

POLITECNICO DI TORINO

Corso di Laurea
in Matematica per l'Ingegneria

Tesi di Laurea

Applicazione delle catene markoviane a modelli epidemiologici: confronto tra varianti COVID-19



Relatori

prof. Paola Siri
prof. Enrico Bibbona
firma dei relatori

.....
.....

Candidato

Alessandro Ciarrocchi
firma del candidato

.....

Anno Accademico 2023-2024

Ai miei genitori

Ai miei nonni

A mia zia

Indice

I	Catene di Markov	9
1	Catene di Markov a tempo discreto	11
1.1	Premesse	11
1.2	Principi fondamentali delle catene di Markov a tempo discreto	11
1.3	Equazioni di Chapman-Kolmogorov	13
1.4	Classificazione degli stati	14
1.5	Random-walk in una dimensione	16
2	Catene di Markov a tempo continuo	17
2.1	Aspetti fondamentali	17
2.2	Corse esponenziali	18
2.3	Tasso di transizione e risultati derivati	20
II	Modello SIR e Applicazioni	23
3	Introduzione al modello SIR	25
3.1	Premesse	25
3.2	Analisi del modello deterministico	26
3.3	Modello SIR stocastico a tempo continuo	27
4	Algoritmi di simulazione deterministici	29
4.1	Metodo di Eulero esplicito	29
4.2	Metodo di Runge-Kutta esplicito a r stadi	30
5	Algoritmi di simulazione stocastici	31
5.1	Algoritmo di Gillespie	32
5.2	Algoritmo del tau-leaping	33
6	Applicazione: Confronto tra varianti di COVID-19	35
6.1	Definizione dei dati	35
6.2	Simulazioni con il Metodo di Eulero	36
6.3	Simulazioni con il metodo di Runge-Kutta	38
6.4	Simulazioni con il metodo di Gillespie	39
6.5	Conclusioni	43

A	Codici	45
A.1	Codice con metodo di Eulero	45
A.2	Codice con metodo di Runge-Kutta a 4 stadi	46
A.3	Codice con metodo di Gillespie	48

Introduzione

Le catene di Markov, introdotte per la prima volta dal matematico russo Andrei Markov (1856-1922) alla fine del XIX secolo, costituiscono un potente strumento matematico in grado di rappresentare le diverse dinamiche di un sistema attraverso transizioni di stato probabilistiche, basate esclusivamente sullo stato corrente. La proprietà di assenza di memoria si manifesta come una caratteristica distintiva delle catene di Markov, conferendo a tali modelli una brillante capacità di adattarsi ad una vasta gamma di fenomeni dinamici le cui evoluzioni sono principalmente legate alle condizioni attuali e non alla storia passata. Le catene di Markov trovano applicazione in numerosi settori differenti. In economia, ad esempio, vengono utilizzati per modellare il comportamento dei mercati finanziari e prevedere l'evoluzione di prezzi e tassi di interesse. In fisica sono alla base della teoria cinetica dei gas e dei modelli di diffusione molecolare. Nell'informatica invece vengono utilizzati in algoritmi di apprendimento automatico e nella compressione dei dati. Un'applicazione particolarmente significativa delle catene di Markov è in epidemiologia, un campo che studia la diffusione delle malattie infettive all'interno delle popolazioni. Grazie alla loro struttura matematica, le catene di Markov permettono di modellare in modo rigoroso e flessibile la probabilità di transizione tra diversi stati di salute in una popolazione. Attraverso questi modelli, è possibile prevedere come un'epidemia si propagerà nel tempo, stimare il numero di individui che saranno infettati in futuro, e valutare l'impatto di diverse strategie di intervento, come la vaccinazione o la quarantena.

La tesi è organizzata in sei capitoli principali, ciascuno dei quali affronta un aspetto specifico delle catene di Markov e delle loro applicazioni nei modelli epidemiologici. Il primo capitolo fornisce una panoramica delle catene di Markov a tempo discreto, introducendo le definizioni fondamentali, come gli stati e le matrici di transizione. Viene esplorata la proprietà di "assenza di memoria" e sono presentate le equazioni di Chapman-Kolmogorov, che costituiscono la base teorica per l'analisi dei processi Markoviani. Il secondo capitolo estende la discussione alle catene di Markov a tempo continuo e illustra la differenza con quelle a tempo discreto. I primi due capitoli utilizzano come riferimento principale il testo di [Ross \[2019\]](#), che offre una trattazione completa delle catene di Markov e dei processi stocastici in generale. Il terzo capitolo è dedicato al modello SIR (Suscettibili-Infetti-Recuperati), uno dei modelli più utilizzati in epidemiologia per descrivere la diffusione delle malattie infettive. Il modello viene presentato sia nella sua forma deterministica che in quella stocastica a tempo continuo. Si esplorano le dinamiche del modello, analizzando come i vari parametri influenzano l'andamento dell'epidemia. Per la trattazione dei

modelli epidemiologici, si fa riferimento in particolare al testo di [Allen \[2008\]](#), che fornisce un'introduzione dettagliata ai modelli epidemiologici stocastici. Il quarto capitolo si concentra sui metodi di simulazione deterministici, utilizzati per risolvere le equazioni differenziali che descrivono i modelli epidemiologici. Vengono presentati i metodi di Eulero e di Runge-Kutta, evidenziando vantaggi e limitazioni di ciascuno. Il capitolo fa ampio riferimento al testo di [Burden and Faires \[2011\]](#), che è una risorsa fondamentale per comprendere e implementare tecniche di simulazione numerica. Il quinto capitolo introduce gli algoritmi di simulazione stocastici: in particolare vengono trattati l'algoritmo di Gillespie ed il metodo del tau-leaping, entrambi utilizzati per simulare processi stocastici che coinvolgono transizioni tra stati in un modello epidemiologico. Questo capitolo si basa, tra gli altri, sul lavoro di [Banks et al. \[2012\]](#), che discute in dettaglio i metodi di simulazione per catene di Markov a tempo continuo. Il sesto e ultimo capitolo è dedicato a un'applicazione pratica dei modelli discussi nei capitoli precedenti: l'analisi e il confronto tra diverse varianti del virus COVID-19. Attraverso i modelli e gli algoritmi sviluppati, sono state effettuate numerose simulazioni per studiare come ciascuna variante influenzi la dinamica dell'epidemia. È particolarmente interessante osservare la notevole capacità del virus di adattarsi rapidamente all'ambiente esterno, mutando con estrema velocità per garantire la propria sopravvivenza e continuare a diffondersi all'interno della popolazione.

Parte I

Catene di Markov

Capitolo 1

Catene di Markov a tempo discreto

I concetti presenti nei primi due capitoli sono descritti in dettaglio in [Ross \[2019\]](#).

1.1 Premesse

Fissiamo per tutto l'elaborato uno spazio di probabilità (Ω, \mathcal{F}, P) dove Ω rappresenta lo spazio campione, \mathcal{F} è una sigma algebra e P è una misura di probabilità definita su \mathcal{F} che assegna a ciascun evento di \mathcal{F} un valore compreso tra 0 e 1, soddisfacendo $P(\Omega) = 1$ e la condizione di sigma additività per eventi mutuamente esclusivi.

Definizione 1.1.1 (Processo stocastico). *Un processo stocastico $\{X(t), t \in T\}$ è una famiglia di variabili casuali definite su uno spazio di probabilità (Ω, \mathcal{F}, P) , dove $X(t)$ indica lo stato del processo al tempo t . Se l'insieme degli indici T è discreto, il processo è a tempo discreto; se è un intervallo reale, è a tempo continuo. Lo spazio degli stati E rappresenta i valori che $X(t)$ può assumere, descrivendo l'evoluzione temporale di un fenomeno casuale nel tempo.*

Nel caso in cui T sia discreto, preferiremo la notazione X_n a $X(n)$ e ipotizzeremo per semplicità $T = \mathbb{N}$.

1.2 Principi fondamentali delle catene di Markov a tempo discreto

Consideriamo un processo che assume al tempo n -esimo ($n = 0, 1, 2, \dots$) un determinato valore X_n e supponiamo di voler analizzare un modello probabilistico per una sequenza di X_0, X_1, X_2, \dots . Un primo approccio potrebbe essere quello di considerare i diversi X_n come variabili indipendenti, ma si otterrebbe con molta probabilità un modello poco realistico e su cui poter fare relativamente poco affidamento per via del fatto che si andrebbero completamente a trascurare le relazioni tra le variabili.

Al contrario, considerare un modello che tenga conto di tutta la storia passata potrebbe condurre ad una complessità eccessiva o ad una dipendenza marcata dalle informazioni pregresse. Questo approccio potrebbe rendere il modello difficile da gestire o interpretare, compromettendo la sua capacità di offrire una visione generale del fenomeno in analisi. Resta dunque semplice notare come le Catene di Markov si rivelino una buona via di mezzo tra le due proposte, basandoci solo sulle informazioni note al tempo presente (X_n) e non su quelle del passato.

Sulla base dei concetti appena analizzati, possiamo dare una definizione formale di catena di Markov a tempo discreto.

Definizione 1.2.1 (Catena di Markov a tempo discreto). *Una catena di Markov a tempo discreto è un processo stocastico $\{X_n, n \in \mathbb{N}\}$ definito su uno spazio di probabilità (Ω, \mathcal{F}, P) e a valori in uno spazio E discreto che soddisfa la proprietà di Markov:*

$$\begin{aligned} P\{X_{n+1} = j | X_n = i, X_{n-1} = i_{n-1}, \dots, X_1 = i_1, X_0 = i_0\} = \\ = P\{X_{n+1} = i_{n+1} | X_n = i\} = \\ = P_{ij}(n), \quad \forall n \in \mathbb{N}, i, j, i_0, \dots, i_{n-1} \in E. \end{aligned}$$

X_n rappresenta lo stato del processo al tempo n , i_n indica uno stato specifico al tempo n e $P_{ij}(n)$ è la probabilità di transizione ad un passo all'istante di tempo discreto n .

