

Approccio stocastico a modelli epidemiologici

Marco Moscardo

Matricola 592109

Relatore: Professoressa Toffolo Gianna Maria

Università di Padova

Facoltà di Ingegneria

Laurea in Ingegneria dell'Informazione

Marzo, 27 2012

Indice

1	Introduzione	5
2	Catene di Markov	7
2.1	Richiami di calcolo delle probabilità	7
2.2	Catene di Markov	9
3	Modelli epidemiologici	15
3.1	Modello SIR	15
3.2	Parametri fondamentali per il controllo di un'epidemia	17
3.3	Modello SI	21
4	Modelli epidemiologici - Simulazione stocastica	23
4.1	Modello SI	23
4.1.1	Simulazione stocastica	25
4.2	Modello SIR	28
4.2.1	Simulazione stocastica	29
4.3	Potenziale effettivo per un processo stocastico	32

Capitolo 1

Introduzione

Lo scopo dell'elaborato é lo studio dell'evoluzione di un'infezione in una popolazione attraverso l'utilizzo di due famosi modelli utilizzati in campo epidemiologico: il modello SI e il modello SIR.

Questi due fanno parte della classe dei modelli compartimentali, i quali suddividono la popolazione, nel nostro caso costante, in piú classi, detti compartimenti, per poi esprimere tramite delle funzioni le relazioni che sussistono tra di esse.

É evidente che le relazioni che si vanno a creare tra due compartimenti dipenderanno anche da alcuni fenomeni casuali e lo strumento che piú si presta alla nostra trattazione sono le catene di Markov. Con esse infatti siamo in grado di prevedere e descrivere qualsiasi processo di diffusione epidemica.

Il processo di diffusione di una malattia avviene attraverso eventi casuali come, per esempio, l'incontro tra una persona sana e un'infecto, ed anche una volta che questo incontro si é verificato, non é detto che l'individuo sano si ammali. Per comprendere questi casi all'interno dei modelli che tratteremo utilizzeremo quindi le catene di Markov.

Per facilitare la comprensione dei passaggi che faremo nella nostra trattazione, presentiamo all'interno del capitolo 2 tutte le nozioni che utilizzeremo attraverso definizioni e teoremi.

Successivamente presenteremo le nozioni fondamentali che caratterizzano i modelli SI e SIR. Lo scopo del capitolo 3 é quindi quello di introdurre i concetti dei due modelli epidemiologici che studieremo partendo da delle assunzioni semplificative fino ad arrivare alla scrittura delle equazioni caratteristiche che descrivono completamente i modelli trattati.

I modelli compartimentali, dei quali fanno parte i modelli SI e SIR, suddividono la popolazione presa in considerazione in classi, nel nostro caso in due e tre classi rispettivamente, che formano una partizione. Quello su cui lavoreremo sono le relazioni che legano queste classi, per poi visionare il numero di individui all'interno di ogni compartimento.

Ovviamente i modelli sono presentati da un punto di vista deterministico,

cioé le funzioni di relazione tra i compartimenti sono costanti.

Nel capitolo 4 applicheremo le definizioni e i teoremi presentati nel capitolo 2 alle nozioni apprese nel capitolo 3 arrivando a definire un processo stocastico vero e proprio. Allora lo scopo del capitolo é quello di trovare le funzioni che rappresentano al meglio le relazioni tra i compartimenti comprendendo i casi aleatori citati prima.

Come conclusione abbiamo simulato la diffusione dell'epidemia per dimostrare la bontá del processo ricavato e abbiamo notato che nei modelli SI non é sempre detto che l'epidemia si estingua, infatti é possibile arrivare ad uno stato di equilibrio dove il numero di infetti e di sani rimane costante con l'avanzare del tempo. Questa conclusione invece non vale per i modelli SIR, dove l'epidemia sicuramente si estinguerá. Il tempo di durata dipende da quanto é contagiosa l'infezione e da quanto tempo un individuo impiega per guarire.

Capitolo 2

Catene di Markov

2.1 Richiami di calcolo delle probabilità

Si definisce *spazio di probabilità* una terna (Ω, \mathcal{A}, P) , dove Ω é un insieme qualunque (in genere pensato come l'insieme dei risultati possibili di un esperimento casuale), \mathcal{A} é detta σ -algebra, ovvero un insieme di insiemi (gli eventi) per i quali si puó calcolare una probabilità e P é appunto una misura di probabilità su Ω ($P : \Omega \rightarrow [0, 1]$).

Per la precisione, una σ -algebra é una famiglia di insiemi tali che:

- $\emptyset \in \mathcal{A}$;
- se $A \in \mathcal{A}$ allora anche il suo complementare $\bar{A} \in \mathcal{A}$;
- unioni numerali di \mathcal{A} appartengono ancora a \mathcal{A} .

Definizione 2.1.1 Dato uno spazio di probabilità (Ω, \mathcal{A}, P) , si definisce una variabile casuale X una funzione misurabile da Ω in \mathbb{R} ($X : \Omega \rightarrow \mathbb{R}$), ovvero

$$\forall A \in \mathcal{B}(\mathbb{R}), \exists B \in \mathcal{A} : X^{-1}(A) = B$$

cióé tale per cui é sempre possibile misurare attraverso P l'insieme dei valori che essa assume ricorrendo alla misura di probabilità P sullo spazio di partenza:

$$P(X \in A) = P\{\omega \in \Omega : \omega \in X^{-1}(A)\} = P(B), \quad A \subset \mathbb{R}, B \in \mathcal{A}$$

Nella notazione di sopra $\mathcal{B}(\mathbb{R})$ é detta σ -algebra di Borel su \mathbb{R} .

Senza scendere in dettagli, diciamo solo che \mathcal{B} é una particolare σ -algebra che lavora sui sottoinsiemi di \mathbb{R} .

Definizione 2.1.2 Dato uno spazio di probabilità (Ω, \mathcal{A}, P) , si definisce processo stocastico una famiglia di variabili casuali $X_t, t \in I$ definita su $\Omega \times I$ a valori in \mathbb{R} . Quindi le variabili aleatorie (misurabili $\forall t \in I$) che costituiscono il processo, sono funzioni del tipo $X(\omega, t) \rightarrow \mathbb{R}$.

Il parametro t può rappresentare qualsiasi grandezza fisica, ma usualmente viene associato al tempo. Nella nostra trattazione t rappresenterà un determinato istante nell'intervallo di tempo considerato. Invece l'insieme I è il *dominio temporale* del processo, cioè l'intervallo di tempo su cui studiamo il processo.

Ora consideriamo un generico processo stocastico $X(\omega, t)$ e poniamo $\omega = \omega_0$ ottenendo così una funzione nella variabile $t \in I$. Allora $X(\omega_0, t)$ rappresenta l'evoluzione del processo e l'insieme dei valori

$$\{X(\omega_0, t), t \in I\}$$

viene chiamato **traiettoria del processo**.

La traiettoria del processo rappresenta l'andamento che il processo stocastico considerato assume durante l'intervallo di tempo I . Nel nostro elaborato rappresenterà la variazione degli individui infetti all'interno della popolazione presa in esame.

Ad esempio, se $I = \mathbb{N}$ e le $X_n, n \in \mathbb{N}$ sono indipendenti e con la stessa legge, siamo di fronte ad una successione di variabili casuali i.i.d. (indipendenti ed identicamente distribuite). Questa successione è un processo stocastico in *tempo discreto*.

Se invece come I prendiamo l'asse dei tempi $[0, \infty)$, allora $(X_t, t \geq 0)$, viene detto processo stocastico in *tempo continuo* e ogni traiettoria del processo rappresenta l'evoluzione temporale del processo X .

Quindi, ogni valore di $\omega \in \Omega$ genera una traiettoria del processo al variare di t . Ciascuna di queste traiettorie è una possibile realizzazione del processo.

Riassunto:

Definizione 2.1.3 *Si definisce processo stocastico in tempo discreto X_{nT} se il dominio temporale I è un insieme numerabile.*

Definizione 2.1.4 *Si definisce processo stocastico in tempo continuo X_t se il dominio temporale I è un sottoinsieme continuo di \mathbb{R} .*

Definizione 2.1.5 *Si definisce realizzazione di un processo la sequenza di stati che il processo visiterà negli n istanti considerati*

Definizione 2.1.6 *Si chiama traiettoria di un processo stocastico l'intera realizzazione di una successione di variabili $X_n, n \geq 0$, cioè:*

$$\{X_0 = i_0, X_1 = i_1, \dots, X_n = i_n\}$$

con $i \in I$

Durante la nostra trattazione prenderemo in considerazione solo processi stocastici in tempo discreto.

Fissato invece un istante $t = t_0$, il processo stocastico diventa una funzione nella variabile $\omega \in \Omega$. In questo caso $X(\omega, t_0)$ ci fornisce la distribuzione del processo stocastico al tempo t_0 .

