

Approccio deterministico e stocastico allo studio di un modello epidemiologico

Michela Pizzolo

Matricola: 611946

Relatore: Prof. Mauro Bisiacco

Università di Padova

Dipartimento di Ingegneria dell'Informazione

Laurea in Ingegneria dell'Informazione

Indice

1	Introduzione	5
2	Il Modello SIR	7
2.1	Dati realmente disponibili	11
2.2	Un caso particolare: nessuna immunità	13
3	Il Modello SIR Probabilistico	17
3.1	I processi stocastici e i processi di Poisson	17
3.2	Descrizione del modello	19
3.3	La costruzione di Sellke	21
3.4	Il caso Markoviano	21
4	Simulazioni del Modello SIR	23
4.1	Parametri deterministici per il controllo di un'epidemia	23
4.2	Simulazioni stocastiche del modello SIR	27

Capitolo 1

Introduzione

Argomento del presente elaborato è lo studio della dinamica generale di diffusione di un'infezione all'interno di una popolazione chiusa attraverso l'utilizzo di un modello molto famoso in campo epidemiologico: il modello SIR. Si tratta di un modello compartimentale che suddivide la popolazione in tre classi, dette appunto compartimenti, per poi esprimere le relazioni tra di esse attraverso delle equazioni differenziali. Il nome SIR deriva dal modo in cui è comune etichettare i tre compartimenti: S, classe dei suscettibili; I, classe degli infetti; R, classe dei rimossi.

Esistono diversi approcci allo studio di un modello epidemiologico. Uno che ha senz'altro ricevuto molte attenzioni in letteratura è l'approccio *deterministico*, in grado di esprimere l'evoluzione di una malattia usando parametri fissi per tutta la durata dell'esposizione. Nel capitolo 2 andremo a studiare il modello deterministico SIR introdotto dagli studiosi W.O. Kermack e A.G. McKendrick, stabilendone le quantità fondamentali e ricavando le principali relazioni tra le variabili che identificano individui sani e malati.

Il punto di vista deterministico si rivela certamente uno strumento potente nell'analisi di un'epidemia, ma diversi aspetti suggeriscono l'opportunità di introdurre nel modello una componente probabilistica, giungendo così a un approccio *stocastico*. Difatti, il modo più naturale di trattare un'infezione reale è quello di parlare *probabilità* che un individuo venga contagiato, in quanto non si potrà avere la certezza che la malattia sia trasmessa o meno. I due approcci rimangono comunque strettamente legati: mentre il modello deterministico è particolarmente efficace quando sono in gioco grandi numeri, quello stocastico riesce a dare una visione accurata dell'epidemia anche negli altri casi, soprattutto negli istanti iniziali e finali, quando la maggior parte della popolazione deve ancora essere colpita oppure è già guarita in modo permanente; e su grandi numeri, buona parte dell'analisi di un modello stocastico consiste nel capire a quale modello deterministico esso converga.

In altre parole, il risultato dell'analisi deterministica può essere interpretato come l'analisi degli *andamenti medi* dell'approccio stocastico. Nel capitolo 3 daremo quindi nozioni sui possibili approcci al complesso modello SIR stocastico, fornendo in primo luogo le basi di Teoria della Probabilità essenziali per comprendere l'argomento. Nel capitolo 4, infine, riprenderemo i concetti analizzati nei primi due capitoli per fornire una visione numerica, conseguente allo sviluppo di alcune simulazioni con relativi grafici, dei due approcci visti.

Capitolo 2

Il Modello SIR

Il primo modello matematico completo di un certo interesse per lo studio delle malattie infettive fu quello ideato nel 1927 da Kermack e McKendrick. Si trattava di un modello *deterministico* volto a descrivere lo sviluppo dell'infezione all'interno di un gruppo chiuso di individui, suddivisi in tre gruppi come indicato graficamente in figura 2.1.

- **S**, la classe dei suscettibili, ovvero gli individui in grado di contrarre la malattia e diventare infetti;
- **I**, la classe degli infetti, ovvero coloro che hanno già contratto la malattia e sono in grado di trasmetterla ad altri individui;
- **R**, la classe dei rimossi, ovvero coloro che, dopo aver contratto la malattia, sono morti oppure sono guariti diventando permanentemente immuni, oppure sono stati isolati rendendo impossibile la trasmissione dell'infezione.

Definiamo $S(t)$, $I(t)$ e $R(t)$ rispettivamente il numero di suscettibili, di infetti e di rimossi al tempo t ; introduciamo inoltre le costanti positive α e β rappresentanti rispettivamente la probabilità di infezione¹ e di guarigione (o morte).

Ipotizzando che la popolazione totale rimanga sempre in quantità costante pari alla somma dei tre compartimenti, ossia

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

¹Il modello è deterministico, quindi sarebbe più logico parlare di frequenza e non di probabilità. Tuttavia, anche in vista di successivi modelli probabilistici, preferiamo anche in questo contesto parlare di probabilità, molto più realistica rispetto ad una diffusione infettiva di tipo deterministico.

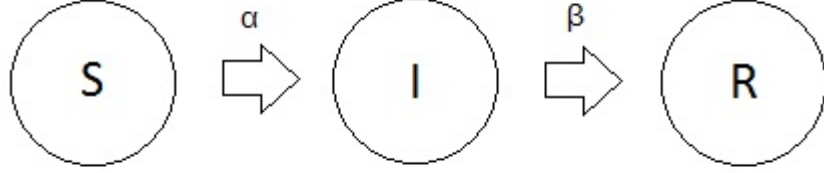


Figura 2.1: Grafo compartimentale del modello SIR.

il modello viene definito nel seguente modo:

$$\begin{aligned} \frac{dS(t)}{dt} &= -\alpha S(t)I(t) \\ \frac{dI(t)}{dt} &= \alpha S(t)I(t) - \beta I(t) \\ \frac{dR(t)}{dt} &= \beta I(t) \end{aligned} \quad (2.1)$$

con stato iniziale pari a $(S(0), I(0), R(0)) = (S_0, I_0, R_0)$, dove valgono

$$I(0) = I_0 \in \mathbb{R}^+$$

stabilendo così che all'istante zero di osservazione del fenomeno sia già presente un gruppo di infetti, altrimenti risolvendo le equazioni si troverebbe $I(t) = 0 \forall t$, e quindi l'infezione non si manifesterebbe;

$$R(0) = R_0 \in \mathbb{R}^+$$

considerando la possibilità che all'interno della popolazione esistano individui intrinsecamente immuni al contagio (ad esempio per cause genetiche o ambientali); infine,

$$S(0) = S_0 = N - I_0 - R_0 \in \mathbb{R}^+$$

ottenuta sfruttando la precedente ipotesi di popolazione totale costante.

Ipotizzando che il contagio avvenga per contatto diretto tra individui, il numero di nuovi infetti sarà proporzionale al numero di incontri tra i membri

della classe S (suscettibili) e della classe I (infetti), ovvero sarà proporzionale al prodotto $S(t)I(t)$, moltiplicato per un fattore α pari alla probabilità che la malattia sia effettivamente trasmessa. Il fattore non lineare $\alpha S(t)I(t)$ suggerisce che l'infezione si propaga con un tasso elevato quando sono presenti *sia* molti suscettibili *sia* molti infetti. Naturalmente, i nuovi infetti usciranno dal compartimento dei suscettibili ed entreranno in quelli degli infetti. Nel caso generale, inoltre, gli infetti avranno una probabilità β di uscire dalla relativa classe, con ciò intendendo sia l'effettiva guarigione dalla malattia, sia la morte. Le persone che guariscono (o muoiono per la malattia) nell'unità di tempo considerata entreranno nella classe R (rimossi).