Nelle catene di Markov, la distribuzione condizionata di ogni futuro stato X_{n+1} , dati gli stati passati $X_0, X_1, X_2, \dots, X_{n-1}$ e lo stato presente X_n , è indipendente dagli stati passati e dipende solamente dallo stato presente.

Definizione 1.2.2 (Catena omogenea). *Una catena di Markov la cui probabilità di transizione ad un passo $P_{ij}(n)$ non dipende da n si dice omogenea.*

Questa caratteristica implica che la dinamica futura del processo dipende solamente dallo stato attuale, senza essere influenzata dal momento temporale specifico in cui si trova. D'ora in avanti, prenderemo in considerazione solamente catene di Markov omogenee e adotteremo la notazione semplificata $P_{ij}(n) = P_{ij}$ per la probabilità di transizione ad un passo di cui ora daremo una definizione rigorosa.

Definizione 1.2.3 (Probabilità di transizione ad un passo). *Data $\{X_n, n \in \mathbb{N}\}$ una catena di Markov omogenea, ogniqualevolta il processo si trova nello stato i al tempo n , esiste una certa probabilità P_{ij} , nota come probabilità di transizione ad un passo, che esso si trovi nello stato j al tempo $n + 1$:*

$$P\{X_{n+1} = j | X_n = i, X_{n-1} = i_{n-1}, \dots, X_1 = i_1, X_0 = i_0\} = P_{ij}$$

per ogni $i_0, i_1, i_2, \dots, i_{n-1}, i, j, n \geq 0$. Denotiamo con \mathbf{P} la **matrice di transizione ad un passo**:

$$\mathbf{P} = \begin{bmatrix} P_{00} & P_{01} & P_{02} & \dots \\ P_{10} & P_{11} & P_{12} & \dots \\ \vdots & \vdots & \vdots & \\ P_{i0} & P_{i1} & P_{i2} & \dots \\ \vdots & \vdots & \vdots & \end{bmatrix}$$

Definizione 1.2.4 (Probabilità di transizione ad n passi). *La probabilità che una catena di Markov omogenea $\{X_n, n \in \mathbb{N}\}$, partendo dallo stato i , giunga nello stato j dopo esattamente n transizioni viene espressa come:*

$$P_{ij}^n = P\{X_{n+k} = j | X_k = i\}, \quad \forall n \geq 0 \quad \forall i, j \geq 0$$

con banalmente $P_{ij}^1 = P_{ij}$.

Osservazione 1.2.5. *Si noti che, poiché le probabilità P_{ij} sono non negative e poiché la catena dovrà spostarsi in un qualsiasi altro stato, abbiamo che:*

$$P_{ij} \geq 0 \quad i, j \geq 0 \quad \sum_{j=0}^{\infty} P_{ij} = 1 \quad i = 0, 1, \dots \quad (1.2.1)$$

Si noti anche che lo spazio degli stati E è preso per semplicità uguale a \mathbb{N} , cosa che faremo anche nel seguito.

1.3 Equazioni di Chapman-Kolmogorov

Per generalizzare quanto appena visto nella sottosezione precedente, introduciamo i seguenti concetti.

Lemma 1.3.1 (Equazioni di Chapman-Kolmogorov). *La probabilità che, partendo da uno stato i , il processo vada in uno stato j dopo $n+m$ transizioni attraverso un percorso che contiene lo stato k all' n -esima transizione è data da:*

$$P_{ij}^{n+m} = \sum_{k=0}^{\infty} P_{ik}^n P_{kj}^m, \quad \forall n, m \geq 0, \forall i, j \in \mathbb{N}.$$

Dimostrazione. Vediamo che

$$\begin{aligned} P_{ij}^{n+m} &= P\{X_{n+m} = j | X_0 = i\} \\ &\stackrel{(1)}{=} \sum_{k=0}^{\infty} P\{X_{n+m} = j, X_n = k | X_0 = i\} \\ &\stackrel{(2)}{=} \sum_{k=0}^{\infty} P\{X_{n+m} = j | X_n = k, X_0 = i\} P\{X_n = k | X_0 = i\} \\ &\stackrel{(3)}{=} \sum_{k=0}^{\infty} P_{kj}^m P_{ik}^n, \end{aligned}$$

dove abbiamo sfruttato la legge delle probabilità totali (1), la definizione di probabilità condizionata (2) e la Definizione 1.2.4 nell'uguaglianza (3). \square

È possibile inoltre scrivere il tutto sotto forma di matrici: denotando $\mathbf{P}^{(n)}$ la **matrice di transizione ad n passi**, vediamo che

$$\mathbf{P}^{(n+m)} = \mathbf{P}^{(n)} \cdot \mathbf{P}^{(m)}$$

intendendo l'operazione di prodotto matriciale. Notiamo in particolare che

$$\mathbf{P}^{(2)} = \mathbf{P}^{(1+1)} = \mathbf{P} \cdot \mathbf{P} = \mathbf{P}^2$$

e per induzione si ottiene

$$\mathbf{P}^{(n)} = \mathbf{P}^{(n-1+1)} = \mathbf{P}^{(n-1)} \cdot \mathbf{P} = \mathbf{P}^n. \quad (1.3.1)$$

La matrice di transizione ad n passi può quindi essere ottenuta moltiplicando la matrice \mathbf{P} per se stessa n volte.

1.4 Classificazione degli stati

Definizione 1.4.1 (Accessibilità). *Uno stato j si dice accessibile da uno stato i ed è indicato come $i \rightarrow j$, se $P_{ij}^n > 0$ per qualche $n \geq 0$. Formalmente scriviamo che:*

$$i \rightarrow j \quad \Longleftrightarrow \quad \exists n \geq 0 \text{ t.c. } P_{ij}^n > 0.$$

Definizione 1.4.2 (Stati comunicanti). *Due stati i, j tra loro accessibili si dicono comunicanti e si denota con $i \leftrightarrow j$.*

Giustificiamo quanto appena visto notando che, se j non è accessibile da i , allora la probabilità che il processo entri almeno una volta in j partendo da i è nulla:

$$\begin{aligned} & P \left\{ \bigcup_{n=0}^{\infty} \{X_n = j\} \mid X_0 = i \right\} \\ & \leq \sum_{n=0}^{\infty} P\{X_n = j \mid X_0 = i\} \\ & = \sum_{n=0}^{\infty} P_{ij}^n = 0. \end{aligned}$$

Proposizione 1.4.3. *La relazione di comunicazione soddisfa:*

1. (riflessività) *Lo stato i comunica con se stesso, $\forall i \geq 0$.*
2. (simmetria) *Se lo stato i comunica con lo stato j , allora lo stato j comunica con lo stato i*

$$i \leftrightarrow j \Longleftrightarrow j \leftrightarrow i.$$

3. (transitività) Se lo stato i comunica con lo stato j e lo stato j comunica con lo stato k , allora anche lo stato i comunica con lo stato k

$$i \leftrightarrow j, j \leftrightarrow k \iff i \leftrightarrow k.$$

Si tratta quindi di una **relazione di equivalenza**.

Possiamo affermare che due stati comunicanti appartengono alla stessa *classe*; il concetto di accessibilità quindi partiziona lo spazio degli stati in classi.

Definizione 1.4.4 (Irriducibilità). Una catena di Markov si dice **irriducibile** se lo spazio degli stati è costituito da una sola classe (ovvero tutti gli stati comunicano tra loro).

Per ogni stato i denotiamo ora con f_i la probabilità che, partendo da i , il processo faccia ritorno in i . Se $f_i = 1$ allora lo stato i si dice **ricorrente**, mentre se $f_i < 1$ allora si dice **transiente**. Per quanto detto finora, è possibile affermare che se lo stato i è ricorrente allora, partendo dallo stato i , il processo tornerà nello stato i infinite volte. Al contrario, se i è transiente, ogni volta che il processo si troverà nello stato i si presenterà una probabilità $1 - f_i > 0$ che il processo non rientri più in i . La probabilità che il processo torni nello stato i esattamente n volte è pari a $f_i^{n-1}(1 - f_i)$. In altre parole, se lo stato i è uno stato transiente allora, partendo da i , il numero di volte che il processo tornerà nello stato i può essere rappresentato da una distribuzione *geometrica* con media $1/(1 - f_i)$ (e quindi è quasi certamente un numero finito).

Proposizione 1.4.5. Un qualsiasi stato i può essere

- ricorrente se $\sum_{n=1}^{\infty} P_{ii}^n = \infty$,
- transiente se $\sum_{n=1}^{\infty} P_{ii}^n < \infty$.

Corollario 1.4.6. Se i è uno stato ricorrente e comunicante con j , allora anche j è ricorrente.

Teorema 1.4.7. In una catena di Markov con un numero finito di stati, essi non possono essere tutti transienti.

Dimostrazione. Supponiamo per assurdo che una catena sia costituita da M stati e che essi siano tutti transienti. Dunque, con probabilità 1, dopo un intervallo di tempo T_i lo stato i -esimo non verrebbe più visitato. Calcolando tale valore per tutti gli M stati del processo notiamo che dopo un intervallo di tempo $T = \max\{T_0, T_1, \dots, T_M\}$ non verrebbe più visitato nessuno stato. Siamo dunque giunti ad un assurdo in quanto il processo deve necessariamente trovarsi in un qualche stato dopo il tempo T . Pertanto almeno uno stato deve essere il ricorrente \square

Osservazione 1.4.8. Dal Corollario 1.4.6 e dal Teorema 1.4.7 possiamo dedurre che tutti gli stati di una catena di Markov finita e irriducibile sono ricorrenti.