Ad ogni processo, ad esempio $(X_t, t \geq 0)$, si può associare, per ogni t , una σ -algebra che indichiamo con $\mathcal{F}_t = \sigma(X_s; 0 \leq s \leq t)$ (la σ -algebra generata da X_t), cioè la più piccola σ -algebra di \mathcal{A} che rende $X(\omega, s)$ misurabile per ogni $0 \leq s \leq t$. Questa σ -algebra (che deve contenere anche gli insiemi di misura nulla di \mathcal{A}) è l'insieme più piccolo di sottoinsiemi di Ω che ci permette di calcolare tutte le probabilità relative ad eventi che risuardano X_t .

Definizione 2.1.7 *La famiglia di σ -algre ($\mathcal{F}_t, t \geq 0$) viene chiamata **filtrazione naturale del processo** ed è tale per cui $\mathcal{F}_s \subset \mathcal{F}_t$, per $s < t$.*

Le filtrazioni sono dunque successioni di σ -algre crescenti.

Vediamo un esempio. Pensiamo al lancio di una moneta. I risultati sono T o C con uguale probabilità. Possiamo associare a questo esperimento la σ -algebra costruita su $[0, 1]$ nel seguente modo: associamo al risultato T l'intervallo $[0, 1/2)$ e a C l'intervallo $[1/2, 1]$ e le relative probabilità alle lunghezze degli intervalli corrispondenti, ovvero:

$$P(T) = P([0, 1/2)) = 1/2, \quad P(C) = P([1/2, 1]) = 1/2$$

allora:

$$\mathcal{F} = \{\emptyset, [0, 1/2), [1/2, 1], [0, 1]\}$$

è una σ -algebra.

L'importanza delle filtrazioni è data dal fatto che esse sono un modo per descrivere l'aumento di informazione portata dal processo che si ha con il trascorrere del tempo.

2.2 Catene di Markov

Prima di dare la definizione formale di catena di Markov, introduciamo alcuni concetti che ci serviranno per comprendere a fondo l'importanza di questo strumento.

Definizione 2.2.1 *Si definisce probabilità di transizione $P(i \rightarrow j)$ come*

$$P(i \rightarrow j) = P(X_{n+1} = j | X_n = i), \quad i, j \in I$$

cioè la probabilità che il prossimo stato sia j dato che lo stato corrente è i .

Definizione 2.2.2 *Le catene di Markov sono identificate da:*

- *l'insieme degli stati $I = \{1, 2, \dots, m\}$;*
- *l'insieme delle possibili transizioni é dato dalle coppie (i, j) tale per cui la loro $P(i \rightarrow j) > 0$;*
- *i valori assunti da $P(i \rightarrow j)$ sono positivi.*

Definizione 2.2.3 *Si definisce catena di Markov un processo stocastico a tempo discreto $(X_n)_{n \in T}, T \subseteq \mathbb{N}_0$ tale che per $i, j \in I, n \in \mathbb{N}$, vale la seguente proprietá:*

$$P(X_{n+1} = j | X_n = i, X_{n-1} = i_{n-1}, \dots, X_1 = i_1) = P(X_{n+1} = j | X_n = i). \quad (2.1)$$

La proprietá appena espressa é anche nota come *proprietá di Markov* e sta a significare che le probabilitá di transizione $P(i \rightarrow j)$ non dipendono dalle transizioni avvenute in precedenza, ma solamente dallo stato in cui il sistema si trova nell'istante i , quindi le probabilitá di transizione definiscono completamente il processo di Markov.

Riassumendo:

un processo stocastico markoviano o processo di Markov é un processo stocastico nel quale la probabilitá di transizione che determina il passaggio ad uno stato di sistema dipende unicamente dallo stato di sistema immediatamente precedente (proprietá di Markov) e non dal come si é giunti a tale stato.

Definizione 2.2.4 *Si definisce catena di Markov omogenea una catena di Markov in cui i due membri della relazione 2.1 non dipendono da n .*

Nel nostro caso, questo significa che le probabilitá di transizione $P(i \rightarrow j)$ non dipenderanno dal tempo t e quindi tratteremo sempre, in questo elaborato, catene di Markov omogenee.

Definizione 2.2.5 *Un vettore $\pi = (\pi_i)_{i \in \chi}$, con χ l'insieme degli indici di π , é detto stocastico su χ se*

$$\begin{aligned} \pi_i &\geq 0 \quad \forall i \in \chi \\ \sum_{i \in \chi} \pi_i &= 1 \end{aligned}$$

In questo elaborato un vettore stocastico rappresenterá le condizioni iniziali in cui si trova il sistema, cioé prima che l'epidemia inizi a diffondersi nella popolazione. Inoltre l'insieme χ corrisponderá all'insieme degli stati I .

Supponiamo ora che l'insieme degli stati I contenga N elementi. Possiamo quindi inserire tutte le probabilità di transizione $P(i \rightarrow j)$, con $i, j = 1, 2, \dots, N$ all'interno di una matrice quadrata $N \times N$. Definiamo così una *matrice delle transizioni* M , $N \times N$ tale che:

$$M_{ij} = (p_{ij})_{i,j \in I \times I} \quad \forall i, j \in I$$

in cui $p_{ij} = P(i \rightarrow j)$.

Definizione 2.2.6 Si definisce matrice stocastica o di Markov una matrice delle transizione che soddisfa la seguente condizione:

$$\sum_{j \in I} P(i \rightarrow j) = 1 \quad \forall i \in I$$

Osserviamo subito che le righe della matrice stocastica sono dei vettori stocastici.

Definizione 2.2.7 Una matrice di transizione si definisce bistocastica se gli elementi delle colonne sommano a 1 e vale anche la definizione 2.2.6.

Definizione 2.2.8 Si definisce matrice di ricorrenza la matrice trasposta della matrice di ricorrenza M , cioè:

$$R = M^T \tag{2.2}$$

Questa matrice è comunemente utilizzata nello studio di una certa tipologia di sistemi; infatti, con essa, è possibile rappresentare la dinamica di un sistema che si evolve nel tempo. Supponiamo di avere un processo omogeneo discreto X_n che rappresenta un sistema, allora posso esprimere la relazione tra due stati consecutivi del processo con la seguente equazione:

$$X_{n+1} = R X_n \tag{2.3}$$

dove i termini R_{ij} rappresentano l'effetto che la variabile j -esima ha su quella i -esima nell'intervallo di tempo $[n, n + 1]$.

Inoltre, nella matrice R la notazione per indicare la probabilità di transizione da uno stato i ad uno stato j è $P(j \rightarrow i)$ invece che $P(i \rightarrow j)$ (questa inversione è una diretta conseguenza dell'utilizzo della matrice di ricorrenza, definita come la trasposta della matrice delle transizione M). Questo comporterà che nella matrice di ricorrenza saranno le colonne ad essere vettori stocastici.

Sfruttando l'osservazione appena fatta, immaginiamo di conoscere la probabilità $p_k(t)$ che il sistema al tempo t si trovi nello stato k -esimo. Definiamo il vettore stocastico $\pi(t)$ contenente tutte le probabilità $p_k(t)$, $\forall k =$

$1, 2, \dots, N$, allora per quanto detto prima possiamo ricavare il vettore $\pi(t + dt)$, contenente le $p_k(t + dt)$. Attraverso la 2.3 otteniamo:

$$\pi(t + dt) = R\pi(t) \quad (2.4)$$

con R matrice di ricorrenza del processo (supponiamo che le probabilità di transizione siano indipendenti dal tempo, così che R sia costante). Più in generale possiamo dire che dopo N volte dt avremmo che:

$$\pi(t + Ndt) = R^N \pi(t) \quad (2.5)$$

Definizione 2.2.9 *Si definisce distribuzione iniziale π_0 di una catena di Markov, se vale*

$$P(X_0 = i_0) = \pi_0.$$

Questa definizione stabilisce che la probabilità di una traiettoria é completamente caratterizzata dalla distribuzione di probabilità dello stato iniziale X_0 e dalla matrice di transizione.

Teorema 2.2.1 *Un processo stocastico X_n , con $0 \leq n \leq N$ é una catena di Markov con distribuzione iniziale $\pi(0) = \pi_0$ e matrice di transizione M se e solo se $\forall i_0, \dots, i_N \in I$ si ha*

$$P(X_0 = i_0, X_1 = i_1, \dots, X_N = i_N) = \pi_{i_0} P(i_0 \rightarrow i_1) P(i_1 \rightarrow i_2) \dots P(i_{N-1} \rightarrow i_N) \quad (2.6)$$

Grazie a questo teorema siamo in grado di esprimere la relazione che si crea tra la distribuzione di probabilità iniziale π_0 e la probabilità di visitare un determinato i_k stato al tempo $t = n$. Entrando nel dettaglio possiamo esprimere $P(X_n = i_k)$ come:

$$P(X_n = i_k) = \sum_{j=1}^N P(X_n = i_k | X_{n-1} = i_j) P(X_{n-1} = i_j) = \sum_{j=1}^N P(j \rightarrow k) P(X_{n-1} = i_j) \quad (2.7)$$

ed, volendo esprimerla in forma matriciale, definendo $\pi(n)$ il vettore stocastico relativo a:

$$\pi(n) = (P(X_n = i_1), P(X_n = i_2), \dots, P(X_n = i_N)),$$

si può esprimere $\pi(n) = R\pi(n - 1)$, con R matrice di ricorrenza, definita come la trasposta della matrice di transizione M . Applicando ricorsivamente questa definizione, otteniamo:

$$\pi(n) = R^N \pi(0) \quad (2.8)$$

che é lo stesso risultato che abbiamo ottenuto con l'equazione 2.5.