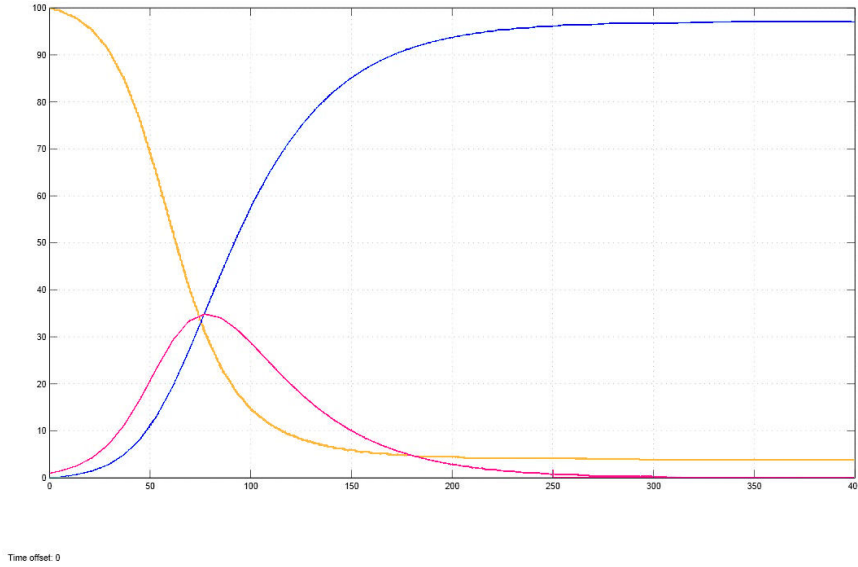


Figura 2.2: Simulazione del modello (2.1) per valori di $\alpha = 0.001$, $\beta = 0.03$, $S_0 = 100$, $I_0 = 1$, $R_0 = 0$. In ascissa è riportato il tempo, mentre l'ordinata specifica l'andamento di $S(t)$ (curva gialla), $I(t)$ (curva rosa) e $R(t)$ (curva blu). Si può notare che, per $t \rightarrow \infty$, $I(t) \rightarrow 0$ ma $S(t) \rightarrow S_1 \neq 0$: alcuni suscettibili non sono pertanto soggetti a infezione.

Come già sottolineato, possiamo esprimere $R(t)$ in funzione dei parametri N (costante), $S(t)$ e $I(t)$. Tralasciando quindi la terza equazione del sistema (2.1) e concentrandoci solo sulle prime due, notiamo che vale sempre $\frac{dS(t)}{dt} < 0$, e che se $S(t) < \frac{\beta}{\alpha}$ allora anche $\frac{dI(t)}{dt} < 0$. Questo fatto può essere interpretato assumendo l'esistenza di un valore di soglia $\gamma = \frac{\beta}{\alpha}$: se S_0 è maggiore di tale

valore, l'infezione si propaga fino a che il valore di $S(t)$ non scende sotto quello di soglia; se fin dall'inizio S_0 non supera la soglia, il numero di infetti non aumenterà.

Grazie alla semplicità delle equazioni del modello (2.1), siamo in grado di riscriverle esplicitamente in forma parametrica, allo scopo di esprimere $I(t)$ in funzione di $S(t)$: abbiamo

$$\frac{dI(t)}{dS(t)} = \frac{\alpha S(t)I(t) - \beta I(t)}{-\alpha S(t)I(t)} = -1 + \frac{\beta}{\alpha S(t)}$$

Moltiplicando entrambi i membri per $dS(t)$ e integrando, otteniamo

$$I(S) = c_0 - S + \frac{\beta}{\alpha} \log[S] \quad (2.2)$$

dove c_0 è una costante esprimibile in funzione dei valori iniziali S_0 e I_0 :

$$c_0 = I_0 + S_0 - \frac{\beta}{\alpha} \log(S_0) .$$

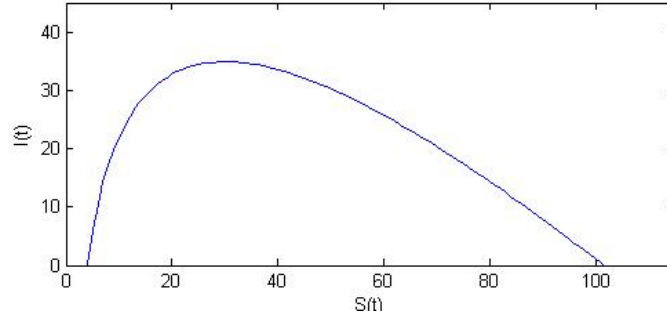


Figura 2.3: Grafico della (2.2) per valori di $\alpha = 0.001$, $\beta = 0.03$, $S_0 = 100$, $I_0 = 1$. In ascissa è riportato $S(t)$, mentre l'ordinata mostra l'andamento di $I(t)$. La curva va pensata percorsa da destra a sinistra al crescere di t , come segue da $\frac{dS(t)}{dt} \leq 0$. Notiamo che $I(t)$ parte da zero e si annulla nuovamente in un tempo t_1 finito, quindi per $t \rightarrow \infty$ vale $I(t) \rightarrow 0$. D'altra parte $S(t_1) \neq 0$, dunque $S(t) \rightarrow S_1 = S(t_1) \neq 0$ per $t \rightarrow \infty$: come nella figura (2.2), anche in questo caso vediamo che alcuni suscettibili non sono soggetti a infezione.

2.1 Dati realmente disponibili

Supponendo di voler studiare l'infezione a partire da dati reali e non sperimentali, molto spesso troviamo la necessità di avere a disposizione un calcolo in funzione di $R(t)$: difatti, è molto probabile che le informazioni sulla malattia consistano sostanzialmente nel numero di malati che si rivolgono al sistema sanitario. Assumendo di avere a disposizione medicinali validi ed efficaci ed eventualmente di disporre della possibilità di isolare i malati più gravi, possiamo considerare tutte queste persone come rimosse: sotto tali ipotesi infatti verranno subito guarite dai farmaci oppure messe in condizioni tali da non poter infettare altri individui. Ritornando a considerare il sistema nella sua interezza e osservando che $\frac{dS}{dR} = -\frac{S(t)}{\gamma}$, si ricava che

$$S(t) = S_0 e^{-\left(\frac{R(t)-R_0}{\gamma}\right)} \quad (2.3)$$

dove, per $R_0 = 0$, si ottiene semplicemente

$$S(t) = S_0 e^{-\frac{R(t)}{\gamma}}$$

Procediamo ora a ricavare una soluzione, anche approssimata, dell'equazione appena ottenuta. Utilizzando il sistema di equazione del modello e ricordando che $R(t) = N - S(t) - I(t)$, possiamo scrivere che

$$\frac{dR(t)}{dt} = \beta I(t) = \beta(N - R(t) - S(t)) \quad (2.4)$$

D'altra parte, grazie all'equazione (2.3), siamo in grado di esprimere $S(t)$ in funzione di $R(t)$. Operando la seguente trasformazione sulla seconda parte del prodotto:

$$(N - R(t) - S_0 e^{-\left(\frac{R(t)-R_0}{\gamma}\right)}) = (N - R(t) - c_0 e^{-\frac{R(t)}{\gamma}})$$

possiamo quindi riscrivere la variazione di $R(t)$ come:

$$\frac{dR(t)}{dt} = \beta(N - R(t) - c_0 e^{-\frac{R(t)}{\gamma}}) \quad (2.5)$$

La (2.5) è chiaramente un'equazione trascendente, ovvero un'equazione composta da funzioni trascendenti nell'incognita $R(t)$ e, pertanto, non riconducibile ad un'espressione polinomiale. Per questo motivo non può essere risolta direttamente, ma è necessario utilizzare i metodi dell'analisi numerica. In questo caso notiamo che, a condizione che il fattore $\frac{R(t)}{\gamma}$ sia piccolo

(come effettivamente spesso avviene), possiamo approssimare l'esponenziale con il suo sviluppo in serie di Taylor. Ricordando che, per $|x| \ll 1$,

$$e^{-x} \simeq 1 - x + \frac{x^2}{2}$$

e ponendo per semplicità $R_0 = 0$, riscriviamo la (2.4) come

$$\frac{dR(t)}{dt} = \beta[N - R(t) - S_0(1 - \frac{R(t)}{\gamma} + \frac{(\frac{R(t)}{\gamma})^2}{2})]$$

e successivamente, riordinando i termini, come

$$\frac{dR(t)}{dt} = \beta[(N - S_0) + (\frac{S_0}{\gamma} - 1)R(t) - \frac{S_0}{2\gamma^2}R(t)^2] \quad (2.6)$$

La soluzione dell'equazione (2.6) può essere calcolata esattamente. Infatti, in termini generali, l'equazione

$$\frac{dx}{dt} = A + Bx + Cx^2$$

ha soluzione

$$x(t) = -\frac{B}{2C} + \frac{\sqrt{-B^2 + 4AC}}{2C} \tan \left[\frac{1}{2}\sqrt{-B^2 + 4AC}t + k_0 \right]$$