1.5 Random-walk in una dimensione

Consideriamo una catena di Markov i cui stati corrispondono agli interi $i = 0, \pm 1, \pm 2, \dots$ e le probabilità di transizione da uno stato all'altro sono definite come

$$P_{i,i+1} = p = 1 - P_{i,i-1}, \quad i \in \mathbb{Z}. \quad (1.5.1)$$

In ogni transizione il processo può muoversi verso destra con probabilità p o verso sinistra con probabilità $1 - p$. Non è difficile notare che tutti gli stati sono comunicanti, dunque per il Corollario 1.4.6 essi sono o tutti transienti o tutti ricorrenti. Per verificare ciò consideriamo lo stato 0 e valutiamo se $\sum_{n=1}^{\infty} P_{00}^n$ è una quantità finita o infinita.

Notiamo che, partendo da 0, è impossibile ritornarci dopo un numero dispari di transizioni, le quali ricordiamo consistono in un incremento/decremento di 1; di conseguenza:

$$P_{00}^{2n-1} = 0, \quad n \geq 1, \quad (1.5.2)$$

mentre, con un numero di $2n$ transizioni, il processo tornerà in 0 solo con n incrementi e n decrementi; poiché si verificherà un incremento con probabilità p e un decremento con probabilità $1 - p$, vediamo che

$$P_{00}^{2n} = \binom{2n}{n} p^n (1-p)^n = \frac{(2n)!}{n!n!} (p(1-p))^n, \quad n \geq 1, \quad (1.5.3)$$

da cui, mediante l'approssimazione di Stirling ed altri passaggi algebrici, otterremo

$$P_{00}^{2n} \sim \frac{(4p(1-p))^n}{\sqrt{n\pi}}. \quad (1.5.4)$$

Quindi

$$\sum_{n=1}^{\infty} P_{00}^n < \infty \quad \Longleftrightarrow \quad \sum_{n=1}^{\infty} \frac{(4p(1-p))^n}{\sqrt{n\pi}} < \infty. \quad (1.5.5)$$

Tuttavia, $4p(1-p) \leq 1$ con uguaglianza verificata per $p = 1/2$ e, di conseguenza, $\sum_{n=1}^{\infty} P_{00}^n = \infty$ se e solo se $p = 1/2$. In conclusione, la catena è ricorrente se $p = 1/2$ (e il processo prende il nome di *random walk simmetrico*), altrimenti è transiente.

Capitolo 2

Catene di Markov a tempo continuo

Per il primo e il terzo paragrafo del corrente capitolo, si farà riferimento al capitolo 6 del libro di [Ross \[2019\]](#), mentre i concetti presenti nel secondo paragrafo sono espressi nel dettaglio in [Durrett \[2016\]](#).

2.1 Aspetti fondamentali

In questo capitolo, tratteremo le catene di Markov a tempo continuo. Questo tipo di processo si distingue da quello precedente per la sua capacità di consentire transizioni non solo in punti discreti del tempo, ma in ogni istante possibile.

Definizione 2.1.1 (Catena di Markov a tempo continuo). *Un processo $\{X(t), t \geq 0\}$ a valori in uno spazio discreto E (che supponiamo pari ad \mathbb{N}) prende il nome di catena di Markov a tempo continuo se, per ogni $s, t \geq 0$ e per ogni $i, j, x(u)$, con $0 \leq u < s$, interi non negativi, vale la proprietà di Markov:*

$$\begin{aligned} P\{X(t+s) = j | X(s) = i, X(u) = x(u), 0 \leq u < s\} \\ = P\{X(t+s) = j | X(s) = i\}. \end{aligned}$$

In altre parole, le catene di Markov a tempo continuo non sono altro che processi stocastici in cui la distribuzione condizionata del futuro $X(t+s)$, noto il presente $X(s)$ e il passato $X(u)$ (con $0 \leq u < s$) dipende esclusivamente dal presente.

Di seguito, mettiamo in evidenza due proprietà distintive delle catene di Markov a tempo continuo.

Proprietà 2.1.2. *In una catena di Markov a tempo continuo:*

1. *Ogniqualevolta il processo entra in uno stato i , l'intervallo di tempo che trascorre nello stato corrente prima di compiere una transizione in un qualsiasi altro stato segue una distribuzione esponenziale di parametro v_i .*

2. (mancanza di memoria) Se il processo rimane in uno stato per un intervallo di tempo $T_i > s$, la probabilità che non effettui alcuna transizione per un intervallo $T_i > t + s$ è indipendente da s . Formalmente:

$$P\{T_i > s + t | T_i > s\} = P\{T_i > t\}. \quad (2.1.1)$$

Analogamente a quanto visto nelle catene di Markov a tempo discreto, chiamiamo **probabilità di transizione** e la denotiamo con

$$P_{ij}(s, t) = P\{X(t + s) = j | X(s) = i\} \quad (2.1.2)$$

la probabilità che un processo attualmente nello stato i passi in uno stato j in un intervallo di tempo t . Come già visto nelle catene di Markov a tempo discreto, se $P\{X(t + s) = j | X(s) = i\}$ è indipendente da s e quindi $P_{ij}(s, t) = P_{ij}(t)$, si dice che la catena è **omogenea**.

2.2 Corse esponenziali

L'argomento di questo paragrafo sarà utile per giustificare l'introduzione dei tassi di transizione nel paragrafo 2.3 ed il loro utilizzo negli algoritmi di simulazione successivi.

Siano $T_i \sim \exp(\lambda_i)$ con $i = 1, \dots, n$ variabili esponenziali indipendenti. Affinchè il minimo tra i T_i superi un determinato valore t , è necessario che $T_i > t$ per ogni $i = 1, \dots, n$. Quindi vediamo che:

$$\begin{aligned} P(\min(T_1, \dots, T_n) > t) &= P(T_1 > t, \dots, T_n > t) \\ &= \prod_{i=1}^n P(T_i > t) = \prod_{i=1}^n e^{-\lambda_i t} = e^{-(\lambda_1 + \dots + \lambda_n)t}. \end{aligned} \quad (2.2.1)$$

Da questo ragionamento, possiamo concludere che $\min(T_1, \dots, T_n)$ segue una distribuzione esponenziale con parametro pari alla somma dei parametri di tutti i parametri $\lambda = \sum_{i=1}^n \lambda_i$.

Da questo ragionamento siamo riusciti a capire qual è la durata di una "corsa" tra n variabili esponenziali. Vogliamo adesso capire "chi finisce prima". Considerando S e T due variabili esponenziali rispettivamente di parametri λ e μ , riconduciamoci al caso in due variabili; studiamo il problema rispetto ad S sfruttando l'indipendenza e le formule di densità e di distribuzione per ottenere

$$\begin{aligned} P(S < T) &= \int_0^\infty f_S(s) P(T > S) ds \\ &= \int_0^\infty \lambda e^{-\lambda s} e^{-\mu s} ds \\ &= \int_0^\infty \lambda e^{-s(\lambda + \mu)} ds \\ &= \frac{\lambda}{\lambda + \mu} \int_0^\infty (\lambda + \mu) e^{-(\lambda + \mu)s} ds = \frac{\lambda}{\lambda + \mu}, \end{aligned}$$

dove nell'ultimo passaggio abbiamo usato il fatto che $(\lambda + \mu)e^{-(\lambda+\mu)s} \mathbb{1}_{[0,+\infty)}$ è una funzione di densità, quindi l'integrale tra 0 e ∞ è pari ad 1.

Detto ciò, enunciamo la seguente proposizione per ricondurci al caso in più variabili.

Proposizione 2.2.1. *Siano T_i con $i = 1, \dots, n$ variabili esponenziali di parametro λ_i , allora la probabilità che una specifica variabile T_k sia la prima a raggiungere il valore minimo può essere espressa come:*

$$P(T_k = \min(T_1, \dots, T_n)) = \frac{\lambda_k}{\sum_{i=1}^n \lambda_i}. \quad (2.2.2)$$

Dimostrazione. Siano $S = T_i$ e $U = \min_{j \neq i}(T_j)$. La (2.2.1) implica che U sia una variabile esponenziale di parametro $\mu = (\lambda_1 + \dots + \lambda_n) - \lambda_i$; scrivendo μ in questa forma ci siamo ricondotti al caso in due variabili, perciò

$$P(T_i = \min(T_1, \dots, T_n)) = P(S < U) = \frac{\lambda_i}{\lambda_i + \mu} = \frac{\lambda_i}{\lambda_1 + \dots + \lambda_n}$$

e la dimostrazione è conclusa. \square

Teorema 2.2.2. *Siano $\tau_1, \tau_2, \dots, \tau_n$ variabili esponenziali indipendenti con parametro λ . La somma $T_n = \tau_1 + \dots + \tau_n$ è una variabile di legge Gamma(n, λ). In particolare la funzione di densità di T_n è data da*

$$f_{T_n} = \lambda e^{-\lambda t} \frac{(\lambda t)^{n-1}}{(n-1)!} \quad (2.2.3)$$

per ogni $t \geq 0$ e nulla altrimenti.

