscuro significato.

2.2 Output del programma

Ciascun file ha una funzione specifica: la maggior parte dei codici sorgente e degli header file editati a compile time produce un output adeguatamente commentato e arricchito da stringhe letterali per rendere meno virtuale e noiosa l'interazione. L'utente troverà, decidendo di eseguire il sorgente principale `SeirModelEpidemic`, una breve scritta esplicativa che lo guiderà attraverso il programma, consentendogli di immettere un numero arbitrario di suscettibili ed esposti iniziali, per poi inserire il numero di giorni di cui vuole visualizzare lo sviluppo della malattia. A questo punto verrà mostrata in output una rudimentale tabella, le cui colonne saranno occupate dai 4 parametri caratteristici del modello SEIR e da r_0 . Successivamente comparirà il primo "bivio", che porterà l'utente a scegliere tra la possibilità di accedere all'uso del file `derivate.hpp`, la cui funzione si lascerà scoprire direttamente all'utente, oppure di declinare questa opzione. Scegliendo l'opzione "N" verrà mostrato un secondo bivio, in cui si potrà avere la possibilità di visualizzare una mappa esemplificativa dell'andamento della malattia in un certo campione ristretto di popolazione, supponendo apriori che una persona infetta, identificata da una x , contagi gli otto puntini contigui. Questa espansione è in verità un'implementazione del modello SIR, quindi leggermente semplificata. Se l'utente non vorrà neanche visualizzare questa opzione, nel messaggio di arrivederci gli sarà consigliato di eseguire almeno il sorgente `histo.cpp`.

2.3 Interpretazione e commento dei risultati

I risultati numerici ottenuti nella tabella visualizzata in output, a prescindere dai valori iniziali, mostrano un andamento caratteristico e correlato alle ipotesi di costruzione del modello SEIR. Per completezza, si è scelto di trasporre i numeri in uno sviluppo grafico: nelle immagini pdf `sir_model`, `seir_model` e `seir_model_2` si troveranno i tre grafici prodotti. Il primo e il secondo non sono altro che un confronto tra un andamento del modello SIR e del modello SEIR, supponendo che la malattia non si estingua, nonostante il numero di riproduzione base sia ad un certo punto minore di uno, e vada ad infettare tutta la popolazione iniziale di suscettibili. Il terzo file, invece, trova proprio questo presupposto: ad un certo punto la curva dei suscettibili prenderà un drastico andamento costante, in accordo con la semplificazione iniziale. Le curve che descrivono l'andamento degli infetti, in tutti e tre i file, e degli esposti, negli ultimi due, presentano un punto di massimo in cui la funzione trova una tangenza orizzontale: questo valore corrisponde, nel caso del modello SIR, esattamente all'istante in cui $r_0=1$, mentre nel modello SEIR, implementato secondo le nostre ipotesi, a un istante temporale in cui r_0 è leggermente minore di 1. Si noterà anche come la curva che descrive l'andamento dei rimossi non tende a crescere come nel caso del modello SEIR, bensì ad un certo punto avrà una tangenza a retta decrescente a causa del tasso di mortalità. Per semplicità, ogni grafico è stato tracciato considerando un campione iniziale di suscettibili di 10000 unità, con un numero iniziale di esposti di 4 soggetti. Ogni curva è stata realizzata per mezzo del programma `gnuplot` versione 5.2. A seguito di varie prove sulla scelta del tipo da apporre ai parametri cardine di tutto il progetto, contenuti nella struct dichiarata nel file `seirstruct.hpp`, si è optato per il `double`, più corretto statisticamente, provvisto di molte cifre significative per maggior completezza. La scelta, chiaramente, è molto inverosimile e poco adeguata dato che si tratta di persone, esseri umani interi e non scomponibili come omini del lego; logicamente si sarebbe dovuto scrivere il tipo `int`. La tabella forniva però numeri profondamenti irreali: i suscettibili tendevano molto velocemente allo zero e i valori finali non erano compatibili con le condizioni iniziali; parte della popolazione sembrava come essersi persa senza un reale motivo. Questo comportamento è probabilmente dovuto ai limiti del tipo `int`; il compilatore ha infatti troncato moltissimi valori.