Il valore k_0 è una costante legata al valore iniziale $x(0) = x_0$ ed esprimibile come

$$k_0 = \arctan\left[\frac{B + 2Cx_0}{\sqrt{-B^2 + 4AC}}\right].$$

Sfruttando la formula generale, siamo perciò in grado di ottenere la soluzione dell'equazione (2.6) in funzione di termini costanti ϕ e k definiti in questo modo:

$$\phi = \frac{1}{k} \operatorname{arctanh}\left[\frac{S_0}{\gamma} - 1\right]$$

$$k = \sqrt{\left(\frac{S_0}{\gamma} - 1\right)^2 + 2\frac{S_0}{\gamma^2}(N - S_0)}$$

La soluzione della (2.6) risulta

$$R(t) = \frac{\alpha^2}{S_0} \left[\left(\frac{S_0}{\gamma} - 1 \right) + k \tanh\left(\frac{k\beta t}{2} - \phi\right) \right].$$

Differenziando infine quest'ultima otteniamo il tasso di rimozione, ovvero i nuovi rimossi in funzione del tempo:

$$\frac{dR(t)}{dt} = \frac{\beta k^2 \gamma^2}{2S_0 \cosh^2[\frac{\beta k t}{2} - \phi]} . \quad (2.7)$$

Alcuni dei parametri da cui dipende l'equazione (2.7), ovvero β , γ e S_0 (tramite k e ϕ), non sono generalmente noti in un'epidemia reale; l'approccio usato di frequente è quello di determinarli da un punto di vista statistico. Come già accennato in precedenza, i calcoli appena effettuati sono validi solamente a patto che $\frac{R(t)}{\gamma} \ll 1$. Quando questa condizione non è verificata, è necessario procedere col calcolo numerico; anche qui però, grazie a misurazioni reali non riportate in questa sede, ci si può rendere conto di come nel caso generale le equazioni sopra esposte descrivano con sufficiente accuratezza il fenomeno.

2.2 Un caso particolare: nessuna immunità

Prendiamo ora in esame un caso più semplice rispetto a quello finora considerato: il modello SI, adatto per quelle infezioni non letali che non prevedono alcuna immunità, sia essa temporanea o permanente. La popolazione all'istante zero è composta da un nucleo di individui infetti (classe I), mentre tutti i rimanenti sono considerati suscettibili (classe S); con l'evolversi dell'epidemia, i suscettibili possono ammalarsi, nel qual caso entreranno a far parte delle file degli infetti. Questi ultimi, al momento della guarigione, non diverranno rimossi (classe R) come nel caso precedente, ma al contrario ritorneranno ad essere suscettibili e quindi passibili di ammalarsi nuovamente. È il caso di infezioni come il semplice raffreddore, o le malattie veneree. Questa situazione viene descritta dal modello

$$\begin{aligned} \frac{dS(t)}{dt} &= -\alpha S(t)I(t) + \beta I(t) \\ \frac{dI(t)}{dt} &= \alpha S(t)I(t) - \beta I(t) \end{aligned} \quad (2.8)$$

Poichè $S(t) + I(t) = N$, possiamo condensare le due equazioni in una sola:

$$\frac{dS(t)}{dt} = \beta N - (\alpha N + \beta)S(t) + \alpha S^2(t)$$

Tale equazione presenta come soluzione

$$S(t) = \frac{\beta e^{\alpha N t} + c_0 N e^{\beta t}}{\alpha e^{\alpha N t} + c_0 e^{\beta t}} \quad (2.9)$$

in cui c_0 è una costante dipendente dalle condizioni iniziali e data da

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

ricordando che $\gamma = \frac{\beta}{\alpha}$.

È possibile riscontrare che in tale modello, per $I_0 \neq 0$, il numero di suscettibili e di infetti si assesta per $t \rightarrow \infty$ ad un valore costante non nullo. Di contro, quando I_0 è nullo, l'infezione non si diffonde e il numero di infetti resta ovviamente a zero. Si può vedere che la soluzione stazionaria a cui il sistema tende corrisponde a $S(t) = \gamma$: difatti, raccogliendo $I(t)$ e α dalla prima delle due equazioni che descrivono il sistema, si può ricavare

$$\frac{dS(t)}{dt} = \alpha I(t)(\gamma - S(t))$$

in cui, sapendo che α e $I(t)$ sono quantità sempre positive ed escludendo il caso banale $I(t) = 0$, notiamo che $S(t)$ cresce quando è minore di γ e decresce quando è maggiore.

Tali risultati possono essere visualizzati nei grafici 2.4 e 2.5. Il primo rappresenta l'evoluzione di un modello SI per un'infezione caratterizzata da un valore di S_0 decisamente superiore a γ . In accordo con l'analisi effettuata, $S(t)$ subisce un decremento fino ad assestarsi, all'infinito, al valore $\gamma = 200$. Il secondo grafico manifesta invece un comportamento opposto: in questo caso, poichè $S_0 < \gamma$, si assiste ad un aumento di $S(t)$, che cresce fino a raggiungere il valore all'equilibrio $\gamma = 200$. Si noti come, mantenendo costante $N = 1000$ ma partendo da dati iniziali molto diversi per S_0 e I_0 , ci siano nei due casi gli stessi valori per $S(t)$ e $I(t)$ all'equilibrio.

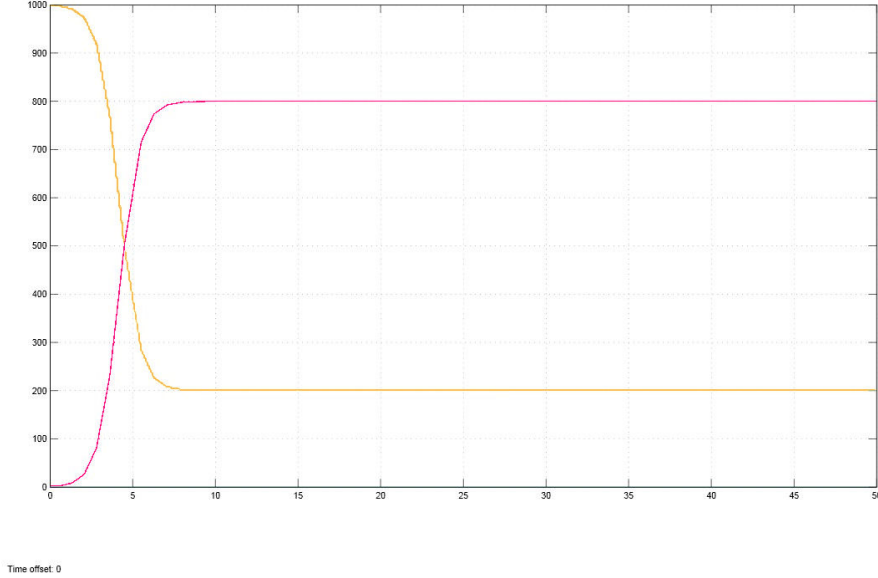


Figura 2.4: Grafico della (2.8) per valori di $\alpha = 0.002$, $\beta = 0.4$, $S_0 = 999$, $I_0 = 1$. La curva gialla rappresenta l'andamento di $S(t)$, quella rosa l'andamento di $I(t)$. Si può notare come il numero di suscettibili *decresca* col passare del tempo (indicato in ascissa) e tenda all'infinito al valore $\gamma = \frac{\beta}{\alpha} = 200$, mentre $I(t)$ si assesta al valore $N - S(\infty) = 1000 - 200 = 800$.

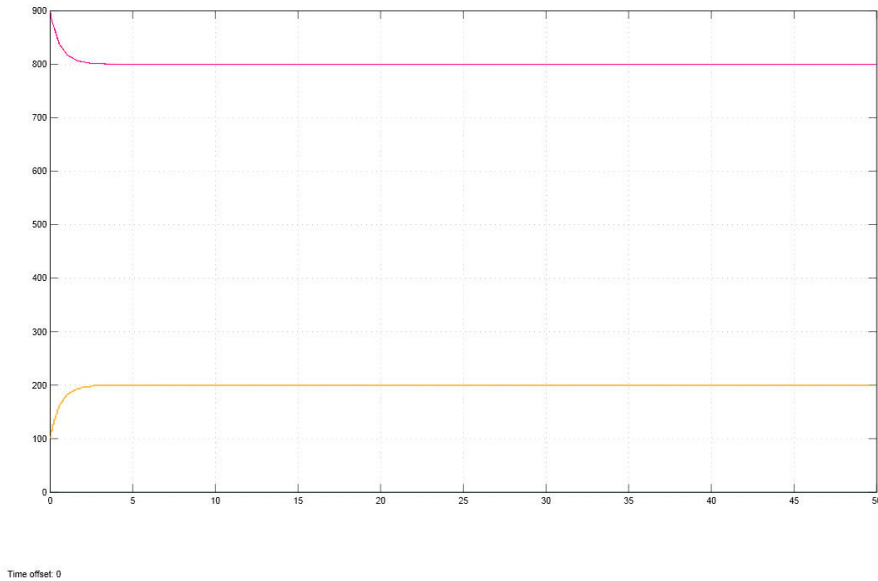


Figura 2.5: Grafico della (2.8) per valori di $\alpha = 0.002$, $\beta = 0.4$, $S_0 = 100$, $I_0 = 900$. La curva gialla rappresenta l'andamento di $S(t)$, quella rosa l'andamento di $I(t)$. Si può notare come il numero di suscettibili *aumenti* col passare del tempo (indicato in ascissa) e tenda all'infinito al valore $\gamma = \frac{\beta}{\alpha} = 200$, mentre $I(t)$ si assesta al valore $N - S(\infty) = 1000 - 200 = 800$.