Dimostrazione. La dimostrazione si fa per induzione. Preso $n = 1$, T_1 segue una distribuzione esponenziale di parametro λ e la sua funzione di densità è

$$f_{T_1}(t) = \lambda e^{-\lambda t}.$$

Per eseguire il passo induttivo, supponiamo che la formula sia vera per n e quindi che $T_{n+1} = T_n + \tau_{n+1}$. Suddividendo il problema rispetto a T_n e usando l'indipendenza di T_n e τ_{n+1} , otteniamo

$$f_{T_{n+1}}(t) = \int_{-\infty}^{\infty} f_{T_n}(s) f_{\tau_{n+1}}(t-s) ds.$$

Sostituendo la (2.2.3) nel primo termine e la funzione di densità dell'esponenziale nel secondo, si avrà che

$$f_{T_{n+1}}(t) = \int_0^t \lambda e^{-\lambda s} \frac{(\lambda s)^{n-1}}{(n-1)!} \lambda e^{-\lambda(t-s)} ds = e^{-\lambda t} \lambda^{n+1} \int_0^t \frac{s^{n-1}}{(n-1)!} ds = \lambda e^{-\lambda t} \frac{\lambda^n t^n}{n!}$$

concludendo la dimostrazione. \square

2.3 Tasso di transizione e risultati derivati

Nelle catene di Markov omogenee a tempo continuo, rappresentate da $X = \{X(t), t \geq 0\}$, le transizioni non avvengono ad ogni istante temporale discreto, ma seguono un modello in cui il processo rimane fermo per un periodo di tempo esponenzialmente distribuito prima di muoversi. Introduciamo quindi la notazione $T_i \sim \text{Exp}(v_i)$ per denotare la variabile aleatoria che rappresenta il tempo di permanenza nello stato i , dove v_i indica il tasso con cui il processo effettua una transizione quando si trova in i . Inoltre, definiamo P_{ij} come la probabilità che la catena effettui la transizione dallo stato i allo stato j .

Definizione 2.3.1 (Tassi di transizione). Per ogni coppia di stati i, j , noti il tasso v_i e la probabilità P_{ij} che dallo stato i si passi allo stato j , definiamo con

$$q_{ij} = v_i P_{ij} \quad (2.3.1)$$

il tasso di transizione istantaneo, ovvero il tasso con cui la catena, attualmente nello stato i , effettua una transizione nello stato j .

Da quanto detto, se effettuiamo una sommatoria su j tenendo fisso i , possiamo scrivere che

$$v_i = \sum_j v_i P_{ij} = \sum_j q_{ij} \quad (2.3.2)$$

dove la prima uguaglianza è giustificata dal fatto che $\sum_j P_{ij} = 1$; vediamo poi che

$$P_{ij} = \frac{q_{ij}}{v_i} = \frac{q_{ij}}{\sum_j q_{ij}}. \quad (2.3.3)$$

Da tale relazione si evince una proprietà fondamentale delle catene di Markov a tempo continuo: le probabilità di transizione da uno stato i a uno stato j possono essere completamente determinate a partire dal tasso istantaneo di transizione tra i due stati.

Lemma 2.3.2.

$$\lim_{h \rightarrow 0} \frac{1 - P_{ii}(h)}{h} = v_i \quad (2.3.4)$$

$$\lim_{h \rightarrow 0} \frac{P_{ij}(h)}{h} = q_{ij} \quad \text{con } i \neq j \quad (2.3.5)$$

Dimostrazione. Notiamo che $1 - P_{ii}(h)$ in (2.3.4) rappresenta la probabilità che un processo, inizialmente in i , si trovi in uno stato differente dopo un intervallo di tempo h ; ciò è come dire che avverrà almeno una transizione nell'intervallo di tempo $h + o(h)$. Formalizzando tale ragionamento avremo

$$1 - P_{ii}(h) = v_i h + o(h), \quad (2.3.6)$$

concludendo così la dimostrazione della prima parte del lemma. Per quanto riguarda l'equazione (2.3.5), notiamo che il valore $P_{ij}(h)$ può anche essere espresso come la probabilità

che, nell'intervallo di tempo h , si verifichi una transizione e che essa porti la catena esattamente dallo stato i allo stato j , sommando al tutto un intervallo di tempo infinitesimo. In formula:

$$P_{ij}(h) = hv_i P_{ij} + o(h) \quad (2.3.7)$$

ed anche la seconda parte del lemma è conclusa. \square

Lemma 2.3.3 (Equazioni di Chapman-Kolmogorov). *Per ogni $s \geq 0$, $t \geq 0$,*

$$P_{ij}(t+s) = \sum_{k=0}^{\infty} P_{ik}(t)P_{kj}(s).$$

Dimostrazione. La dimostrazione è analoga a quella del Lemma 1.3.1. \square

Combinando quanto visto nei Lemmi 2.3.2 e 2.3.3 ed assumendo l'interscambiabilità di limiti e sommatorie otteniamo i seguenti teoremi:

Teorema 2.3.4 (Equazioni backward di Kolmogorov). *Per ogni stato i, j , e per ogni $t \geq 0$,*

$$P'_{ij}(t) = \sum_{k \neq i} q_{ik} P_{kj}(t) - v_i P_{ij}(t).$$

Teorema 2.3.5 (Equazioni forward di Kolmogorov). *Per ogni stato i, j , e per ogni $t \geq 0$,*

$$P'_{ij}(t) = \sum_{k \neq j} q_{kj} P_{ik}(t) - v_j P_{ij}(t).$$

Osservazione 2.3.6. *L'ipotesi di interscambiabilità è sempre valida nelle equazioni backward, mentre è ammissibile nelle forward solamente sotto opportune condizioni di regolarità.*

Parte II

Modello SIR e Applicazioni

Capitolo 3

Introduzione al modello SIR

Quanto scritto nel presente capitolo è riportato in [Allen \[2008\]](#).

3.1 Premesse

Il modello SIR è un modello epidemiologico che descrive la diffusione di una malattia infettiva attraverso la partizione della popolazione in 3 gruppi (*Suscettibili* (S), *Infetti* (I) e *Rimossi* (R)) e le dinamiche di spostamento degli individui tra di essi. In particolare, ogni individuo appartenente alla categoria dei suscettibili, entrando in contatto con un infetto, ha una certa probabilità di contrarre la malattia e quindi di passare al gruppo degli infetti; dopo un determinato periodo di tempo, o a causa della morte dell'individuo o per lo sviluppo dell'immunità di quest'ultimo, passerà al gruppo dei rimossi.

Le dinamiche del modello SIR sono regolate da diversi parametri determinati in modo specifico per ogni malattia, poiché riflettono le sue caratteristiche e le sue dinamiche di diffusione. Nel nostro caso indicheremo con

- $\beta \geq 0$ il **tasso di contagio**, ovvero la velocità di trasmissione della malattia da un individuo infetto ad uno suscettibile, interpretabile anche come il numero medio di suscettibili che un infetto può contagiare in ogni istante.
- $\gamma \geq 0$ il **tasso di rimozione**, ovvero la velocità con cui un individuo passa al gruppo dei rimossi (il suo inverso $1/\gamma$ rappresenta la durata media della malattia).

Definiamo le seguenti funzioni del tempo:

- $S(t)$ rappresenta il numero di suscettibili ad ogni istante;
- $I(t)$ rappresenta il numero di infetti ad ogni istante;
- $R(t)$ rappresenta il numero di rimossi ad ogni istante.

Supponendo che la popolazione totale N resti costante nel tempo, possiamo scrivere

$$N = S(t) + I(t) + R(t) \quad \forall t \geq 0. \quad (3.1.1)$$

3.2 Analisi del modello deterministico

Nel modello deterministico, utilizziamo la convenzione che la popolazione sia molto grande e si possano rappresentare $S(t)$, $I(t)$ e $R(t)$ con numeri reali. Questo perchè immaginiamo $S(t) = Ns(t)$, $I(t) = Ni(t)$ e $R(t) = Nr(t)$ dove r, s, i soddisfano $r(t) + s(t) + i(t) = 1$ e rappresentano la proporzione di appartenenti ad ogni compartimento. Dire che la popolazione N è costante nel tempo, nel modello deterministico equivale a scrivere che

$$\frac{dN}{dt} = \frac{dS(t)}{dt} + \frac{dI(t)}{dt} + \frac{dR(t)}{dt} = 0. \quad (3.2.1)$$

Le seguenti equazioni differenziali descrivono invece le evoluzioni temporali del modello SIR in questione:

$$\frac{dS}{dt} = -\frac{\beta}{N}SI \quad (3.2.2)$$

$$\frac{dI}{dt} = \frac{\beta}{N}SI - \gamma I \quad (3.2.3)$$

$$\frac{dR}{dt} = \gamma I \quad (3.2.4)$$

imponendo le condizioni iniziali $S(0) \geq 0, I(0) \geq 0, R(0) \geq 0$. Notiamo che, sommando le tre equazioni differenziali, la (3.2.1) è immediatamente verificata.

Analizziamole nello specifico:

- Nella (3.2.2) il termine $-\frac{\beta}{N}SI$, ovvero il prodotto tra la velocità di infezione β normalizzato rispetto alla dimensione della popolazione N ed il numero totale di contatti avvenuti tra suscettibili S e infetti I , spiega una diminuzione del numero di suscettibili quando questi incontrano gli infetti.
- Nella (3.2.3) il valore $-\frac{\beta}{N}SI$ della prima equazione viene cambiato di segno in quanto rappresenta la velocità con cui gli individui passano dal gruppo dei suscettibili a quello degli infetti, che aumentano. Mentre il termine $-\gamma I$ indica che la quantità di individui infetti sta diminuendo poiché questi ultimi diventano rimossi.
- Infine, nella (3.2.4) il termine $-\gamma I$ cambia di segno in quanto gli individui infetti passano al gruppo dei rimossi seguendo il tasso di guarigione γ .