Definizione 2.2.10 *Si definisce matrice ergodica una matrice che non ammette sottospazi invarianti generati da vettori che rappresentano gli stati.*

Questa definizione sta a significare che ogni stato del sistema é raggiungibile da qualsiasi altro stato, interno al sistema, per una qualche realizzazione del processo, dove con raggiungibile intendiamo che esiste una realizzazione che mi permette di visitare un qualsiasi stato del sistema partendo da qualsiasi altro stato, sempre del sistema.

Teorema 2.2.2 *Una matrice M ergodica possiede autovettori che hanno moduli compresi tra 0 e 1, l'autovettore di modulo massimo é $\lambda_0 = 1$ ed ha molteplicitá uno.*

L'importanza di questo teorema viene evidenziata quando M é diagonale, di conseguenza lo sará anche la matrice di ricorrenza R . Prima di addentrarci nella spiegazione dell'importanza di questo teorema, andiamo a dare il teorema di diagonalizzabilitá e la definizione di matrice diagonale.

Teorema 2.2.3 (Teorema di diagonalizzabilitá) *Una matrice quadrata M con n righe é diagonalizzabile se e solo se valgono entrambe le condizioni:*

- (1) *la somma delle molteplicitá algebriche dei suoi autovettori é n ;*
- (2) *le molteplicitá algebriche e geometriche di ogni autovalore sono coincidenti.*

Definizione 2.2.11 (Matrice diagonalizzabile) *Una matrice M é diagonalizzabile se esiste una matrice invertibile A tale che*

$$P^{-1}MP = \begin{pmatrix} \lambda_0 & 0 & \dots & 0 \\ 0 & \lambda_1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_n \end{pmatrix}$$

Ritornando al teorema 2.2.2, il fatto che la matrice R sia diagonalizzabile implica che esiste una matrice A che la diagonalizza e quindi possiamo definire la matrice Λ tale che

$$\Lambda = A^{-1}RA = \text{diag}(\lambda_0, \dots, \lambda_{n+1})$$

Definiamo la quantitá $\phi = A^{-1}\pi$ ed effettuando un cambio di variabili alla 2.5 otteniamo che

$$\phi(t + Ndt) = \Lambda^N \phi(t)$$

ma per come abbiamo definito Λ sappiamo che sará la matrice diagonale con gli autovettori di R sulla diagonale principale.

Ora facendo tendere $N \rightarrow \infty$ avremmo che gli elementi sulla diagonale si annulleranno tranne quello unitario e quindi per $t \rightarrow \infty$ avremo che $\phi(t) \rightarrow$

ψ_0 , con ψ l'autovettore corrispondente all'autovettore unitario λ_0 . Allora per $\phi = A^{-1}\pi$ avremo che

$$\pi(t) = A\phi(t) \rightarrow A\psi_0.$$

Questa relazione é fondamentale poiché se noi andiamo a trovare il valore di ψ_0 e la matrice che diagonalizza R permettono di prevedere l'andamento del sistema corrispondente alla catena di Markov.

Capitolo 3

Modelli epidemiologici

Dopo aver introdotto le proprietà dei processi aleatori e delle catene di Markov che utilizzeremo, presentiamo due tipologie di modelli epidemiologici: il modello SIR e il Modello SI. Ovviamente non esistono solo questi due modelli per lo studio delle epidemie, ma in questa trattazione ci limiteremo a studiare il comportamento di questi due perché sono alla base di tutti gli altri. Successivamente a questo capitolo, applicheremo i concetti presentati nel capitolo precedente ai modelli che ora introdurremo.

3.1 Modello SIR

In questa tipologia di modelli la popolazione viene suddivisa in 3 compartimenti che definiscono una partizione della popolazione presa in esame.

I 3 compartimenti rappresentano le tre possibili situazioni in cui si può trovare un individuo durante lo studio dell'epidemia. Le situazioni sono:

- *suscettibili*, sono coloro che sono suscettibili all'infezione, cioè che si possono infettare;
- *infetti*, sono coloro che hanno contratto l'infezione
- *rimossi*, sono coloro che hanno già contratto l'infezione e sono guariti, ottenendo così l'immunità.

Con immunità intendiamo che il paziente non potrà più essere infettato.

Nel modello SIR, l'individuo che entra a far parte del compartimento dei rimossi otterrà una immunità permanente (in altre tipologie di modelli, come ad esempio il modello SIRS, l'immunità è temporanea, cioè l'individuo dopo un certo arco di tempo torna ad essere suscettibile all'infezione).

Per una più semplice comprensione del modello SIR consideriamo le seguenti assunzioni:

- (a) il tempo di incubazione dell'infezione sia nullo;
- (b) la probabilità di incontro tra due individui qualsiasi della popolazione

sia uguale;

(c) il contagio avvenga per contatto diretto;

(d) la probabilità di guarigione sia costante per unità di tempo.

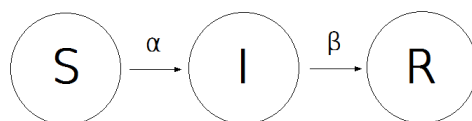
Indichiamo con N il numero di individui totali della popolazione che supponiamo costante, allora per la definizione di partizione sappiamo che al tempo t avremo:

$$N = S(t) + I(t) + R(t) \quad \forall t \quad (3.1)$$

Riassumendo:

abbiamo assunto che il contagio avvenga per contatto diretto dato che gli incontri tra due qualsiasi individui sono equiprobabili. Il numero dei nuovi contagiati, ed anche di nuovi infettivi, per unità di tempo sarà proporzionale al numero di contatti tra individui nella classe S (susceptibili) ed individui nella classe I (infetti), ossia proporzionale al prodotto SI . D'altra parte, nella stessa unità di tempo vi saranno degli individui malati e quindi infettivi che guariscono, e per l'ultima assunzione fatta, ciò avverrà in proporzione al numero di malati.

Ora siamo in grado di rappresentare graficamente le relazioni che legano i vari compartimenti del modello SIR:



dove α rappresenta la probabilità che un individuo suscettibile possa infettarsi, β rappresenta la probabilità che un infetto guarisca e ottenga l'immunità.

Quindi le equazioni che caratterizzano il modello SIR sono:

$$\begin{cases} dS/dt = -\alpha SI \\ dI/dt = \alpha SI - \beta I \\ dR/dt = \beta I \end{cases} \quad (3.2)$$

con α e β positivi, e con condizioni iniziali $I(0) = I_0$ (esiste nella popolazione un piccolo gruppo di infetti), $R(0) = R_0$ non necessariamente nullo (possibile presenza di individui geneticamente immuni all'infezione), per $S(0)$ sfruttiamo la relazione 3.1 ottenendo $S(0) = S_0 = N - I_0 - R_0$. Come notiamo dalle equazioni 3.2, il numero di suscettibili, infetti e rimossi varia nel tempo e secondo alcuni coefficienti. Nel nostro caso questi coefficienti, nello specifico α e β , sono dei valori costanti, ma questo non sta a significare che non possano assumere valori variabili nel tempo. Questi due coefficienti svolgono un ruolo fondamentale nello studio delle infezioni, ed ora cercheremo di capirne il motivo.

3.2 Parametri fondamentali per il controllo di un'epidemia

Parliamo di epidemia quando il numero di infetti diventa maggiore di una certa quantità $I_0 > 0$, cioè l'infezione si propaga.

Ritornando alle equazioni SIR, notiamo che la quantità dS/dt è sempre minore di zero. Vediamo allora che se $S < \gamma = \beta/\alpha$ allora anche dI/dt sarà negativa. Da questo possiamo dedurre che se il numero dei suscettibile iniziale è inferiore al valore β/α , allora l'epidemia non si propagherà. Nel caso in cui S_0 sia maggiore del valore soglia, allora il numero di infetti crescerà fino a quando S sarà minore di γ . Quindi il valore massimo di infetti I^* sarà raggiunto quando $S = \gamma$. Definiamo così il parametro I^* che rappresenta il picco massimo di infetti raggiunto durante l'epidemia.