Testing

Al fine di garantire la correttezza di alcune formule utilizzate, si sono scritti dei codici di testing, utilizzando il framework per la scrittura del test `doctest`. Nel caso del file `Derivate.test.cpp`, si potrà verificare la correttezza delle funzioni definite nello header `derivate.hpp`. Eseguendo questo codice di testing si potrà studiare la conformità del segno di un coefficiente angolare ricavato a partire da 4 parametri della tabella, in un certo istante di tempo, con quello del valore numerico visualizzato eseguendo l'apposita espansione del codice `SeirModelEpidemic.cpp` (per ulteriori chiarimenti si consiglia di utilizzare questo test). Il sorgente visualizzerà eventuali messaggi di errori in base a delle condizioni di default scritte a compile-time.

Eseguendo invece il programma `Doctest.infomodel.cpp` si avrà una verifica matematica di altre formule, confrontate con la loro forma estesa. Ad esempio, si potrà verificare la correttezza dell'equazione approssimata per ricavare \mathbf{P} , oppure che il periodo medio della durata della malattia può essere espresso come il reciproco del parametro γ e viceversa. Tali verifiche hanno un piccolo errore, come potrà essere visualizzato dall'utente, per il quale i valori confrontati non saranno esattamente uguali a ogni cifra significativa. Per ricavare le formule di \mathbf{P} sono state assunte apriori alcune costanti esemplificative; ovviamente, la corretta verifica vale per ogni ipotesi numerica.

Infine, si cita la presenza di un'ulteriore struttura dati inclusa nel primo programma di testing e appartenente allo header `structconsants.hpp`: la struct presenta alcune costanti che ricalcano i valori già precedentemente attribuiti ai parametri caratteristici del modello SEIR, come λ , μ e tutti gli altri; qualora ne venissero cambiati i valori nel codice sorgente principale o nello header incluso, il doctest esplicherà l'errore.

Conclusioni

La corretta esecuzione dei file di testing e, soprattutto, i grafici e i dati numerici ricavati nelle tabelle, a prescindere dalle condizioni iniziali, mostrano con successo che l'implementazione realizzata è fedele al modello teorico SEIR. Nonostante tutte le assunzioni e ipotesi di base, nonché anche piccoli errori di approssimazione, il risultato ottenuto in tutti gli output è stato quello previsto. Possiamo notare come il file `histo.cpp` mostri chiaramente un andamento ritrovabile nelle curve di sviluppo riferite a ciascun parametro nei grafici realizzati. La presenza di alcune ipotesi, però, porta inevitabilmente a vincoli: ad esempio nella scelta di valori adeguati al numero di popolazione; ciò potrebbe produrre, soprattutto nelle tabelle, dei dati molto lontani dalla realtà e soprattutto matematicamente scorretti, come ad esempio dei valori di suscettibili negativi dopo un certo numero di istanti temporali. Infine, il modello poteva essere ovviamente arricchito con molti ulteriori parametri e reso più vicino a un possibile reale andamento di un'epidemia: si sarebbe potuto rendere ogni costante non più tale ma in funzione del tempo o altri parametri o, ad esempio, realizzare una mappa con diverse librerie grafiche presenti all'intero di linux.

Bibliografia e repository git

Fonti utilizzate: [wikipedia.org/wiki/Modelli-matematici-in-epidemiologia](https://it.wikipedia.org/wiki/Modelli_matematici_in_epidemiologia);

P.Acquistapace, E.Scandurra *Controllo ottimale di epidemie di tipo SIR mediante vaccinazione. Il caso dell'ebola in Africa occidentale (2014)*

B.Cifra, L.Lamberti, S.Marone. *SIR: Suscettibili, Infetti, Rimossi, un modello matematico di epidemie.*

Repository git di riferimento: <https://github.com/francescobuccheri00/seirproject>