Definiamo ora con R_0 il **numero di riproduzione di base**, ovvero il numero medio di nuovi casi di infezione generati da un singolo individuo infetto in una popolazione completamente suscettibile, durante il suo periodo infettivo:

$$R_0 = \frac{\beta}{\gamma}; \quad (3.2.5)$$

la quantità $\frac{1}{\gamma}$ rappresenta la durata media dell'infezione (fino a guarigione o morte, anche per ragioni non legate alla malattia in questione).

Considerando che con $R_0 \frac{S(0)}{N}$ si indica il **numero di sostituzione iniziale**, ovvero il numero medio di secondi contagi causati da un individuo infetto durante il suo periodo di infezione all'inizio del diffondersi della malattia, il seguente teorema esprime come l'indice R_0 sia strettamente legato alle dinamiche del sistema.

Teorema 3.2.1. *Siano $S(t), R(t), I(t)$ soluzioni del modello SIR,*

1. *se $R_0 \leq 1$, allora la malattia si estinguerà completamente dalla popolazione raggiungendo un equilibrio libero dalla malattia.*
2. *Se $R_0 > 1$, si raggiungerà un equilibrio endemico, ovvero la malattia persisterà all'interno della popolazione in modo costante senza causare né un aumento esponenziale di nuove infezioni né la sua totale estinzione nel lungo periodo.*
3. *Se $R_0 \frac{S(0)}{N} > 1$ allora si verificherà un incremento iniziale degli individui infetti, determinando un comportamento epidemico; al contrario, se $R_0 \frac{S(0)}{N} \leq 1$, il numero di individui infetti $I(t)$ decrescerà gradualmente nel tempo. In entrambi i casi $I(t)$ convergerà a zero, indicando così un equilibrio libero da malattia.*

3.3 Modello SIR stocastico a tempo continuo

Nella corrente sezione, esamineremo il modello SIR dal punto di vista stocastico, considerando la variabile temporale t un valore reale positivo. In particolare, $\{(S(t), I(t), R(t)); t \geq 0\}$ è modellizzato da una CTMC (Continuous-Time Markov Chain) ed è costituito da due variabili indipendenti $S(t), I(t)$. Dalla seguente combinazione lineare è infatti possibile determinare la variabile dipendente $R(t)$:

$$R(t) = N - S(t) - I(t).$$

Possiamo dunque considerare la funzione di densità congiunta associata al processo bivariato $\{(S(t), I(t)); t \geq 0\}$, ad ogni istante temporale $t \geq 0$ fissato:

$$p_{(s,i)}(t) = P\{S(t) = s, I(t) = i\}. \quad (3.3.1)$$

Per definire la probabilità di transizione del modello SIR a tempo continuo, è necessario considerare un intervallo di tempo h sufficientemente piccolo tale che, durante quell'intervallo, avvenga al più un cambiamento di stato. In altre parole, deve verificarsi al più una nuova infezione o una guarigione.

Definizione 3.3.1 (Probabilità di transizione nel modello SIR). *Siano $S(t) = s$ e $I(t) = i$ rispettivamente il numero di suscettibili e di infetti all'istante di tempo t , la probabilità che $S(t+h) = s+k$ e $I(t+h) = i+j$ siano rispettivamente il numero di suscettibili ed infetti dopo un intervallo di tempo h è*

$$P_{(s,i)(s+k,i+j)}(h) = P\{(S(t+h), I(t+h)) = (s+k, i+j) | (S(t), I(t)) = (s, i)\} \quad (3.3.2)$$

rappresenta la probabilità di transizione della catena di Markov.

Visto che nel modello SIR preso in considerazione abbiamo scelto di uguagliare il tasso di nascita con il tasso di morte, è importante mettere in evidenza che una nascita si verifica se e solo se c'è anche una morte. La differenza fondamentale tra i due eventi è che le nascite interessano solamente l'insieme dei suscettibili S (un individuo appena nato è per definizione suscettibile alla malattia), mentre le morti per cause diverse dalla malattia che abbiamo opreso in esame possono verificarsi in una qualsiasi partizione della popolazione. Le possibili transizioni sono quindi individuate dai seguenti valori della coppia (k, j) :

- $(k, j) = (-1, 1)$: infezione di un individuo suscettibile (i suscettibili diminuiscono, gli infetti aumentano);
- $(k, j) = (0, -1)$: guarigione di un individuo infetto (i suscettibili non variano, gli infetti diminuiscono ed aumentano i rimossi, ma non lo evidenziamo);

La probabilità di transizione può essere quindi espressa in forme differenti, usando le stesse regole viste per il modello deterministico, nel modo seguente:

$$P_{(s,i)(s+k,i+j)}(h) = \begin{cases} (\beta is/N)h + o(h), & \text{se } (k, j) = (-1, 1) \\ \gamma ih + o(h), & \text{se } (k, j) = (0, -1) \\ 1 - (\beta is/N + \gamma i)h + o(h), & \text{se } (k, j) = (0, 0) \\ 0, & \text{altrimenti.} \end{cases} \quad (3.3.3)$$

Ricordando che h deve essere scelto sufficientemente piccolo tale che la probabilità di transizione sia un numero compreso in $[0, 1]$, notiamo che ogni caso la probabilità di transizione deve essere costituita da un termine $o(h)$ che tiene conto dell'azione di altri fattori che potrebbero influenzare il sistema ma che vengono considerati trascurabili rispetto a quelli principali. Da (3.3.3) mettiamo in evidenza i tassi di transizione q_{ij} la cui trattazione è stata discussa al paragrafo 2.3:

$$q_{(s,i)(s+k,i+j)} = \begin{cases} \beta \frac{is}{N}, & \text{se } (k, j) = (-1, 1) \\ \gamma i, & \text{se } (k, j) = (0, -1). \end{cases} \quad (3.3.4)$$

Capitolo 4

Algoritmi di simulazione deterministici

Il capitolo corrente è dedicato all'analisi di due metodi deterministici impiegati per la risoluzione di equazioni differenziali ordinarie: il metodo di Eulero esplicito e il metodo di Runge-Kutta. Questi approcci, fondamentali nell'analisi numerica, saranno qui brevemente discussi, con riferimenti per ulteriori approfondimenti disponibili in [Burden and Faires \[2011\]](#).

4.1 Metodo di Eulero esplicito

Il metodo di Eulero esplicito è un metodo numerico per la risoluzione di equazioni differenziali ordinarie che consiste nell'approssimare la soluzione della funzione incognita avanzando a piccoli passi temporali. Partendo da una condizione iniziale y_0 al tempo t_0 , definiamo

$$\frac{dy}{dt} = f(t, y) \quad (4.1.1)$$

dove $f(t, y)$ è una funzione opportuna. Nello specifico, per passare dal valore y_n al tempo t_n al valore successivo y_{n+1} al tempo t_{n+1} , si utilizza la formula

$$y_{n+1} = y_n + hf(t_n, y_n), \quad (4.1.2)$$

dove h rappresenta il passo temporale. Questo metodo viene definito esplicito perchè il nuovo valore y_{n+1} viene calcolato direttamente dalla conoscenza del valore precedente y_n e dalla funzione f al tempo corrente. Il costo computazionale del metodo di Eulero esplicito è proporzionale al numero di passi temporali necessari per coprire l'intervallo di integrazione desiderato. Ogni passo richiede il calcolo della funzione $f(t_n, y_n)$ e un'operazione di somma, rendendo il metodo piuttosto efficiente dal punto di vista del tempo di esecuzione. Tuttavia, il metodo di Eulero ha delle limitazioni in termini di stabilità e accuratezza, specialmente quando la soluzione varia rapidamente o in presenza di problemi rigidi. La

scelta del passo h risulta quindi cruciale per garantire risultati affidabili; passi troppo grandi possono portare a instabilità numerica e errori significativi, mentre passi troppo piccoli aumentano il costo computazionale senza un beneficio proporzionale in precisione.

4.2 Metodo di Runge-Kutta esplicito a r stadi

Il metodo di Runge-Kutta a r stadi è una tecnica numerica leggermente più sofisticata del metodo di Eulero esplicito impiegata per risolvere equazioni differenziali ordinarie. A partire da una condizione iniziale y_0 al tempo t_0 , il metodo procede calcolando una serie di stadi intermedi, che sono valutazioni successive della derivata della soluzione, rappresentata dalla funzione $f(t, y)$, definita come in (4.1.1). Questi stadi intermedi, indicati con k_i per $i = 1, 2, \dots, r$, vengono determinati mediante la valutazione di f in punti specifici, scelti in base alle iterazioni precedenti e al tempo attuale. Ogni k_i è calcolato utilizzando una combinazione lineare delle valutazioni precedenti, pesata da coefficienti noti come coefficienti di Butcher. La formula generale per uno stadio intermedio può essere espressa come

$$k_i = f(t_n + c_i h, y_n + h \sum_{j=1}^{i-1} a_{ij} k_j), \quad (4.2.1)$$

dove h è il passo temporale, c_i e a_{ij} sono i coefficienti che definiscono la struttura del metodo, e y_n è l'approssimazione corrente della soluzione. Dopo aver calcolato tutti gli stadi intermedi, la soluzione approssimata al tempo successivo, y_{n+1} , è ottenuta sommando il valore corrente y_n a una combinazione ponderata degli stadi k_i , secondo la formula

$$y_{n+1} = y_n + h \sum_{i=1}^r b_i k_i, \quad (4.2.2)$$

dove b_i sono i coefficienti che ponderano ciascuno stadio. Questi coefficienti sono scelti per ottimizzare la precisione dell'approssimazione, minimizzando l'errore di troncamento locale.