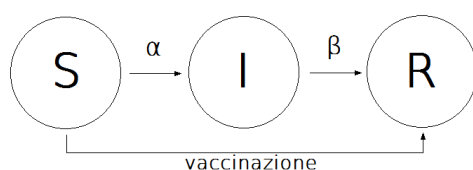
Definiamo anche il parametro *tasso relativo di rimozione* γ che misura la velocità di rimozione/guarigione di un individuo rispetto all'infezione. L'inverso è chiamato *tasso di contatto*.

Nello studio di un'epidemia teniamo anche conto del *tasso riproduttivo dell'infezione*, definito come $r = (\alpha/\beta)S_0$ e misura quante infezioni secondarie sono prodotte da ogni infezione primaria da quando l'infezione è stata introdotta nella popolazione.

Vaccinazione

Per un'infezione data, l'unica classe sulla quale possiamo agire per evitare l'epidemia è la classe dei suscettibili. Questo può avvenire tramite una campagna di vaccinazioni che conferisce l'immunità.

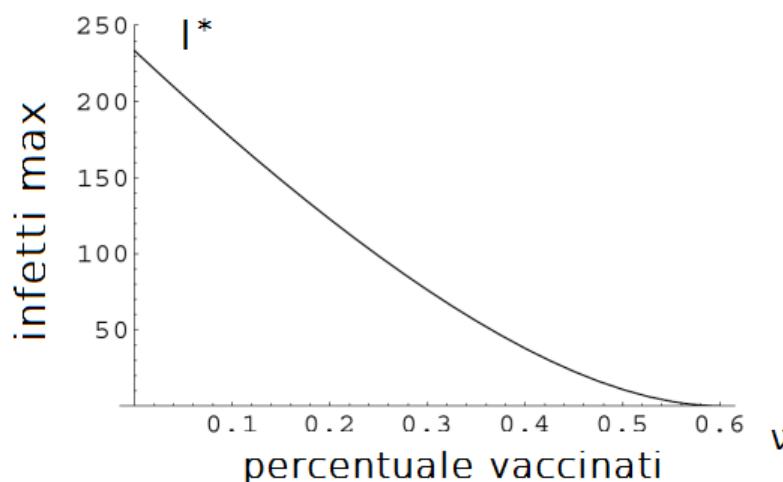
Ora rappresentiamo graficamente il nuovo modello compartimentale, contenente anche la possibilità di una eventuale vaccinazione.



In assenza di vaccinazioni, i suscettibili iniziali sono $\hat{S}_0 \simeq N$, invece vaccinando una percentuale ν della popolazione, il numero di suscettibili diviene

$$S_0 = S_0(\nu) := (1 - \nu)\hat{S}_0 \simeq (1 - \nu)N$$

quindi diminuendo S_0 opportunamente, possiamo evitare il diffondersi dell'epidemia o almeno limitare I^* . Infatti diminuendo la popolazione di suscettibili, siamo in grado di contenere l'infezione e in alcuni casi, possiamo anche non farla iniziare.



Dalla figura notiamo che per una certa percentuale di popolazione vaccinata siamo in grado di ridurre a zero il picco massimo di infetti I^* . Per trovare questa percentuale dobbiamo prima ricavare il valore di I^* .

Gravità dell'epidemia

Nello studio di un'epidemia é fondamentale conoscere il picco massimo di infetti, definito precedentemente come I^* , e per calcolare questo valore abbiamo prima bisogno di conoscere il numero R di rimossi, ma attraverso la relazione 3.1 possiamo ricavarcelo. A questo punto esprimiamo I in funzione di S . Quindi dal sistema 3.2 ricaviamo:

$$\frac{dI}{dS} = \frac{\alpha SI - \beta I}{-\alpha SI} = -1 + \frac{\beta}{\alpha S}$$

Integrando l'equazione trovata arriviamo alla soluzione cercata:

$$I(S) = c_0 - S + (\beta/\alpha) \log(S). \quad (3.3)$$

con c_0 dipendente dalle condizioni iniziali e piú precisamente:

$$c_0 = I_0 + S_0 + (\beta/\alpha) \log(S_0). \quad (3.4)$$

Tralasciamo per il momento il numero massimo di infetti raggiunti nell'epidemia e ci concentriamo sulla porzione di popolazione che durante l'epidemia sará infettata. Questa stima é fondamentale nello studio di un'epidemia poiché ci consente di attuare una campagna vacciazioni tale da arginare sostanzialmente l'infezione.

Definiamo ora il parametro S_∞ che indica il numero di individui suscettibili che al termine dell'epidemia non saranno diventati infetti. Ovviamente l'epidemia si considererá terminata quando il numero di infettivi presenti nella classi I sará diventato zero.

Con l'utilizzo della 3.3 e della 3.4 possiamo trovare il numero di suscettibili

3.2. PARAMETRI FONDAMENTALI PER IL CONTROLLO DI UN'EPIDEMIA 19

S_∞ per un t molto grande, tendente all'infinito.

Ritornando alle relazioni precedenti esplicitiamo la differenza tra S_0 e S_∞ che rappresenta la porzione di popolazione che verrà colpita effettivamente durante l'epidemia, quindi dalla 3.3 ricaviamo:

$$-(\beta/\alpha) \log(S_\infty) + S_\infty = c_0 = I_0 + S_0 - (\beta/\alpha) \log(S_0);$$

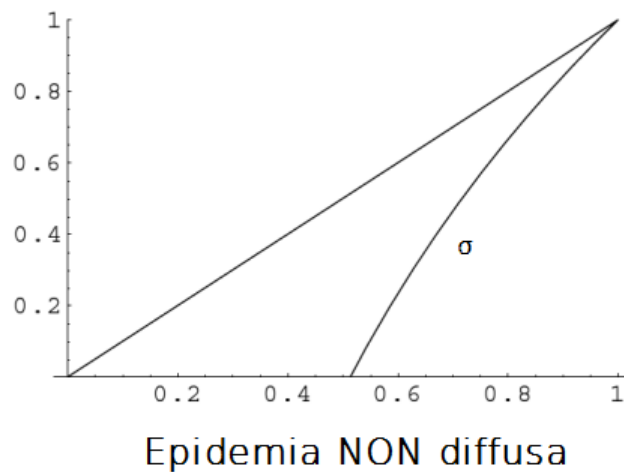
supponendo che $I_0 \ll S_0$ così da poterlo ignorare e esplicitare la differenza cercata, otteniamo:

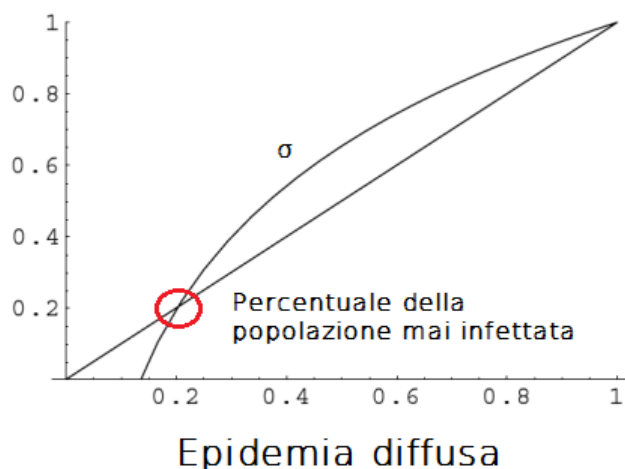
$$\log[(S_0/S_\infty)^\gamma] = S_0 - S_\infty. \quad (3.5)$$

Se indichiamo con σ il rapporto S_∞/S_0 , la 3.5 si riscrive come

$$\sigma = 1 + \log[\sigma^{\gamma/S_0}] = 1 + (\gamma/S_0) \log(\sigma) \quad (3.6)$$

Come vediamo, l'equazione qui sopra é trascendente e questo implica che non é possibile risolverla esattamente ma é facile risolverla numericamente con la precisione voluta. Per una migliore comprensione ne studiamo una curva rappresentativa in due casi.





Nelle due figure sono rappresentate le soluzioni numeriche dell'equazione 3.6 in due diversi casi.

Nella prima figura é rappresentata la situazione in cui il valore S_0 non supera il valore soglia γ e di conseguenza l'epidemia non si diffonde. Notiamo che la curva di σ rimane sotto al valore soglia.

Nella seconda immagine invece, l'infezione si diffonde infatti la curva σ supera il valore soglia. Al termine dell'epidemia, la percentuale di suscettibili che non avrà contratto l'infezione sarà data dal punto di intersezione tra la curva σ e la retta che rappresenta il valore soglia γ .