Capitolo 3

Il Modello SIR Probabilistico

Prima di procedere all'analisi in chiave probabilistica di un modello SIR, riteniamo opportuno introdurre brevemente il concetto di processo stocastico e in particolare di processo di Poisson, per consentire un'adeguata comprensione degli argomenti che tratteremo in seguito.

3.1 I processi stocastici e i processi di Poisson: qualche definizione

Introduciamo ora alcuni concetti che ci aiuteranno a definire i processi stocastici.

Definizione 3.1. Sia Ω un insieme non vuoto, $\Omega \neq \emptyset$. Una σ -algebra F è una famiglia di sottoinsiemi di Ω tale che:

- $\emptyset \in F$;
- Se $S \in F$ allora $S^c := (\Omega \setminus S) \in F$;
- Per ogni successione $(S_n)_{n \in \mathbb{N}}$ di elementi di F , $\bigcup_{n=1}^{\infty} S_n \in F$.

Definizione 3.2. Una *misura sulla σ -algebra F di Ω* è un'applicazione

$$P : F \rightarrow [0, \infty]$$

tale che:

- $P(\emptyset) = 0$;

- Per ogni successione $(S_n)_{n \in \mathbb{N}}$ di elementi di F , a due a due disgiunti, vale:

$$P\left(\bigcup_{n \geq 1} S_n\right) = \sum_{n \geq 1} P(S_n)$$

Se $P(\Omega) < \infty$, diciamo che P è una *misura finita*. Inoltre, se vale:

- $P(\Omega) = 1$, allora diciamo che P è una *misura di probabilità*.

Definizione 3.3. Uno *spazio di probabilità* è una terna (Ω, F, P) con F σ -algebra su Ω e P misura di probabilità su F .

Definizione 3.4. Una *variabile aleatoria* (v.a.) sullo spazio di probabilità (Ω, F, P) è una funzione misurabile X da Ω a valori in $\mathbb{R}^N$, ossia una funzione

$$X : \Omega \rightarrow \mathbb{R}^N \text{ t.c. } X^{-1}(H) \in F, H \in B$$

con B detta σ -algebra dei Borelliani. Nel caso $N = 1$, X è detta *v.a. reale*; quando $N > 1$, si parla invece di *vettore aleatorio*.

Definizione 3.5. Con il nome di *processo stocastico* ci riferiamo ad un modello che permette di studiare un fenomeno aleatorio che evolve nel corso del tempo. Consideriamo:

- Uno spazio di probabilità (Ω, F, P) ;
- Uno spazio misurabile (E, B) , dove E è chiamato *insieme degli stati* (del processo);
- Una famiglia $(X_t)_{t \in T}$ di variabili aleatorie definite su (Ω, F, P) a valori in E .

L'insieme T degli indici è chiamato *spazio dei tempi*. Per $\omega \in \Omega$ e $t \in T$ la quantità $X_t(\omega)$ è chiamata *stato del processo all'istante t* . Dato $\omega \in \Omega$, l'applicazione che ad ogni t in T fa corrispondere $X_t(\omega)$, lo stato del processo all'istante t , è chiamata *traiettoria di ω* .

Si distinguono più tipi di processi a seconda che T e E siano a valori continui o discreti. Se $T \subset \mathbb{N}$, diciamo che il processo è a *tempi discreti*. Se $T = [0, 1]$, o coincide con $\mathbb{R}$, il processo si dice a *tempi continui*. A seconda che E sia finito, numerabile o continuo, si dice che il processo rispettivamente ha un numero finito di stati, uno spazio di stati numerabili o uno spazio di stati continuo. Quando $T = \mathbb{N}$ (tempo discreto), il processo stocastico è descritto da una serie $X_0, X_1, X_2, \dots$ di variabili aleatorie a valori in (E, B) . Se E è finito o numerabile, queste variabili aleatorie sono necessariamente variabili

aleatorie discrete. Consideriamo un processo a tempi discreti, o ancora una serie di variabili aleatorie (X_n) ($n \geq 0$). Per dare in modo chiaro la nozione di processo all'istante n , introduciamo l'insieme di eventi che si producono fino all'istante n (incluso), come l'insieme rappresentato dal vettore $(X_0, \dots, X_n)$.

Ci proponiamo di studiare un fenomeno aleatorio in istanti casuali, chiamati *istanti* $-\tau$. Nelle applicazioni, questi sono i momenti in cui si verificano certi eventi specifici, come possono essere ad esempio il tempo in cui un materiale radioattivo emette particelle, il tempo con cui richieste da computer periferici arrivano a un server o il tempo in cui delle automobili arrivano a una stazione di servizio. Chiameremo questi eventi *i salti*, che si verificano quindi nei suddetti *istanti* $-\tau$.

Definizione 3.6. Indichiamo con $N(t)$ ($t \geq 0$) il numero di *salti* che si verificano nell'intervallo di tempo $[0, t]$, supponendo che $N(0) = 0$. Il processo $N(t) : t \geq 0$ è chiamato *processo di conteggio*. Un processo di conteggio è detto a incrementi indipendenti se il numero dei salti che si verificano in intervalli di tempo disgiunti sono indipendenti.

Si noti come $N(t)$ sia un processo a tempo continuo, nel senso che non esistono istanti privilegiati; le variazioni di $N(t)$ possono avvenire in qualunque $t \in \mathbb{R}$.

Definizione 3.7. Un processo di conteggio $N(t) : t \geq 0$ è chiamato *processo di Poisson*, di intensità $\lambda > 0$, se si verificano le seguenti proprietà:

1. $N(0) = 0$;
2. Il processo è a incrementi indipendenti;
3. Il numero di *salti* che si verificano in un intervallo di tempo di lunghezza $t \geq 0$ segue la distribuzione di Poisson di parametro λt , vale a dire che per ogni $s \geq 0$ ed ogni $t \geq 0$ si ha:

$$P\{N(s+t) - N(s) = n\} = e^{-\lambda t} \frac{(\lambda t)^n}{n!} \quad (n \geq 0)$$

3.2 Descrizione del modello

Procediamo ora a presentare un modello stocastico molto semplice per studiare la diffusione di un'epidemia entro particolari condizioni conservative. In particolare, si richiede che la popolazione sia approssimabile come un gruppo chiuso e omogeneo, di entità costante, e che non siano tenuti in considerazione

periodi di latenza, variazioni nell'incisività della malattia e casi di immunità parziale. In altre parole, assumiamo che gli infetti diventino istantaneamente contagiosi, senza alcun periodo di incubazione della malattia; che il periodo di osservazione considerato sia sufficientemente breve da rendere trascurabili variazioni del numero di individui per merito di nascite o morti; che le persone guarite dalla malattia o geneticamente immuni non possano essere infettate.

Nell'istante zero del fenomeno epidemiologico, chiamiamo m il numero iniziale di individui infetti e n il numero di suscettibili. I periodi in cui perdura la malattia degli infetti sono v.a. indipendenti e identicamente distribuite tra loro secondo la distribuzione di probabilità di una v.a. I . Durante tale periodo, ogni malato può entrare in contatto con dei suscettibili seguendo le leggi probabilistiche di un processo di Poisson di intensità $\frac{\lambda}{n}$; in tal caso, l'individuo suscettibile si ammala istantaneamente e subito acquisisce l'abilità di infettare altre persone. Un individuo viene considerato rimosso quando non è più in grado di trasmettere la malattia e diventa conseguentemente immune a futuri attacchi. Infine, l'epidemia cessa di esistere quando il numero di persone infette diventa pari a zero. Tutti i processi di Poisson sono considerati indipendenti tra loro e dai periodi di tempo.