Il metodo di Runge-Kutta a r stadi risulta essere quindi un metodo più avanzato e preciso rispetto al metodo di Eulero esplicito poiché impiega più valutazioni della derivata della funzione all'interno di ciascun passo temporale. Questo permette di ottenere una soluzione con un livello di accuratezza superiore e una maggiore stabilità. Tuttavia, il vantaggio in termini di precisione comporta un costo computazionale più elevato, poiché sono necessarie numerose valutazioni della funzione per ogni passo temporale. Di conseguenza, mentre il metodo di Runge-Kutta fornisce risultati più accurati e affidabili, il metodo di Eulero esplicito può essere più vantaggioso per problemi meno complessi o quando si ha a disposizione una capacità computazionale limitata.

Capitolo 5

Algoritmi di simulazione stocastici

Esistono numerosi metodi stocastici che possono essere impiegati per eseguire simulazioni numeriche di modelli basati sulle CTMC, ognuno dei quali presenta caratteristiche e specificità che possono renderlo più o meno adatto a determinati problemi (esigenze di precisione, efficienza computazionale...). Di conseguenza, non è possibile predeterminare con certezza quale algoritmo risulterà essere il più efficace senza prima analizzare in dettaglio il contesto specifico e le esigenze della simulazione da realizzare. In questo capitolo analizzeremo solamente i seguenti:

- **l'algoritmo di Gillespie** (sviluppato da Dan Gillespie negli anni 70')
- **l'algoritmo di Tau-Leaping esplicito**

I due algoritmi sono descritti in dettaglio in [Banks et al. \[2012\]](#).

Gli algoritmi in questione sono comunemente utilizzati per simulare l'evoluzione temporale di modelli stocastici in chimica cinetica, scienza che analizza i movimenti e le interazioni delle molecole durante le reazioni chimiche. In aggiunta, essi trovano applicazione anche nella simulazione numerica di modelli stocastici che coprono un'ampia gamma di applicazioni biologiche. Un esempio è la modellazione della diffusione di un'epidemia secondo il modello SIR descritto nella sezione precedente. In questo specifico caso, definiamo la variabile di stato del sistema come:

$$X(t) = \{S(t), I(t)\} \tag{5.0.1}$$

dove si indicano il numero di individui appartenenti rispettivamente alle categorie di suscettibili e infetti all'istante di tempo t . Ricordando quanto visto nella sezione relativa ai tassi di transizione (2.3), procediamo nel descrivere i due algoritmi.

L'algoritmo di Gillespie è noto per la sua capacità di fornire una simulazione esatta delle dinamiche stocastiche dei sistemi, basata sulla considerazione di ogni singola transizione tra gli stati del sistema. Tuttavia, questa precisione implica un costo computazionale significativo, specialmente quando si affrontano modelli complessi o simulazioni a lungo

termine. Per mitigare questa problematica, è stato sviluppato l'algoritmo del tau-leaping, che conserva i principi fondamentali dell'algoritmo di Gillespie ma introduce un approccio deterministico basato sull'approssimazione del numero di transizioni che si verificano entro intervalli di tempo più grandi, noti come τ . In questo modo il costo computazionale della simulazione viene significativamente ridotto, consentendo l'analisi di modelli caratterizzati da un elevato numero di transizioni. Tuttavia, è importante sottolineare che l'uso del tau-leaping comporta una leggera perdita di precisione rispetto all'algoritmo di Gillespie.

In entrambi gli algoritmi considereremo i tassi di nascita e morte pari a zero in quanto supponiamo che la durata della malattia in esame sia molto breve, al massimo di qualche mese; in un intervallo di tempo così ristretto la popolazione cambia in maniera infinitesima a causa di nascite o decessi per altre motivazioni.

5.1 Algoritmo di Gillespie

- **Step 1:** inizializziamo lo stato del sistema $X_0 = X(0)$
- **Step 2:** calcoliamo ora i tassi di transizione $q_{(s,i)(s+k,i+j)}$ possibili a partire da uno stato (s, i) , considerando che queste ultime in un modello SIR sono fondamentalmente 2:
 1. Infezione: un individuo viene contagiato e passa dal gruppo dei suscettibili S a quello degli infetti I . Il tasso di infezione verrà indicato come $q_{(s,i)(s-1,i+1)}$
 2. Guarigione o morte: un individuo guarisce o muore a causa della malattia presa in esame, passando dal gruppo degli infetti I a quello dei rimossi R . Il tasso di guarigione è espresso come $q_{(s,i)(s,i-1)}$. Notiamo inoltre che essi non dipendono dalla variabile di stato del sistema in quanto stiamo lavorando con un catene di Markov omogenee.
- **Step 3:** calcoliamo i tassi $v_{(s,i)}$ di uscita dallo stato (s, i) .
- **Step 4:** simuliamo ora il tempo τ che manca alla transizione successiva seguendo una distribuzione esponenziale di media $\frac{1}{v_{(s,i)}}$. Per generare il tempo τ usiamo la formula

$$\tau = \frac{-\ln(r_1)}{v_{(s,i)}} \quad (5.1.1)$$

dove r_1 è un numero casuale uniformemente distribuito tra 0 e 1. Questa è una tecnica standard per generare variabili aleatorie da una distribuzione esponenziale.

- **Step 5:** identifichiamo quale delle possibili transizioni si verificherà, ricordando che ad ognuna di esse è associata una probabilità di transizione $P_{(s,i)(s+k,i+j)}$ (2.3.3). Dopo aver generato un nuovo numero casuale r_2 uniformemente distribuito tra 0 e 1, dobbiamo decidere quale transizione avverrà tra le possibili transizioni dallo stato (s, i) . Per fare questo, sommiamo le probabilità delle transizioni disponibili in modo ordinato; procedendo in questo modo, ne troveremo una che, sommata a tutte le precedenti, porterà a superare il valore di r_2 . La transizione associata a tale probabilità sarà quella che si verifica.

- **Step 6:** aggiorniamo il nuovo tempo $t = t + \tau$ e il nuovo stato del sistema.
- **Step 7:** iteriamo gli Step 2-6 finchè vale

$$t \leq t_{stop}.$$

5.2 Algoritmo del tau-leaping

Prima di descrivere step-by-step l'algoritmo di tau-leaping è necessario dare la seguente definizione.

Definizione 5.2.1 (Transizione critica). *Data una transizione $(s, i) \rightarrow (s + m_1, i + m_2)$, noto $\mathbf{m} = (m_1, m_2)$ il vettore che rappresenta la variazione di individui di ciascuna categoria, definiamo $L_{\mathbf{m}}$ come segue:*

$$L_{\mathbf{m}} = L_{\mathbf{m}}(s, i) = \min \left\{ \left\lfloor \frac{s}{\|\mathbf{m}\|} \right\rfloor, \left\lfloor \frac{i}{\|\mathbf{m}\|} \right\rfloor \right\}$$

dove $\|\mathbf{m}\|$ è la norma del vettore \mathbf{m} e $\lfloor \cdot \rfloor$ è la funzione parte intera. L'equazione ci indica dunque il numero massimo di volte che la transizione può verificarsi prima che una qualsiasi componente di $X(t)$ diventi negativa. La transizione si dice critica se, fissato un certo limite n_c , si ha

$$L_{\mathbf{m}} \leq n_c.$$

Procediamo ora con la descrizione dell'algoritmo:

- **Step 1:** identifichiamo le transizioni critiche e raccogliamole nell'insieme J_{cr} definito quindi come

$$J_{cr} = \{\mathbf{m} | L_{\mathbf{m}} < n_c\}; \quad (5.2.1)$$

le transizioni non critiche saranno invece nell'insieme J_{ncr} di seguito

$$J_{ncr} = \{\mathbf{m} | L_{\mathbf{m}} \geq n_c\}. \quad (5.2.2)$$

- **Step 2:** calcoliamo ora un intervallo τ_1 durante il quale le transizioni possono essere simulate approssimativamente senza causare grandi errori nel comportamento del sistema. Tale intervallo dovrebbe essere sufficientemente piccolo per mantenere l'accuratezza, ma abbastanza grande da permettere una simulazione efficiente. Definiamo ora i valori medi e le varianze relative ai suscettibili e agli infetti:

$$\begin{aligned} \mu_S(s, i) &= \sum_{\mathbf{m} \in J_{ncr}} m_1 q_{(s, i)(s+m_1, i+m_2)} \\ \mu_I(s, i) &= \sum_{\mathbf{m} \in J_{ncr}} m_2 q_{(s, i)(s+m_1, i+m_2)} \\ \sigma_S^2(s, i) &= \sum_{\mathbf{m} \in J_{ncr}} m_1^2 q_{(s, i)(s+m_1, i+m_2)} \\ \sigma_I^2(s, i) &= \sum_{\mathbf{m} \in J_{ncr}} m_2^2 q_{(s, i)(s+m_1, i+m_2)} \end{aligned}$$

Indichiamo con ϵ un valore che controlli l'entità dell'errore introdotto nell'approssimazione delle transizioni ed i parametri g_s e g_i che dipendono dal contesto specifico del modello, possiamo trovare τ_1 seguendo la seguente formulazione

$$\tau_1 = \min \left\{ \frac{\max\{\epsilon s/g_s\}}{\mu_S(s, i)}, \frac{\max\{\epsilon i/g_i\}}{\mu_I(s, i)}, \frac{\max\{\epsilon s/g_s, 1\}^2}{\sigma_S^2(s, i)}, \frac{\max\{\epsilon i/g_i, 1\}^2}{\sigma_I^2(s, i)} \right\}. \quad (5.2.3)$$