Ritorniamo ora sul parametro I^* . Ricordando quanto detto in precedenza, sappiamo che il picco massimo di infetti verrà raggiunto quando $S = \gamma = \beta/\alpha$ e, con l'utilizzo della 3.3, possiamo anche valutare quanto sarà diffusa l'epidemia al momento di massima intensità: abbiamo infatti

$$I^* = I(\gamma) = c_0 - \gamma + \gamma \log(\gamma)$$

e ricordando la definizione della costante c_0 espressa nella relazione 3.4 e supponendo nuovamente che $I_0 \ll S_0$, possiamo arrivare a definire che

$$I^* \simeq (S_0 - \gamma) + \gamma[\log(\gamma) - \log(S_0)] = (S_0 - \gamma) - \gamma \log(S_0/\gamma). \quad (3.7)$$

Grazie alla relazione 3.7 siamo ora in grado di trovare la percentuale di popolazione suscettibile che dovremmo vaccinare per non far diffondere l'infezione e ridurre a zero il parametro I^* . Quindi, per la 3.7 ed ricordando che $S_0 = (1 - \nu)N$ scriviamo

$$I^*(\nu) \simeq [(1 - \nu)N - \gamma] - \gamma \log[(1 - \nu)N/\gamma]$$

dove abbiamo espresso I^* in funzione del parametro ν che rappresenta la percentuale di popolazione vaccinata.

Ricordiamo che un'epidemia si diffonde se $S_0 > \gamma$ quindi a noi interessa il caso massimo in cui l'epidemia non si diffonde, cioè $S_0 = \gamma$. Allora per come avevamo definito S_0 nella sezione della vaccinazione abbiamo che l'epidemia non inizierà mai se vacciniamo

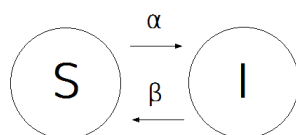
$$\nu_0 := 1 - \gamma/N$$

dove con ν_0 intendiamo la percentuale di popolazione che *inizialmente* dovremo vaccinare per evitare che, all'arrivo di un individuo ammalato nella popolazione, l'infezione si diffonda.

3.3 Modello SI

Ora presentiamo il modello piú semplice dei modelli epidomologici, infatti in questa tipologia non é prevista una immunitá permanente o temporanea. Con questo intendiamo che coloro che sono appena guariti saranno di nuovo suscettibili all'infezione (un esempio é il raffreddore).

La rappresentazione per compartimenti mostrerá due dei 3 compartimenti mostrati nel modello SIR, cioè quello dei *suscettibili* S e quello degli *infettivi* I . Le dinamiche di iterazione tra le due classi sono simili a quelle presentate prima, solo che ora un individuo, una volta che é guarito, torna ad essere suscettibile all'infezione.



dove α rappresenta la probabilità che un individuo suscettibile possa infettarsi, β rappresenta la probabilità che un infetto guarisca e ritorni ad essere suscettibile all'infezione.

In questo caso le equazioni SI saranno:

$$\begin{cases} dS/dt = -\alpha SI + \beta I \\ dI/dt = \alpha SI - \beta I \end{cases} \quad (3.8)$$

Dato $S + I = N$ possiamo esprimere S in funzione del tempo:

$$dS/dt = \beta N - (\alpha N + \beta)S + \alpha S^2$$

e successivamente integrando, otteniamo:

$$S(t) = \frac{\beta e^{\alpha N t} + c_0 N e^{\beta t}}{\alpha e^{\alpha N t} + c_0 e^{\beta t}},$$

dove c_0 dipende dai valori iniziali,

$$c_0 = \alpha \left(\frac{S_0 - \beta/\alpha}{N - S_0} \right)$$

In questo modello riscontriamo che il numero di infetti e suscettibili tende ad un valore costante diverso da zero, partendo da un $I_0 \neq 0$. Nel caso in cui il numero di infetti sia nullo, allora non esiste l'infezione e quindi non ci potranno essere infetti.

È facile vedere che la soluzione stazionaria alla quale il modello SI tende è quando $S = \gamma$, infatti possiamo ricavare dalla prima delle equazioni di sistema 3.8 che

$$dS/dt = (\beta - \alpha S)I = \alpha I(\gamma - S)$$

dove, sapendo che α positiva e $I \geq 0$, abbiamo per soluzioni non banali ($I \neq 0$) che S cresce quando è minore di γ e decresce quando è maggiore di γ .

Capitolo 4

Modelli epidemiologici - Simulazione stocastica

Consideriamo ora i modelli che descrivono l'evoluzione istantanea e non la quantità media di un processo stocastico.

Nel capitolo precedente siamo partiti da uno stato iniziale ed attraverso le equazioni caratteristiche dei due modelli, abbiamo descritto l'andamento del numero dei nuovi infetti. Questa descrizione sarà quindi un valore atteso dato da una certa distribuzione di popolazione e una certa probabilità di infettarsi.

Quello che vedremo in questo capitolo saranno le fluttuazioni, cioè come si comporta il processo stocastico istantaneamente rispetto all'andamento medio trovato.

4.1 Modello SI

Ricordiamo che nel modello SI trattato nel capitolo precedente, la variabile principale di studio non era il picco massimo di infetti ma la condizione di equilibrio alla quale tendeva il sistema.

Nel modello che cerchiamo di formulare in questo capitolo, consideriamo che ogni membro sano (*suscettibile*) abbia una probabilità di ammalarsi dopo aver incontrato un infetto pari a αdt in un tempo infinitesimo dt , ed ogni malato abbia una probabilità βdt di guarire nello stesso tempo.

Quindi avremo che se al tempo t ci sono k malati e quindi $s = n - k$ sani, la probabilità che nessuno dei sani si ammali è:

$$(1 - \alpha dt)^{sk} = 1 - (\alpha ks)dt + o(dt).$$

Se definiamo $P_a(m; s)$ come la probabilità che tra s membri sani ci siano m membri che si ammaliano, avendone già k infetti, avremo che:

$$\binom{s}{m} (1 - \alpha dt)^{(s-m)k} (\alpha k dt)^{km}$$

che é di ordine $(dt)^m$; in generale avremo:

$$P_a(m; s) = \begin{cases} 1 - \alpha s k dt + o(dt) & m = 0 \\ \alpha s k dt + o(dt) & m = 1 (e s \neq 0) \\ o(dt) & m > 1 \end{cases}$$

Vediamo le probabilità di guarigione di q dei k malati presenti al tempo t . Per trovare la probabilità di guarigione di q dei k malati sfruttiamo lo stesso ragionamento utilizzato in precedenza.

Avremo allora:

$$P_g(q; k) = \begin{cases} 1 - k\beta dt + o(dt) & q = 0 \\ k\beta dt + o(dt) & q = 1 (e k \neq 0) \\ o(dt) & q > 1 \end{cases}$$

A questo punto abbiamo tutto il necessario per poter esprimere la probabilità $P(k \rightarrow j)$ di avere j malati al tempo $t + dt$ se al tempo t abbiamo k malati. Essa é:

$$P(k \rightarrow j) = \begin{cases} k\beta dt + o(dt) & \text{per } j = k - 1 \\ 1 - [k\beta + (n - k)k\alpha]dt + o(dt) & \text{per } j = k \\ (n - k)k\alpha dt & \text{per } j = k + 1 + o(dt) \\ o(dt) & \text{altrimenti} \end{cases}$$

Notiamo che le probabilità appena mostrate non comprendono i casi limite $k = 0$ e $k = n (s = 0)$. Il calcolo di queste, per $k = n$ diventa:

$$P(n \rightarrow n) = 1 - n\beta dt, P(n \rightarrow n - 1) = n\beta dt.$$

Per $k = 0$ c'è una sola probabilità non nulla, cioè $P(0 \rightarrow 0) = 1$; in poche parole se non ci sono infetti, l'epidemia non si potrà mai propagare.

Quello che ci interessa ora é riuscire a dare una descrizione delle quantità che sono sostanzialmente determinate da quello che nella teoria dei processi stocastici si chiama un tempo di uscita (*stopping time*). Ad esempio se quando l'infezione si estingue ci saranno delle persone mai toccate dal contagio. Grazie alla notazione che abbiamo ricavato dalle probabilità di guarigione e di infezione, possiamo costruire una matrice di Markov M e di conseguenza la matrice di ricorrenza R associata ad essa. Sfruttando la relazione tra gli elementi della matrice di ricorrenza e le probabilità di transizione, e le definizioni appena date, abbiamo:

$$R_{ij} = [1 - (j\beta + (n - j)j\alpha)dt]\delta_{i,j} + (j\beta dt)\delta_{i,j-1} + [(n - j)j\alpha dt]\delta_{i,j+1}$$

con $0 < i \leq n$ e $0 < j \leq n$. Invece per le considerazioni fatte, per $i = 0$ e $j = 0$ avremo che $R_{ij} = 1$.

Da come é stata definita la matrice di ricorrenza, possiamo vedere facilmente che quando il sistema si trova nello stato con 0 infetti, esso rimarrá in questa

condizione, detta *situazione di equilibrio*: infatti, come detto in precedenza, se non ci sono infetti allora l'epidemia non potrà diffondersi.