Il modello appena descritto viene denominato *modello SIR*, dove le lettere S, I, R indicano rispettivamente la classe dei *suscettibili*, *infetti* e dei *rimossi*, le classi già definite nel capitolo 2.

Seguendo la notazione di Ball (1995) chiamiamo tale processo $E_{n,m}(\lambda, I)$ e ne indichiamo con τ la media e con σ^2 la varianza. Assumendo che l'unico modo per contrarre l'infezione sia quello di entrare in contatto con individui malati, definiamo con $\frac{\lambda}{n}$ la frequenza di incontri, dove λ è il tasso costante di contatti tra un particolare infetto con individui suscettibili; si noti che tale valore non dipende dall'entità totale della popolazione.

Definiamo inoltre due variabili molto importanti dal punto di vista epidemiologico. Z è il numero di suscettibili iniziali che entro il termine dell'epidemia sono diventati infetti: si tratta di un numero casuale ma finito che assume valori compresi tra 0 e n . La seconda variabile, M_0 , è definita come il valore atteso dei contagi generati da un singolo individuo infetto all'interno di una popolazione ragionevolmente grande. Per il modello in considerazione $M_0 = \lambda\tau$, dal momento che τ è la durata media della malattia di un individuo ed è proprio durante tale periodo che il malato incontra individui sani (secondo il tasso λ), contagiandoli. Esistono diversi approcci allo studio di tale modello.

3.3 La costruzione di Sellke

Il metodo di Sellke (1983) consiste nel tenere traccia della pressione infettiva totale degli individui infetti, mentre a ogni suscettibile viene associato un valore critico di esposizione all'infezione; appena tale valore viene superato il suscettibile viene considerato infetto. Chiamiamo gli infetti iniziali $-(m-1), -(m-2), \dots, 0$ e i suscettibili $1, 2, \dots, n$. Siano $I_{-(m-1)}, I_{-(m-2)}, \dots, I_n$ variabili casuali identicamente distribuite e indipendenti tra loro, in accordo con la distribuzione di probabilità della variabile casuale I . Sia inoltre $Q_1, Q_2, \dots, Q_n$ una sequenza indipendente di variabili esponenziali casuali indipendenti e identicamente distribuite, con media 1: queste rappresentano i valori di soglia individuali. Per $i = -(m-1), -(m-2), \dots, 0$, il malato contrassegnato da i rimane contagioso per un tempo I_i e viene poi rimosso. Denotiamo con $Y(t)$ il numero di infetti al tempo t , e sia

$$A(t) = \frac{\lambda}{n} \int_0^t Y(u) du$$

definita come la pressione infettiva totale esercitata su un preciso suscettibile fino al tempo t . Per $i = 1, 2, \dots, n$, il suscettibile i diventa infetto quando $A(t)$ raggiunge Q_i . Il j -esimo suscettibile che diventa infetto rimane malato per un tempo I_j e poi viene rimosso. L'epidemia termina quando non sono più presenti individui infetti.

3.4 Il caso Markoviano

Consideriamo il modello standard SIR $E_{n,m}(\lambda, I)$ e denotiamo con $X(t)$ e $Y(t)$ rispettivamente il numero di suscettibili e di infetti al tempo t . Il processo $(X, Y) = \{(X(t), Y(t)); t \geq 0\}$ è un processo di Markov se e solo se il periodo di infezione ha la proprietà di assenza di memoria. Si assuma che I sia distribuita esponenzialmente con intensità γ . In tal caso possiamo descrivere il processo (X, Y) tramite la seguente tabella di transizione:

$$\begin{aligned} (i, j) &\rightarrow (i-1, j+1) \text{ con probabilità } \frac{\lambda i j}{n} \\ (i, j) &\rightarrow (i, j-1) \text{ con probabilità } \gamma j \end{aligned}$$

Infatti, noto che al tempo t sono presenti i suscettibili e j infetti, all'istante successivo uno dei suscettibili potrebbe essersi ammalato in seguito a un eventuale incontro con degli infetti; in alternativa, uno degli infetti potrebbe guarire ed entrare a far parte dei rimossi. Nonostante non sia motivata da considerazioni realistiche, l'ipotesi di considerare il periodo di infezione come

distribuito esponenzialmente semplifica notevolmente lo studio del modello e può essere considerata più che ragionevole.

Capitolo 4

Simulazioni del Modello SIR

Nel capitolo 2 abbiamo studiato il modello deterministico fornito da Kermack e McKendrick, dandone diverse interpretazioni di tipo qualitativo. Riprendiamo ora tale modello per darne alcune considerazioni di carattere più pratico; in seguito, basandoci sui concetti espressi nel capitolo 3, vedremo una tecnica di simulazione stocastica in grado di migliorare la precisione del modello SIR.

4.1 Parametri deterministici per il controllo di un'epidemia

Una popolazione si dice affetta da epidemia se in un dato istante il numero di infetti è più grande del numero di malati iniziali, I_0 . In uno sviluppo infettivo, ciò che principalmente desta preoccupazione è il caso $I(t) \gg I_0$ (per un qualche valore di t), con particolare attenzione al valore assunto dal picco massimo di $I(t)$, che chiameremo I_* , e al suo valore in proporzione alla popolazione totale, di entità costante N .

Abbiamo già studiato come il manifestarsi dell'epidemia sia strettamente legato alle costanti positive α e β , che determinano rispettivamente il tasso di infezione e di guarigione (o morte). In particolare, l'infezione si propaga quando il numero di suscettibili disponibili all'istante iniziale, S_0 , è maggiore del rapporto $\frac{\beta}{\alpha} = \gamma$, che rappresenta la misura di quanto la rimozione dalla malattia sia più veloce dell'infezione stessa; per tale motivo, γ è detto anche *tasso relativo di infezione*, mentre il suo inverso, $\gamma^{-1} = \frac{\alpha}{\beta}$, è chiamato *tasso di contatto*.

La relazione tra S_0 e γ può essere verificata visivamente nei grafici 4.1, 4.2 e 4.3. I tre grafici riportano le soluzioni di $S(t)$ (curva gialla), $I(t)$ (curva rosa) e $R(t)$ (curva blu) nelle equazioni del modello di Kermack e McKendrick per

valori di $\alpha = 0.0005$, $\beta = 0.02$ (perciò $\gamma = 40$), $I_0 = 5$, $R_0 = 0$. Ciò che varia è il valore dei suscettibili iniziali S_0 e, di conseguenza, la taglia della popolazione totale. In figura 4.1, S_0 assume valore 100, che è ben al di sopra del valore di γ , e si assiste ad un inasprirsi dell'infezione fino a raggiungere il picco massimo di valore I_* , oltre il quale il numero di malati comincia a scendere fino ad annullarsi del tutto. In figura 4.2, S_0 assume il valore limite 41, leggermente superiore a γ ; si noti come l'infezione stenti a propagarsi, dando origine ad un picco I_* negli istanti immediatamente successivi a $t = 0$. Infine, la figura 4.3 mostra l'annullarsi quasi immediato dell'infezione in presenza di pochi suscettibili (nell'esempio, $S_0 = 20$): la malattia non si propaga.

Introduciamo ora il numero $r = \gamma^{-1}S_0$, detto *tasso riproduttivo* dell'infezione. Esso misura il numero di infezioni secondarie prodotte da ognuno degli infetti iniziali (quelli considerati in I_0) quando la malattia è introdotta nella popolazione. L'epidemia si diffonde se $r > 1$, mentre declina subito se $r < 1$. In riferimento ai tre casi appena analizzati, si nota che vale $r > 1$ quando S_0 è pari a 100 o 41, mentre vale $r < 1$ quando $S_0 = 20$. Si evidenzia quindi come, a parità di valori per gli altri parametri dell'epidemia, la situazione cambi radicalmente a seconda del valore iniziale dei suscettibili disponibili per essere infettati: minore è tale valore, meno incisiva sarà l'epidemia. Questa considerazione è perfettamente in linea con l'utilità delle campagne di vaccinazione: vaccinare la popolazione, infatti, non modifica la contagiosità dell'infezione, ma va a diminuire il valore di S_0 , cioè delle persone a rischio di ammalarsi, facendo sensibilmente calare l'efficacia dell'epidemia.