- **Step 3:** calcoliamo il tasso totale di tutte le transizioni $v_{(s,i)}$ e confrontiamolo con τ_1 : fissato un $k \in \mathbb{N}$
 - se $\tau_1 > \frac{k}{v_{(s,i)}}$, allora procediamo con il metodo del tau-leaping usando τ_1 come intervallo temporale;
 - se $\tau_1 < \frac{k}{v_{(s,i)}}$ allora eseguiamo un numero $N \in \mathbb{N}$ di iterate con il metodo di Gillespie e ripetiamo il calcolo di τ_1 (tornando quindi allo Step 2).
- **Step 4:** calcoliamo il tasso totale di tutte le transizioni critiche

$$v_{(s,i)}^{cr} = \sum_{\mathbf{m} \in J_{cr}} q_{(s,i)(s+m_1, i+m_2)}.$$

e generiamo un secondo tempo di salto candidato, indicato come τ_2 e ottenuto campionando una variabile aleatoria esponenziale con valore atteso pari a $1/v_{(s,i)}^{cr}$

- **Step 5:** chiamiamo $\tau = \min\{\tau_1, \tau_2\}$. Approssimiamo il numero di transizioni $K_{\mathbf{m}}$ all'interno dell'intervallo di tempo τ campionando una variabile aleatoria di Poisson con valore atteso $q_{(s,i)(s+m_1, i+m_2)}\tau$ per ogni $\mathbf{m} \in J_{ncr}$. Per quanto riguarda invece le transizioni critiche, approssimiamo $K_{\mathbf{m}}$ come segue:
 - se $\tau = \tau_1$, allora $K_{\mathbf{m}} = 0$ per ogni $\mathbf{m} \in J_{cr}$ (quindi non ci sono transizioni critiche);
 - se $\tau = \tau_2$, allora chiamiamo \mathbf{m}^{cr} un campione della variabile aleatoria intera con probabilità puntuale $\frac{q_{(s,i)(s+m_1, i+m_2)}}{v_{(s,i)}^{cr}}$ per ogni $\mathbf{m} \in J_{cr}$. Impostiamo quindi $K_{\mathbf{m}^{cr}} = 1$ (dove \mathbf{m}^{cr} è l'unica transizione critica che si verifica) e $K_{\mathbf{m}} = 0$ per ogni $\mathbf{m} \in J_{cr}$ tale che $\mathbf{m} \neq \mathbf{m}^{cr}$.
- **Step 6:** prendiamo in considerazione il vettore $x + \sum_{\mathbf{m}} K_{\mathbf{m}} v_{\mathbf{m}}$. Se esso risulta avere una componente negativa, allora dimezziamo l'intervallo di tempo τ_1 e torniamo allo Step 3. Altrimenti, procediamo con il salto aggiornando il tempo $t = t + \tau$ e il nuovo stato del sistema

$$x(t + \tau) = x + \sum_{\mathbf{m}} K_{\mathbf{m}} v_{\mathbf{m}}.$$

- **Step 7:** iteriamo gli Step 1-6 finchè risulta valida la condizione $t \geq t_{stop}$.

Capitolo 6

Applicazione: Confronto tra varianti di COVID-19

In questo capitolo, applicheremo le catene di Markov per esaminare come due varianti del COVID-19 si diffondono nel tempo. Utilizzeremo le tecniche discusse in precedenza per confrontare l'andamento e l'impatto delle diverse varianti del virus sul progresso della malattia.

Di seguito ci proponiamo di analizzare i tempi e le velocità con cui una determinata variante, indicata come variante 1, cederà il posto a una nuova, denominata variante 2. Le simulazioni presentate si concentreranno sull'assegnazione di diversi valori ai tassi di contagio e di rimozione per entrambe le varianti, con l'obiettivo di rispondere a due domande chiave:

- **Quesito 1:** affinché tutti gli individui infetti siano infettati dalla variante 2 dopo un certo intervallo di tempo, è necessario che $R_{0,1} < 1$ oppure è sufficiente che $R_{0,2} \gg R_{0,1}$ anche se $R_{0,1} > 1$ (con $R_{0,1}$ e $R_{0,2}$ numeri di riproduzione di base delle rispettive varianti)?
- **Quesito 2:** qual è la probabilità che al tempo t un individuo si infetti con la variante 2 piuttosto che con la variante 1?

6.1 Definizione dei dati

Nell'analisi delle due varianti del virus, assumeremo che i tassi di contagio e di rimozione restino costanti, ossia non varino nell'intervallo di tempo considerato per il passaggio da una variante all'altra. Tuttavia, è importante notare che, realisticamente, tali tassi dovrebbero dipendere dal tempo, poiché il virus si adatta continuamente all'ambiente circostante e la popolazione adotta misure di sicurezza sempre più rigorose con l'aumento della diffusione (ad esempio, obbligo di mascherina, distanziamento sociale, coprifuoco, ecc.). In questo capitolo assumeremo la seguente notazione:

- N popolazione totale
- $S(t)$ popolazione suscettibile all'istante $t \geq 0$
- $I_1(t)$ individui infetti dalla variante 1 al tempo $t \geq 0$;
- $I_2(t)$ individui infetti dalla variante 2 al tempo $t \geq 0$;
- $R_1(t)$ individui rimossi al tempo $t \geq 0$ dopo esser stati infetti dalla variante 1;
- $R_2(t)$ individui rimossi al tempo $t \geq 0$ dopo esser stati infetti dalla variante 2.
- β_1, β_2 i tassi di contagio;
- γ_1, γ_2 i tassi di rimozione
- $R_{0,1}, R_{0,2}$ i numeri di riproduzione di base definiti come

$$R_{0,1} = \frac{\beta_1}{\gamma_1} \quad R_{0,2} = \frac{\beta_2}{\gamma_2}. \quad (6.1.1)$$

Il nuovo modello SIR che seguiremo è un'estensione di quello visto nel paragrafo 3.2 ed è composto dalle seguenti equazioni:

$$\frac{dS(t)}{dt} = -\frac{\beta_1}{N}S(t)I_1(t) - \frac{\beta_2}{N}S(t)I_2(t); \quad (6.1.2)$$

$$\frac{dI_1(t)}{dt} = \frac{\beta_1}{N}S(t)I_1(t) - \gamma_1 I_1(t); \quad (6.1.3)$$

$$\frac{dI_2(t)}{dt} = \frac{\beta_2}{N}S(t)I_2(t) - \gamma_1 I_2(t); \quad (6.1.4)$$

$$\frac{dR_1(t)}{dt} = \gamma_1 I_1(t); \quad (6.1.5)$$

$$\frac{dR_2(t)}{dt} = \gamma_1 I_2(t). \quad (6.1.6)$$

Come già anticipato nella prima parte, le simulazioni considereranno che i tassi di contagio della variante iniziale siano sempre inferiori a quelli della variante successiva. Questo comportamento riflette una tendenza tipica dei virus, dove le nuove varianti, pur essendo spesso meno letali, tendono a diffondersi con maggiore rapidità rispetto alle precedenti. Per quanto riguarda i tassi di guarigione invece è necessario avere maggiori accortezze: secondo la maggior parte degli studi effettuati fino ad oggi, il tempo medio necessario affinché un individuo non sia più infetto dalla variante 2 è inferiore rispetto a quello richiesto per la variante 1, ma restano comunque ancora molti dubbi sulla questione.

Il linguaggio di programmazione adottato è Python.

6.2 Simulazioni con il Metodo di Eulero

Nella sezione corrente valuteremo il modello SIR attraverso il metodo di Eulero esplicito, illustrato nel paragrafo 4.1. Nel codice presente in A.1 è stata scelta una

popolazione totale iniziale composta da 10.000 individui, di cui 1.000 infetti dalla variante 1 e 10 dalla variante 2; la simulazione viene effettuata in un intervallo di tempo di 120 giorni con un passo temporale $dt = 0.1$; i rimossi all'istante iniziale sono pari a zero; il tasso di contagio della variante 1 (β_1) è $1/15$ mentre quello della variante 2 (β_2) è $1/2$; i tassi di rimozione della variante 1 (γ_1) e 2 (γ_2) sono rispettivamente fissati a $1/2$ e $1/10$.

Sono state effettuate simulazioni con 3 diversi valori di β_1 così che i numeri di riproduzione di base R_0^1 possano assumere valori differenti.