Costruiamo l'autovettore $\pi_* = (\pi_*^{(0)}, \dots, \pi_*^{(n)})$ corrispondente all'autovalore 1, che risulta avere come componenti le quantità:

$$\pi_*^{(k)} = \binom{n}{k} \frac{\alpha^k b^{n-k}}{(a+b)^n}$$

e ciò ci fornisce la distribuzione stazionaria per $t \rightarrow +\infty$. Attraverso questo autovettore però siamo anche in grado di ricavarci la media dei malati e la dispersione di k (numero malati) intorno alla sua media.

Con questo approccio siamo allora in grado di dare una informazione statistica del primo a qualsiasi tempo t , e non solo all'equilibrio raggiunto dal sistema. Infatti, determinata la matrice di Markov e la conseguente matrice di ricorrenza, partendo dallo stato iniziale siamo in grado di calcolare la distribuzione di probabilità a qualsiasi tempo successivo, e dunque *a fortiori* di calcolare l'evoluzione temporale di qualsiasi quantità statistica.

La maggiore quantità di informazioni apportate da questo approccio comporta una maggiore difficoltà di calcolo.

4.1.1 Simulazione stocastica

Un altro metodo, alternativo a quello analitico, si basa sulla simulazione, che permette di aggirare gli ostacoli dell'approccio analitico, che porta a equazioni risolvibili solo per sistemi molto semplici. La decisione di adottare un approccio simulativo è dettata dal fatto che con un metodo analitico dovremmo addentrarci troppo in campo statistico, biologico e matematico. Per illustrare la bontà dei risultati ottenuti da una simulazione, riportiamo i risultati ottenuti da una simulazione stocastica (utilizzando il modello SI) su una popolazione di $n = 100$ individui.

Schema della simulazione

Lo schema di questa simulazione è il seguente. Prima di tutto, fissiamo una quantità di tempo dt (sufficientemente piccola) come passo di rilevazione del numero di infetti. Inoltre è necessario controllare che la quantità $[\alpha k(n-k) + \beta]dt$ sia non superiore ad uno per tutti i valori di $k \in [0; n]$; questo è sempre vero per dt abbastanza piccolo.

Possiamo d'ora in poi considerare i parametri $a := \alpha dt$ e $b := \beta dt$ come costanti fondamentali del modello.

Ora, consideriamo al tempo t uno stato con k infettivi, e dunque $(n-k)$ suscettibili. Al tempo $t + dt$, il sistema ha una probabilità bk di trovarsi nello stato con $k-1$ malati (cioè che si sia verificata una guarigione nell'intervallo tra t e $t + dt$), una probabilità $ak(n-k)$ di trovarsi nello stato con $k+1$ malati (uno dei suscettibili è stato infettato), ed ovviamente una

probabilità $[1 - ak(n - k) - bk]$ che il sistema abbia ancora k malati (sistema inalterato). Naturalmente, questo é a rigore corretto solo nel limite $dt \rightarrow 0$, ma trascureremo egualmente tutti i termini e le probabilità di ordine superiore in dt .

Se si estrae un numero casuale r uniformemente distribuito tra 0 ed 1; l'intervallo $I = [0, 1]$ viene suddiviso in tre sottointervalli (con $I_i \cap I_j = \emptyset, i, j = 1, 2, 3, i \neq j$), ossia

$$I = I_+ \cup I_0 \cup I_- = [0, ak(n - k)] \cup [ak(n - k), 1 - bk] \cup [1 - bk, 1]$$

ed il numero k viene incrementato di uno se $r \in I_+$, diminuito di uno se $r \in I_-$ e lasciato invariato se $r \in I_0$.

Come notiamo, la grandezza degli intervalli varia a seconda del numero degli individui sani ed infetti. Questo perché quando ci sono molti soggetti infettati, é molto piú facile incontrare un individuo ammalato e contrarre l'infezione.

Con l'utilizzo di matlab, abbiamo simulato la diffusione di un'epidemia in una popolazione di 100 individui, di cui 95 sani e 5 infetti al tempo t_0 . La simulazione é stata aggiornata per ogni dt di tempo, per un intervallo di tempo $[0, 10]$ s. Riportiamo il codice utilizzato per la simulazione:

```
alpha= 0.01;    probabilit  di infettarsi -> 1%
beta= 0.5;     probabilit  di guarire -> 50%
dt=0.01;      ogni 10 ms effettuo una rilevazione
t=10;         monitoro la popolazione per 10 s
a=alpha * dt;
b=beta * dt;
n=100;        numero popolazione

num = t / dt;    effettuo 1000 rilevazioni
k=[1:num];      array contenente il numero di infetti ad ogni rilevazione
k(1) = 5;       impostazione numero iniziale di infetti
for i=2:num     le rilevazioni le effettuo ogni i volte dt

    r=rand();
    if ( r < ( a * k(i-1) * (n - k(i-1))))

        k(i)=k(i-1)+1;

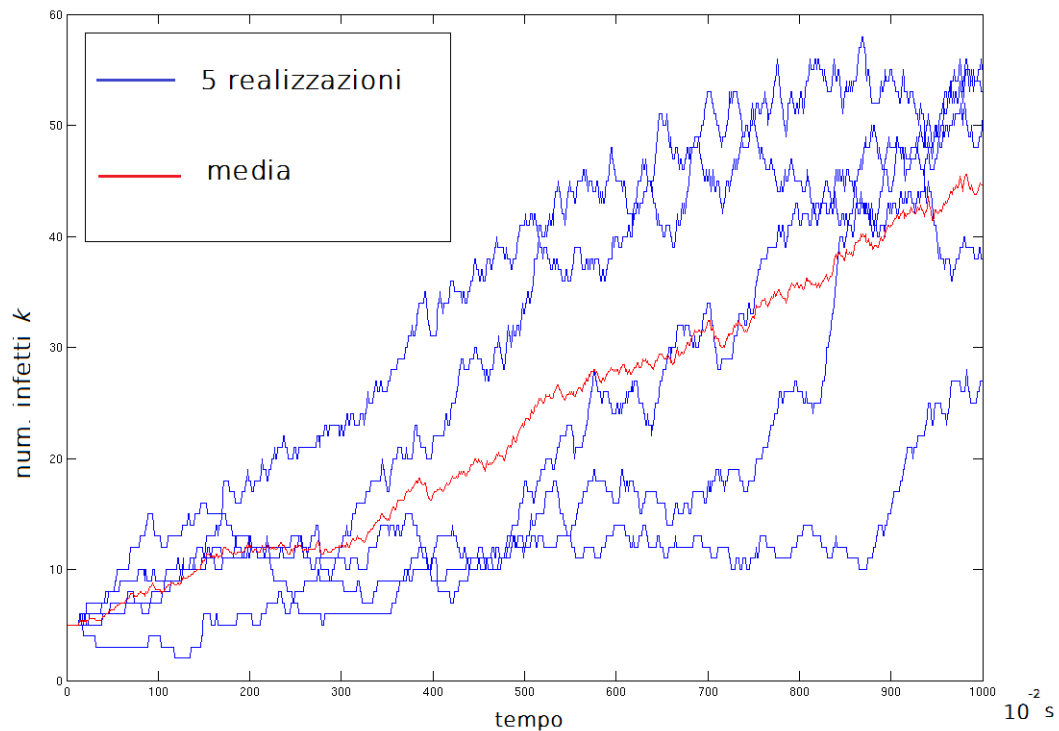
    elseif (r>=(1-b*k(i-1)))

        k(i)=k(i-1)-1;

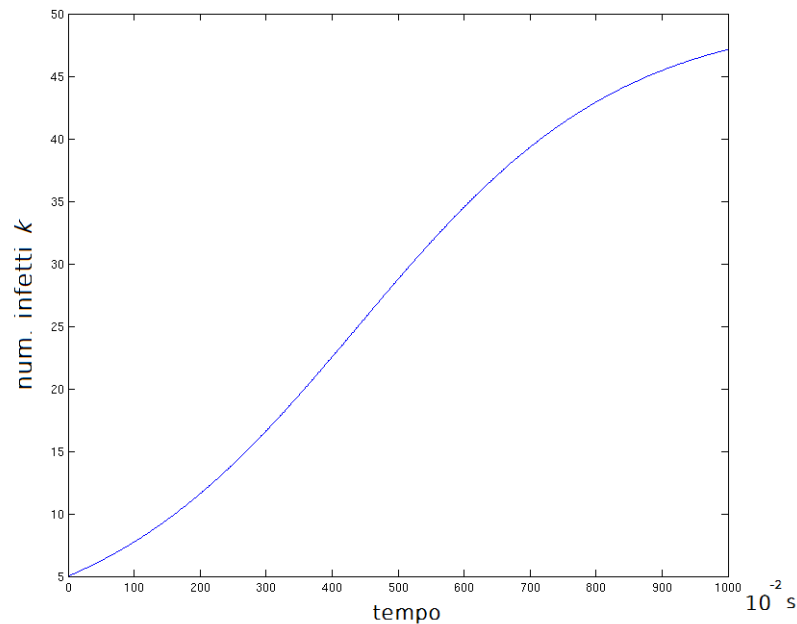
    else k(i) = k(i-1);
end
```

```
end
plot(k);    disegno l'andamento degli infetti
```

in cui $a = \alpha dt$ rappresenta la probabilità che un individuo sano contragga l'infezione nel periodo di tempo dt , $b = \beta dt$ rappresenta la probabilità che un individuo infetti guarisca e sia di nuovo suscettibile all'infezione. Riportiamo il grafico dell'andamento della curva degli infetti k .