Come accennato in precedenza, un parametro rilevante nella valutazione dell'infezione è il picco di massima intensità di $I(t)$, che abbiamo chiamato I_* . L'equazione (2.2) e l'osservazione che il massimo di $I(t)$ si raggiunge per il valore $S(t) = \gamma = \frac{\beta}{\alpha}$ consentono di determinare la massima intensità di diffusione della malattia. Infatti, con la condizione che $S_0 > \gamma$ (in caso contrario, l'epidemia non si verifica), si ha

$$I_* = I(\gamma) = c_0 - \gamma + \gamma \log(\gamma) \quad (4.1)$$

Ricordando inoltre che $c_0 = I_0 + S_0 - \gamma \log(S_0)$ e supponendo che al tempo $t = 0$ si abbia $I_0 \ll S_0$, è lecito fare la seguente approssimazione:

$$c_0 \simeq S_0 - \gamma \log(\gamma)$$

L'equazione (4.1) diventa quindi

$$I_* \simeq (S_0 - \gamma) + \gamma [\log(\gamma) - \log(S_0)] = (S_0 - \gamma) - \gamma \log\left(\frac{S_0}{\gamma}\right).$$

4.1. PARAMETRI DETERMINISTICI PER IL CONTROLLO DI UN'EPIDEMIA 25

Sfruttando le equazioni appena scritte, siamo ora in grado di calcolare il valore del picco degli esempi riportati in precedenza. Nel caso $S_0 = 100$ di figura 4.1, la formula restituisce il valore $I_* = 28.348$, mentre nel caso $S_0 = 41$ di figura 4.2 si ottiene $I_* = 5.012$ (quest'ultimo valore di pochissimo superiore al corrispondente I_0). Nel caso della figura 4.3 sappiamo già che $S_0 = 20$ rappresenta un valore minore del tasso di rimozione ($\gamma = 40$), perciò non è presente alcun picco.

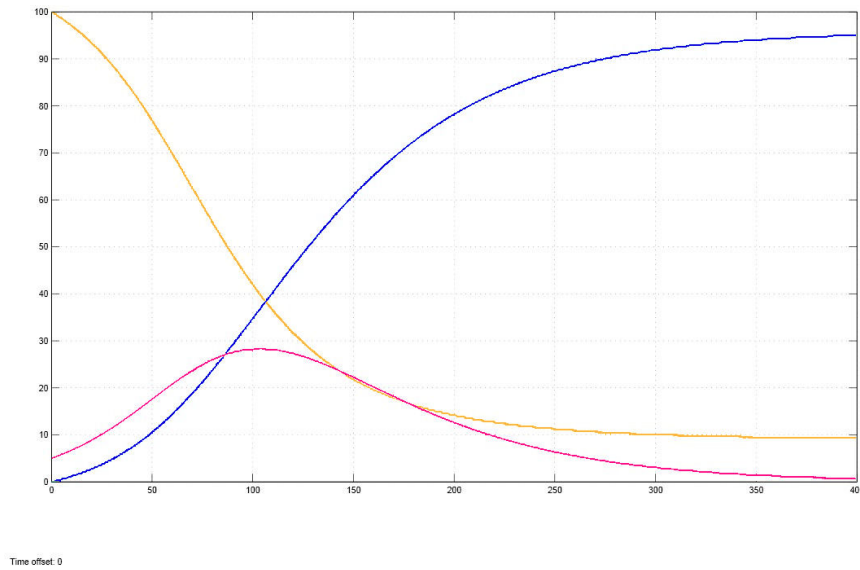


Figura 4.1: Evoluzione dei valori di $S(t)$ (curva gialla), $I(t)$ (curva rosa), $R(t)$ (curva blu) per i seguenti valori: $\alpha = 0.0005$, $\beta = 0.02$ (perciò $\gamma = 40$), $I_0 = 5$, $R_0 = 0$, $S_0 = 100$.

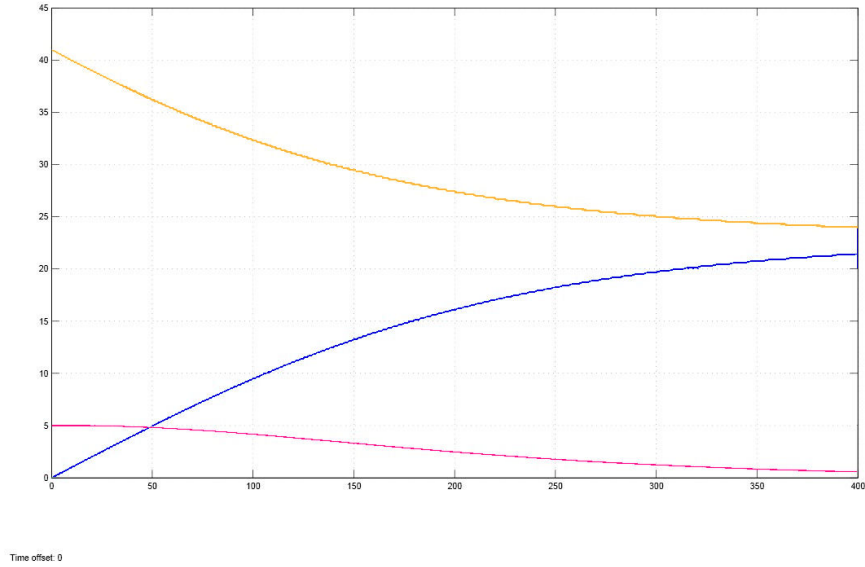


Figura 4.2: Evoluzione dei valori di $S(t)$ (curva gialla), $I(t)$ (curva rosa), $R(t)$ (curva blu) per i seguenti valori: $\alpha = 0.0005$, $\beta = 0.02$ (perciò $\gamma = 40$), $I_0 = 5$, $R_0 = 0$, $S_0 = 41$.

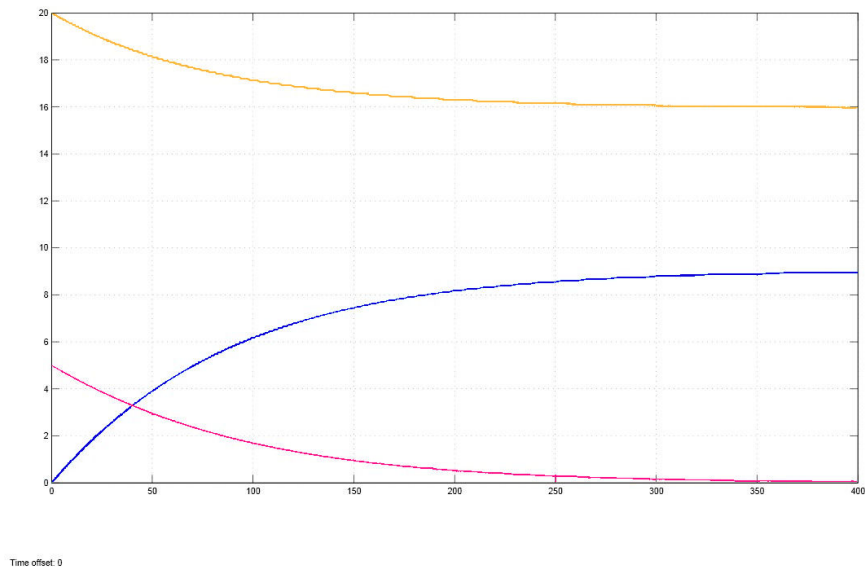


Figura 4.3: Evoluzione dei valori di $S(t)$ (curva gialla), $I(t)$ (curva rosa), $R(t)$ (curva blu) per i seguenti valori: $\alpha = 0.0005$, $\beta = 0.02$ (perciò $\gamma = 40$), $I_0 = 5$, $R_0 = 0$, $S_0 = 20$.

4.2 Simulazioni stocastiche del modello SIR

Il modello di Kermack e McKendrick, studiato nel capitolo 2, prevede un approccio puramente deterministico allo studio dell'epidemia. Mantenendo costanti i parametri e le condizioni iniziali per tutta l'analisi, riesce a produrre una descrizione dell'epidemia tanto più fedele quanto più grande è il numero di soggetti che risultano infetti. Esso resta valido nelle ipotesi che tutti gli individui infettino alla stessa velocità e che sia parimenti fisso il tempo in cui i soggetti malati vengono rimossi dalla popolazione. In altri termini, il modello di Kermack e McKendrick mostra l'evoluzione di *quantità medie*. Nella realtà, tuttavia, le epidemie presentano una notevole variabilità intrinseca, dovuta a numerosi aspetti non considerati all'interno del modello, che ne rendono difficile una stima esatta soprattutto quando il numero di malati in circolazione è scarso, come nelle fasi iniziali e finali della diffusione. In questi casi, un approccio stocastico allo studio dell'epidemia può darne una comprensione maggiore, in quanto tiene in considerazione anche fluttuazioni della dinamica epidemica attorno alle quantità medie.