All'interno del grafico abbiamo voluto inserire piú simulazioni e ne abbiamo calcolato l'andamento medio. Ora confrontiamo quest'ultimo con l'andamento ricavato dal modello deterministico presentato nel capitolo precedente.



Come notiamo le due curve rappresentative hanno lo stesso andamento. La differenza tra il modello deterministico e quello stocastico é che nel modello stocastico, possiamo ricavare piú informazioni riguardanti i valori istantanei.

L'aspetto rilevante della simulazione non é la situazione asintotica ma lo stato di equilibrio su scale di tempo intermedie, sempre però abbastanza durature per non essere influenzate dal dato iniziale. Quello che viene richiesto é che la probabilità di verificarsi fluttuazioni eccezionali sia minore di uno, tendente allo zero.

Infatti la scelta del periodo di campionamento e del periodo di osservazione é fondamentale se si vuole utilizzare il modello SI, poiché se si prendessero campioni troppo poco frequentemente rischieremmo di vedere innumerevoli fluttuazioni intorno al punto di equilibrio, invece se si prendesse un periodo di osservazione troppo breve, rischieremmo di non riuscire a vedere a che valore tenderá il nostro sistema in tempi molto piú grandi.

4.2 Modello SIR

Come per il modello SI andiamo a scrivere le probabilità di transizione da k a j infetti.

Ricordiamo che il modello SIR suddivide la popolazione in tre classi: *suscettibili*, cioè coloro che possono essere infettati, *infettivi*, cioè coloro che hanno contratto la malattia e sono possono trasmettere l'infezione, *rimossi*, cioè coloro che sono guariti dall'infezione e ora sono immuni ad essa.

Supponiamo di avere k infetti, r rimossi, e quindi $s = n - k - r$ sani al tempo t . La probabilità che nessuno dei suscettibili si ammali é:

$$(1 - \alpha dt)^{sk} = 1 - (\alpha sk)dt + o(dt)$$

con α la probabilità che un individuo sano si ammali. Quindi calcoliamo la generica $P_a(m; s)$ che esprime la probabilità che tra s sani, m si infettino é:

$$P_a(m; s) = \binom{s}{m} (1 - \alpha dt)^{(s-m)k} (\alpha dt)^{mk}$$

La probabilità di guarigione di un individuo infetto é:

$$(1 - \beta dt)^k = 1 - k\beta dt + o(dt)$$

dove β é la probabilità che un individuo ammalato guarisca e diventi immune all'infezione. Per quanto riguarda la formazione di un individuo immune all'infezione, la sua probabilità sar  la stessa di guarigione poich  avevamo assunto che gli individui guariti, diventassero immuni permanentemente all'infezione.

Consideriamo al tempo t uno stato, che indicheremo con $|s, k \rangle$, in cui vi sono s suscettibili e k infetti. Al tempo $t + dt$, il sistema ha una probabilit  $\beta k dt$ di trovarsi nello stato $|s, k - 1 \rangle$ (cio  che si sia verificata una guarigione nell'intervallo tra t e $t + dt$), una probabilit  $\alpha sk dt$ di trovarsi nello stato $|s - 1, k + 1 \rangle$ (cio  che uno dei suscettibili sia stato infettato), ed ovviamente una probabilit  $(1 - \alpha sk dt - \beta k dt)$ di essere nello stato $|s, k \rangle$ (cio  niente   cambiato). Formalizzando quanto detto, avremo:

$$\begin{aligned} P(|s, k \rangle \rightarrow |s - 1, k + 1 \rangle) &= \alpha sk dt \\ P(|s, k \rangle \rightarrow |s, k - 1 \rangle) &= \beta k dt \\ P(|s, k \rangle \rightarrow |s, k \rangle) &= 1 - (\alpha sk + \beta k) dt \end{aligned}$$

4.2.1 Simulazione stocastica

Per il processo stocastico, relativo al modello SIR,   preferibile l'utilizzo del metodo della simulazione poich  lo stato del sistema   identificato dal numero S di individui suscettibili e dal numero I di infettivi. Questa dipendenza dai parametri S ed I comporta la costruzione di una matrice di transizione che cresce quadraticamente rispetto al numero di individui della popolazione. Prima di presentare lo schema utilizzato per la simulazione, ricordiamo che per la natura stessa del sistema, l'infezione   destinata a estinguersi e di conseguenza le domande che ci poniamo sono diverse da quelle che ci siamo fatti nel modello SI, cio  non ci interessa sapere il comportamento di equilibrio del sistema. In questo modello potrebbe essere interessante sapere per

quanto tempo il numero di infetti $I(t)$ si mantiene al di sopra di una certa soglia I_{ref} fissata.

Schema per la simulazione

Come detto prima lo stato del sistema sarà determinato da due interi s e k che rappresentano il numero di suscettibili e quello di infetti (il numero di rimossi si ricava dalla relazione $n = s - k - r$). Fissiamo una quantità di tempo dt (sufficientemente piccola) come passo di rilevazione del numero di infetti e suscettibili. È necessario controllare che la quantità $[sk\alpha + \beta]dt$ sia non superiore ad uno per tutti i valori di s e k in $[0, n]$; questo è sempre vero per dt abbastanza piccolo (infinitesimale). Le costanti $a := \alpha dt$ e $b := \beta dt$ saranno le costanti fondamentali del modello.

Si estra un numero casuale r uniformemente distribuito tra 0 e 1; l'intervallo $I = [0, 1]$ viene diviso in tre sottointervalli (valgono i vincoli esposti in precedenza),

$$I = I_+ \cup I_0 \cup I_- = [0, ask] \cup [ask, 1 - bk] \cup [1 - bk, 1]$$

e lo stato $|s, k >$ diviene $|s - 1, k + 1 >$ se $r \in I_+$, diviene $|s, k - 1 >$ se $r \in I_-$, e resta invariato se $r \in I_0$.

Con l'utilizzo di matlab, siamo andati a simulare la diffusione di un'epidemia in una popolazione di 100 individui, di cui 95 sani, 5 infetti e 0 rimossi al tempo t_0 . La simulazione è stata aggiornata per ogni unità di tempo, per un periodo di tempo pari a 10 s. Riportiamo il codice utilizzato per la simulazione:

```
alpha= 0.05;    probabilità di infettarsi → 5%
beta= 0.1;     probabilità di guarire → 10%
dt=0.001;     ogni 1 ms effettuo una rilevazione
t=10;         monitoro la popolazione per 10 s
a=alpha * dt;
b=beta * dt;
n=100;        numero popolazione

num = t / dt;    effettuo 10000 rilevazioni
r=[1:num];      array contenente il numero di rimossi ad ogni rilevazione
k=[1:num];      array contenente il numero di infetti ad ogni rilevazione
s=[1:num];      array contenente il numero di suscettibili ad ogni
rilevazione
k(1) = 5;       impostazione numero iniziale di infetti
s(1)=n - k(1);  impostazione numero iniziale di suscettibili
r(1)=0;         impostazione numero iniziale di rimossi
for i=2:num     le rilevazioni le effettuo ogni i volte dt
```

```

c=rand();
if ( c < ( a * k(i-1) * s(i-1)))

    k(i)=k(i-1)+1;
    s(i)=s(i-1)-1;
    r(i)=r(i-1);

elseif ( c >= ( 1 - b * k(i - 1)))

    r(i)=r(i-1)+1;
    k(i)=k(i-1)-1;
    s(i)=s(i-1);

else k(i) = k(i-1);

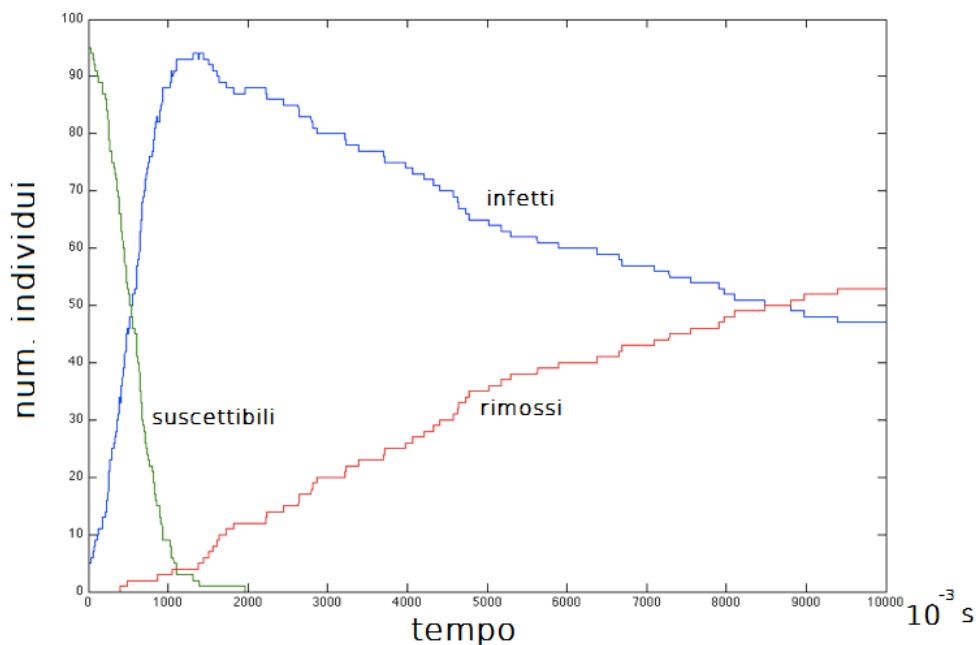
    s(i)=s(i-1);
    r(i)=r(i-1);

end

end
plot(k);    disegno l'andamento del numero di infetti
hold all
plot(s);    disegno l'andamento del numero di suscettibili
hold all
plot(r);    disegno l'andamento del numero di rimossi

```

in cui $a = \alpha dt$ rappresenta la probabilità che un individuo sano contragga l'infezione nel periodo di tempo dt , $b = \beta dt$ rappresenta la probabilità che un individuo infetti guarisca e diventi immune permanentemente. Riportiamo il grafico dell'andamento delle tre classi nel modello SIR.



In verde é rappresentato l'andamento della classe dei sani (S), in blu quello degli infetti (I) e in rosso quello dei rimossi (R).

4.3 Potenziale effettivo per un processo stocastico

Per dare una descrizione migliore del processo stocastico, abbiamo bisogno di descrivere sia il suo valore atteso che la sua dispersione intorno alla media. Questo viene fatto tramite l'utilizzo del *potenziale effettivo*.

Consideriamo il valore atteso delle variazioni δk dei malati nel periodo di tempo dt . Definiamo quindi $\langle \delta k \rangle$ come :

$$\langle \delta k \rangle = P(k \rightarrow k + 1) - P(k \rightarrow k - 1) + o(dt)$$

e dunque la quantità $K(t) := \langle k(t) \rangle$ evolve in prima approssimazione (trascurando la dispersione) secondo la

$$\frac{dK}{dt} = [\alpha(n - K) - \beta]K := f(K).$$

A questo punto possiamo determinare il potenziale effettivo come una funzione $V(K)$ tale che

$$f(K) = -\frac{dV(K)}{dK}.$$

4.3. POTENZIALE EFFETTIVO PER UN PROCESSO STOCASTICO 33

Ora andiamo ad analizzare il potenziale effettivo di uno dei due modelli epidemiologici presentati, ossia il modello SI. Sfruttando il sistema di equazioni descrittive del modello e successivamente integrando la funzione dK in dt , abbiamo che il suo potenziale é:

$$V(K) = \frac{k^3}{3} - \left(\frac{\alpha n - \beta}{2} \right) k^2 + c_0$$

dove la costante c_0 puó essere fissata in modo che $V(K_0) = 0$, con $K_0 = n - \beta/\alpha$ definito come il valore di equilibrio del sistema. Questa condizione é verificata quando

$$c_0 = \frac{(\alpha n - \beta)^3}{6\alpha^2}.$$

Quindi avremo che il potenziale effettivo del modello SI é

$$V(k) = - \frac{(\beta - \alpha(n - k))^2(8\beta - \alpha(n + 2k))}{6\alpha^2}$$

Come detto all'inizio della sezione, daremo una descrizione piú dettagliata della distribuzione dell'infezione, per questo andiamo a calcolare la varianza intorno alla media appena trovata. Ricordiamo la definizione di varianza come

$$\sigma^2(k) = \langle (\delta k)^2 \rangle - \langle \delta k \rangle^2$$

quindi andiamo a calcolare

$$\langle \delta k \rangle^2 = (\beta - \alpha n)^2 k^2 + 2\alpha(\beta - \alpha n)k^3 + \alpha^2 k^4.$$

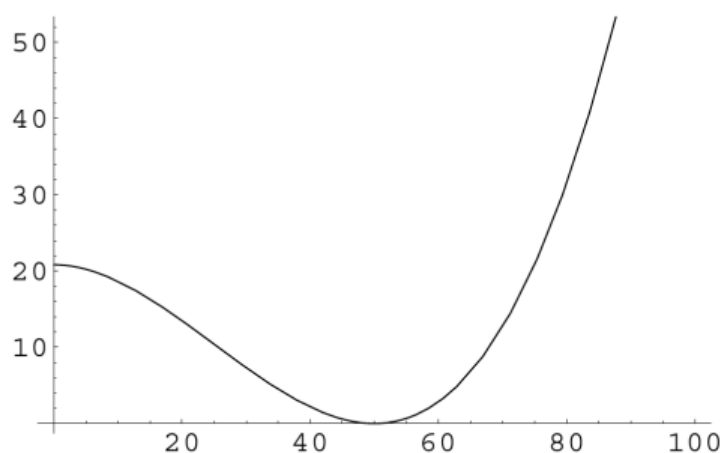
e successivamente

$$\langle (\delta k)^2 \rangle = P(k \rightarrow k + 1) + P(k \rightarrow k - 1) = (\beta + \alpha n)k - \alpha k^2$$

Ricavati i due termini, possiamo calcolare la varianza che risulta essere

$$\sigma^2(k) = (\beta + \alpha n)k - [\alpha + (\beta - \alpha n)^2]k^2 - 2\alpha(\beta - \alpha n)k^3 - \alpha^2 k^4.$$

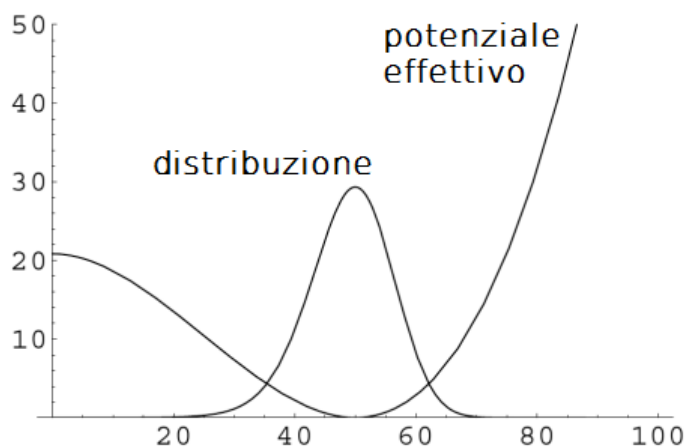
Possiamo ora adottare una approssimazione migliore di quella usata in precedenza, usando le informazioni ottenute, ossia il fatto che nel processo che descrive l'evoluzione della $k(t)$ tra il tempo t ed il tempo $t + dt$, la quantità δk ha media $f[k(t)]$ e varianza (o meglio dispersione) $\sigma^2[k(t)]$.



Con i dati ricavati, possiamo approssimare il processo stocastico ad una distribuzione di Boltzmann; quindi avremo:

$$P(x) = \frac{1}{Z} \exp[-V(x)/\sigma_0] \quad (4.1)$$

con $\sigma_0 = \sigma(k_0)$, dove k_0 rappresenta il minimo del potenziale effettivo e Z rappresenta il fattore di normalizzazione. Con fattore di normalizzazione intendiamo quel valore utilizzato per fare in modo che integrando $P(x)$ da $-\infty$ a $+\infty$ otteniamo 1.



Per rendere confrontabili le due curve, si è moltiplicata la distribuzione per un fattore $F = 500$, così da poter avere delle scale simili.

Come possiamo vedere facilmente dal grafico soprastante, notiamo che il potenziale effettivo diventa nulla quando l'epidemia è raggiunge il picco massimo di infetti. Quindi, con il solo utilizzo del *potenziale effettivo*, in prima

4.3. POTENZIALE EFFETTIVO PER UN PROCESSO STOCASTICO 35

approssimazione, possiamo dare una stima di quando l'epidemia avrà il suo picco massimo di infetti.

Bibliografia

- [1] G. Gaeta, *Modelli matematici in biologia*, Springer Verlag, 2009.
- [2] Dimitri P. Bertsekas, John N. Tsitsiklis, *Introduction to probability*, Athena Scientific, 2008.
- [3] Cristina Ronconi, *Appunti di Geometria*, Univer Editrice, 2009