Per dare un'idea degli effetti delle fluttuazioni casuali sulla forza dell'epidemia, ne consideriamo non uno studio analitico (piuttosto complesso) bensì una simulazione stocastica. Introduciamo allora una componente aleatoria nella forma delle *probabilità di contagio e di rimozione*. Già nel capitolo 2 ci siamo riferiti in questi termini alle costanti α e β (con le opportune considerazioni). Fissiamo ora una unità di tempo dt , sufficientemente piccola, come passo di rilevazione del numero di infetti e definiamo $a := \alpha dt$ e $b := \beta dt$ come costanti fondamentali del modello. Lo stato del sistema sarà determinato da due valori interi che chiamiamo s e k , che indicano rispettivamente il numero di suscettibili e di infetti. Ricordando che la taglia della popolazione è costante e vale N , il numero di rimossi sarà semplicemente $N - k - s$. È necessario assicurarsi che la quantità $[sk\alpha + k\beta]dt$ sia non superiore ad uno per tutti i valori di s e k in $[0, N]$, cosa sempre vera per dt abbastanza piccolo.

La probabilità che nessuno dei suscettibili si ammali è:

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

La generica $P_a(m; s)$, che esprime la probabilità che tra s sani, m si infettino, vale invece:

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

La probabilità di rimozione di un individuo infetto è:

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

Facciamo notare che quella appena calcolata è pari alla probabilità che un individuo diventi permanentemente immune all'infezione, poichè abbiamo assunto come ipotesi che gli individui guariti non siano più infettabili. Consideriamo al tempo t uno stato, che indicheremo con $|s, k\rangle$, in cui vi sono s suscettibili e k infettivi, e analizziamo il sistema all'istante $t + dt$. Nell'intervallo di tempo tra t e $t + dt$ potrebbe verificarsi (con probabilità $\beta k dt$) una guarigione o una morte, e in tal caso ci ritroveremmo nello stato $|s, k - 1\rangle$; oppure, uno dei suscettibili potrebbe ammalarsi (con probabilità $\alpha s k dt$), e quindi il nuovo stato sarebbe $|s - 1, k + 1\rangle$; infine, la situazione potrebbe non cambiare (con probabilità $1 - \alpha s k dt - \beta k dt$) e dunque resteremmo nello stato $|s, k\rangle$. Formalizzando quanto detto, abbiamo:

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

In riferimento allo studio appena effettuato, l'intervallo $J = [0, 1]$ viene diviso in tre sottointervalli:

$$J = J_+ \cup J_0 \cup J_- = [0, \alpha s k dt] \cup [\alpha s k dt, 1 - \beta k dt] \cup [1 - \beta k dt, 1]$$

Successivamente si estrae un numero casuale c uniformemente distribuito tra 0 e 1, e in base a tale valore si procede a determinare lo stato successivo a $|s, k\rangle$: lo stato diviene $|s - 1, k + 1\rangle$ se $c \in J_+$, diviene $|s, k - 1\rangle$ se $c \in J_-$, altrimenti, se $c \in J_0$, rimane $|s, k\rangle$.

Sappiamo già che l'infezione tenderà a estinguersi in un tempo finito, sia per effetto delle fluttuazioni ma anche per la dinamica stessa del sistema; per tale ragione, non ci interessa conoscere il comportamento all'equilibrio del sistema. Potrebbe invece essere interessante avere informazioni sul valore del picco massimo di $I(t)$, o sapere per quanto tempo questo si mantiene al di sopra di un valore di riferimento I_r .

Utilizzando Matlab, scriviamo un programma che svolga il procedimento appena illustrato. Per chiarezza, ne riportiamo il codice qui di seguito.

```

alfa= 0.03;           %probabilita' di ammalarsi
beta= 0.2;            %probabilita' di essere rimossi
dt= 0.0001;          %passo di campionamento: 0.1 ms
t= 10;                %monitoro la popolazione per 10s
a= alfa * dt;
b= beta * dt;
N= 150;               %grandezza della popolazione
n= t / dt;            %effettuo 100'000 rilevazioni

```

```

r= [1:n];           %array del numero di rimossi a ogni rilevazione
k= [1:n];           %array del numero di infetti a ogni rilevazione
s= [1:n];           %array del numero di suscettibili a ogni rilevazione
k(1) = 10;           %numero iniziale di infetti
s(1)= N - k(1);      %numero iniziale suscettibili
r(1)= 0;             %numero iniziale rimossi

for i=2:n
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);
hold all
plot(s);
hold all
plot(r);

```

Riportiamo qui di seguito alcune simulazioni del modello SIR ottenute grazie al procedimento descritto. La figura 4.4 mostra cinque possibili andamenti per $S(t)$, mentre la 4.5 ne identifica altri cinque per $I(t)$. In figura 4.6 viene riportato un esempio di andamento per tutte e tre le variabili $S(t)$, $I(t)$ e $R(t)$. Si veda infine la figura 4.7 per un confronto visivo tra approccio stocastico e approccio deterministico allo studio del modello.

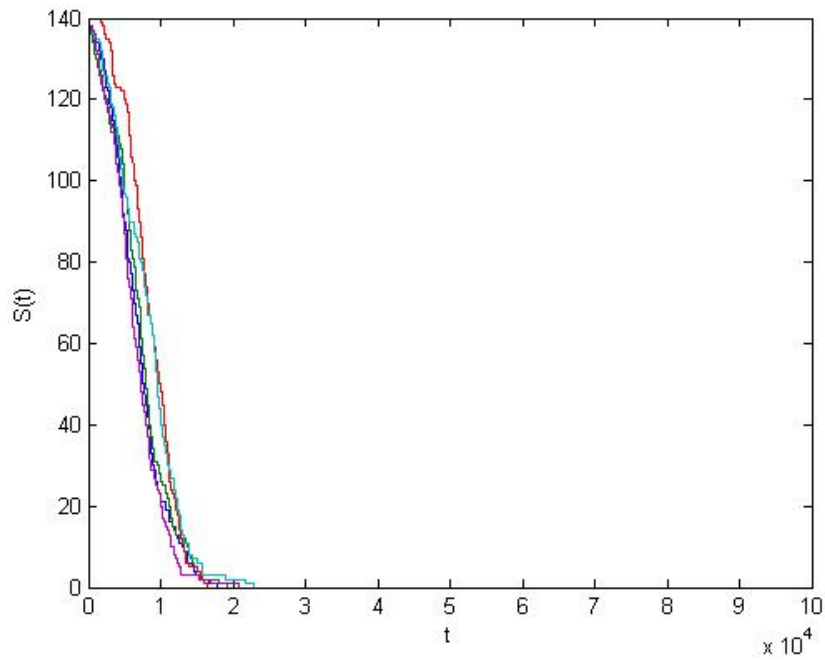


Figura 4.4: Schema di cinque simulazioni stocastiche dell'andamento del numero di suscettibili (ordinate) col passare del tempo (ascisse). Valori utilizzati: $\alpha = 0.03$, $\beta = 0.2$, $N = 150$, $dt = 0.0001$ (passo di campionamento), $t = 10ms$, $I_0 = 10$, $S_0 = 140$, $R_0 = 0$. Si può notare come l'effetto stocastico sia quello di generare curve diverse, ma dall'andamento abbastanza simile, sia nella forma che nei tempi dei fenomeni in gioco.

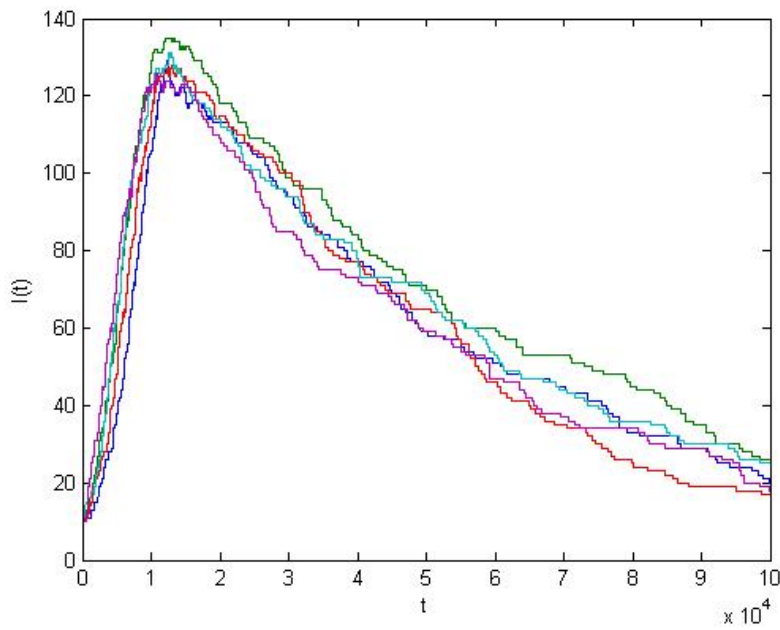


Figura 4.5: Schema di cinque simulazioni stocastiche dell'andamento del numero di infetti (ordinate) col passare del tempo (ascisse). Valori utilizzati: $\alpha = 0.03$, $\beta = 0.2$, $N = 150$, $dt = 0.0001$ (passo di campionamento), $t = 10ms$, $I_0 = 10$, $S_0 = 140$, $R_0 = 0$. Come nella figura precedente, si può notare che le cinque simulazioni stocastiche generano curve piuttosto simili tra loro.

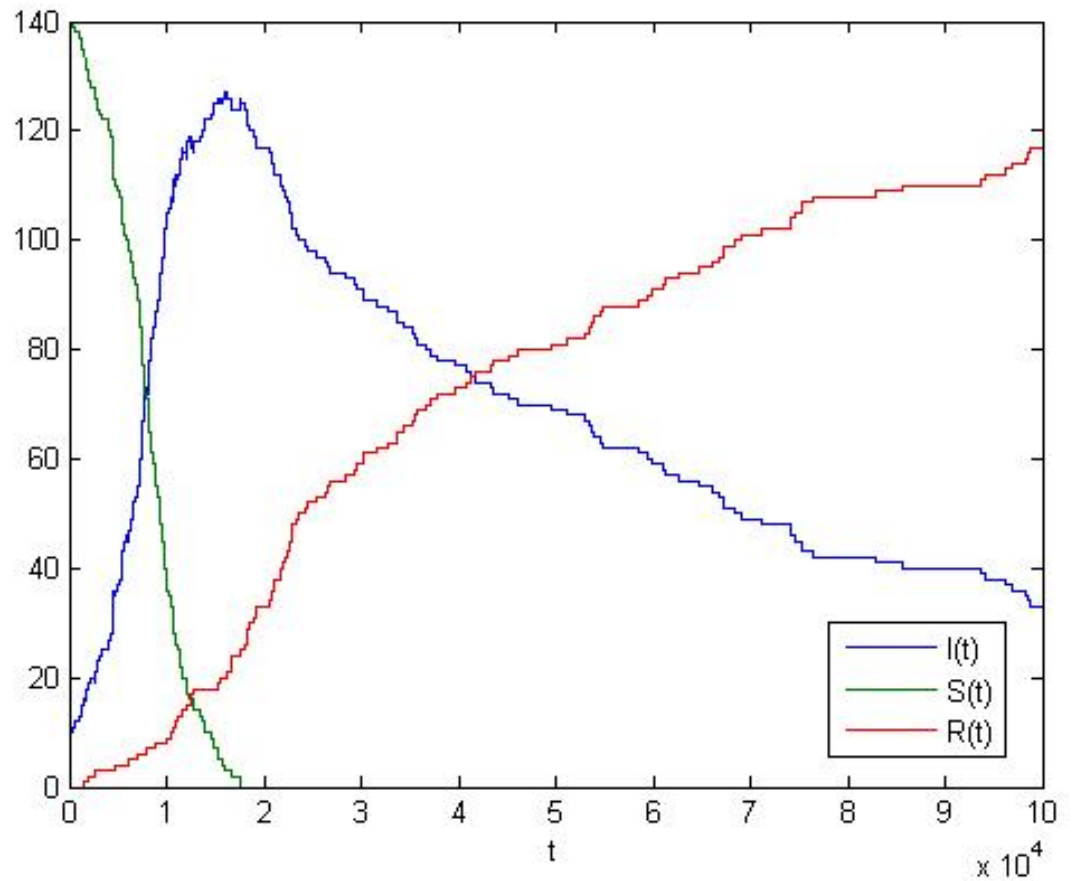


Figura 4.6: Simulazione stocastica dell'andamento del numero di suscettibili (curva verde), infetti (curva blu) e rimossi (curva rossa) al passare del tempo espresso in ascissa. Valori utilizzati: $\alpha = 0.03$, $\beta = 0.2$, $N = 150$, $dt = 0.0001$ (passo di campionamento), $t = 10ms$, $I_0 = 10$, $S_0 = 140$, $R_0 = 0$.

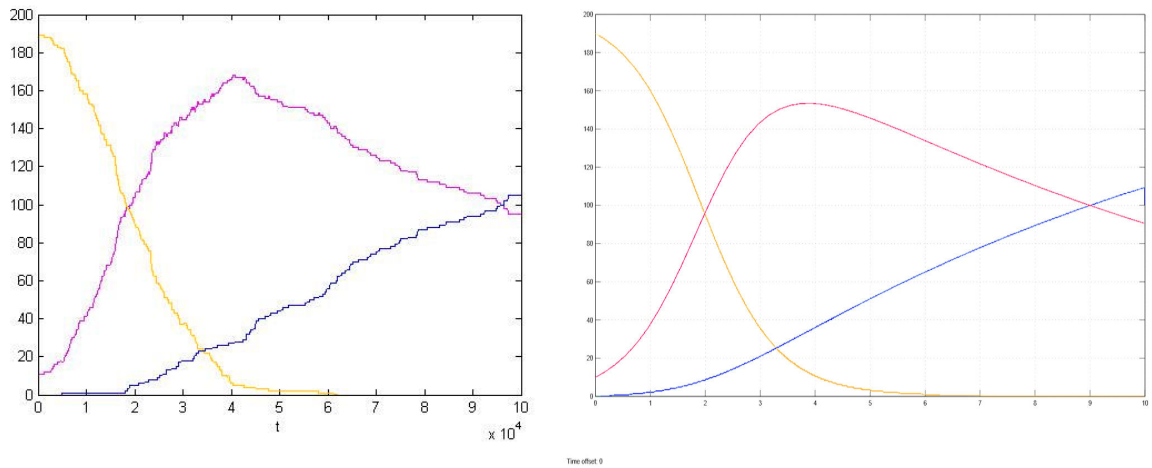


Figura 4.7: Confronto grafico tra i due approcci visti per studiare il modello SIR: a sinistra quello stocastico, a destra quello deterministico. Per entrambe le simulazioni sono stati usati i seguenti parametri: $\alpha = 0.008$, $\beta = 0.1$, $N = 200$, $dt = 0.0001$ (passo di campionamento), $t = 10ms$, $I_0 = 10$, $S_0 = 190$, $R_0 = 0$. In ascissa è riportato il tempo, mentre l'ordinata indica l'evoluzione di $S(t)$ (curva gialla), $I(t)$ (curva rosa) e $R(t)$ (curva blu). In base a considerazioni precedenti, la simulazione stocastica (qui effettuata una sola volta per chiarezza) permette di rappresentare al meglio la variabilità dell'epidemia considerata; si noti tuttavia come l'approccio deterministico risulti abbastanza fedele nel riprodurre l'evoluzione del fenomeno, sia dal punto di vista dei tempi che da quello dell'andamento delle tre variabili.

Bibliografia

- [1] G. Gaeta, *Modelli matematici in Biologia*, Springer, pp. 109-144, 2007.
- [2] H. Andersson e T. Britton, *Stochastic Epidemic Models and their Statistical Analysis*, Springer Verlag, 2000.
- [3] S. M. Ross, *Calcolo delle probabilità*, Apogeo, **II ed.**, 2007.
- [4] D.P. Bertsekas e J.N. Tsitsiklis, *Introduction to Probability*, Athena Scientific **II ed.**, 2008.
- [5] E. Renshaw, *Modelling Biological Populations in Space and Time*, Cambridge University Press, 1993.