Caso 1	Caso 2	Caso 3
$\beta_1 = 1/15$	$\beta_1 = 1/12$	$\beta_1 = 1/10$
$R_0^1 = 0.8$	$R_0^1 = 1.0$	$R_0^1 = 1.25$

Tabella 6.1. Tassi di infezione e numeri di riproduzione di base

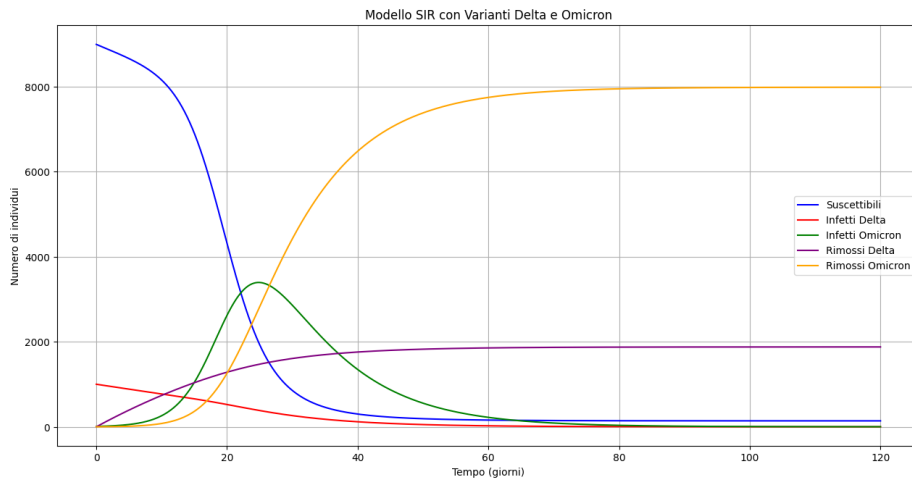


Figura 6.1. Caso $R_0^1 = 0.8$

Si osserva che, in tutti e tre i casi considerati, esiste sempre un istante temporale successivo al quale il numero di individui infetti dalla variante 2 supera quello degli infetti dalla variante 1 fino al termine della simulazione. In particolare, si osserva dalla figura 6.1 che, nel primo caso, tale istante si verifica il giorno 13, nel secondo caso, come illustrato nella 6.2, ha luogo al giorno 14 ed infine, nel terzo caso, come si vede in figura 6.3, tale istante corrisponde al giorno 15.

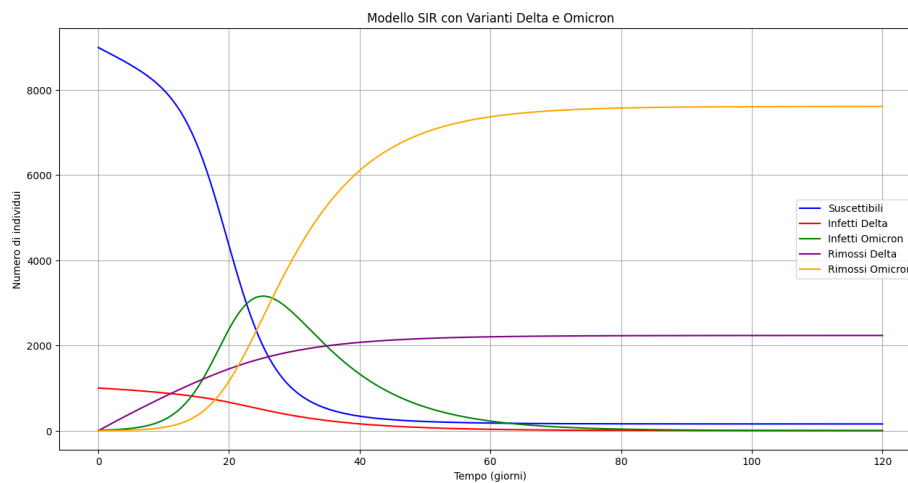


Figura 6.2. Caso $R_0^1 = 1.0$

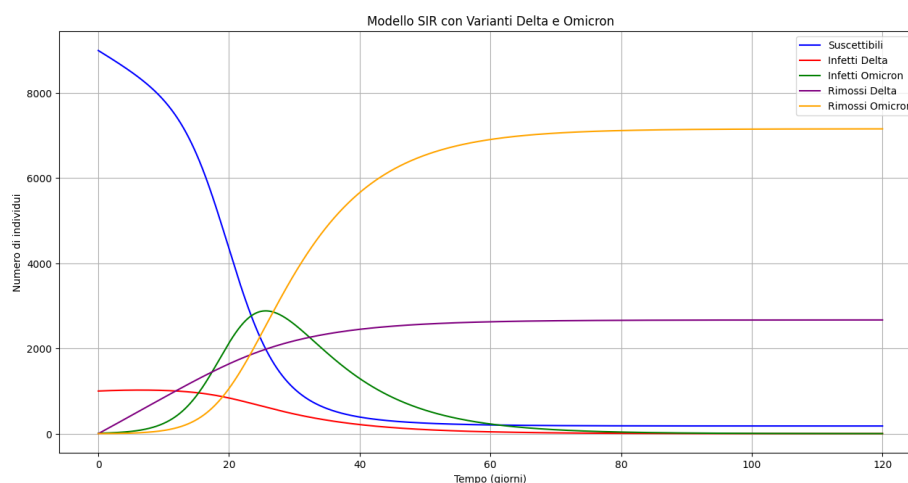


Figura 6.3. Caso $R_0^1 = 1.25$

6.3 Simulazioni con il metodo di Runge-Kutta

Visto che utilizzando un passo temporale $dt = 0.001$ nel metodo di Eulero si ottengono risultati analoghi a quelli trovati con $dt = 0.1$, possiamo dire che la discretizzazione temporale è sufficientemente fine. Questo ci consente di utilizzare il metodo di

Runge-Kutta a 4 stadi con un passo temporale maggiore, ad esempio $dt = 1$, mantenendo l'accuratezza desiderata. Di conseguenza possiamo ottenere un miglioramento computazionale significativo, riducendo il numero di passi di integrazione necessari per coprire l'intervallo temporale di interesse. Nel codice presentato in A.2 abbiamo implementato l'algoritmo di Runge-Kutta impostando i dati iniziali in modo analogo a quelli descritti nel paragrafo precedente (ovviamente fatta eccezione del passo temporale) ed abbiamo simulato tutti e tre i casi riportati nella tabella 6.1.

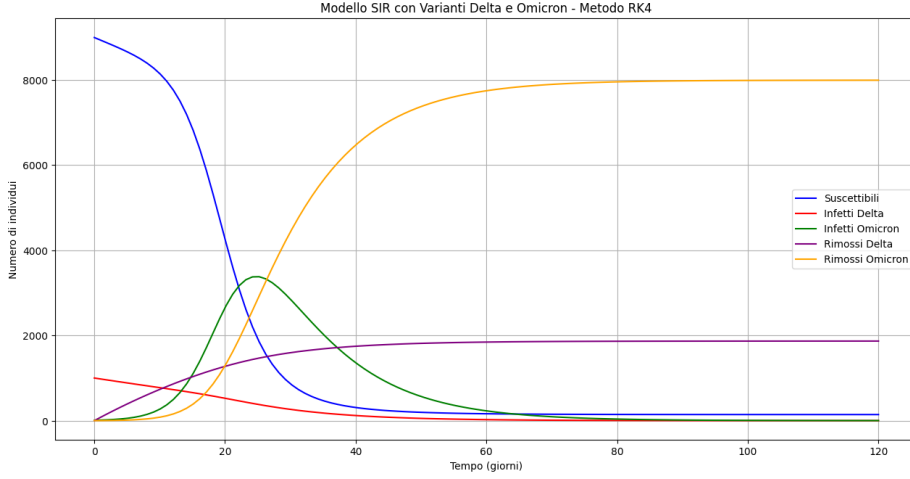


Figura 6.4. Caso $R_0^1 = 0.8$

Dopo aver confrontato i risultati ottenuti con il metodo di Eulero e il metodo di Runge-Kutta di ordine 4, si è osservato che, utilizzando un passo temporale $dt = 1$, il metodo di Runge-Kutta a 4 stadi produce risultati qualitativamente e quantitativamente simili a quelli ottenuti con il metodo di Eulero con un passo più piccolo $dt = 0.001$. Questo suggerisce che il passo temporale $dt = 1$ è sufficientemente piccolo per descrivere accuratamente il comportamento del sistema. Pertanto, è possibile adottare il metodo di Runge-Kutta con $dt = 1$ per migliorare l'efficienza computazionale senza compromettere la precisione dei risultati.

Grazie alle simulazioni effettuate abbiamo raccolto le informazioni necessarie per rispondere al primo quesito. Non è necessario che si verifichi la condizione $R_0^1 < 1$ affinché, ad un certo punto, tutti gli individui infetti siano contagiati dalla variante 2, ma basta che si abbia $R_0^2 \gg R_0^1$.

6.4 Simulazioni con il metodo di Gillespie

I metodi di Eulero e Runge-Kutta appena visti forniscono una soluzione deterministica delle variabili di stato (S , I , R) come funzioni continue nel tempo e la traiettoria

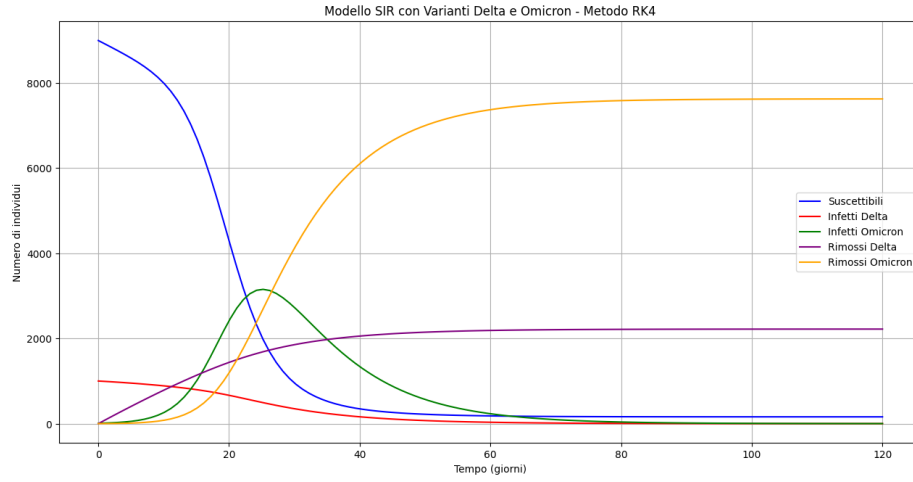


Figura 6.5. Caso $R_0^1 = 1.0$

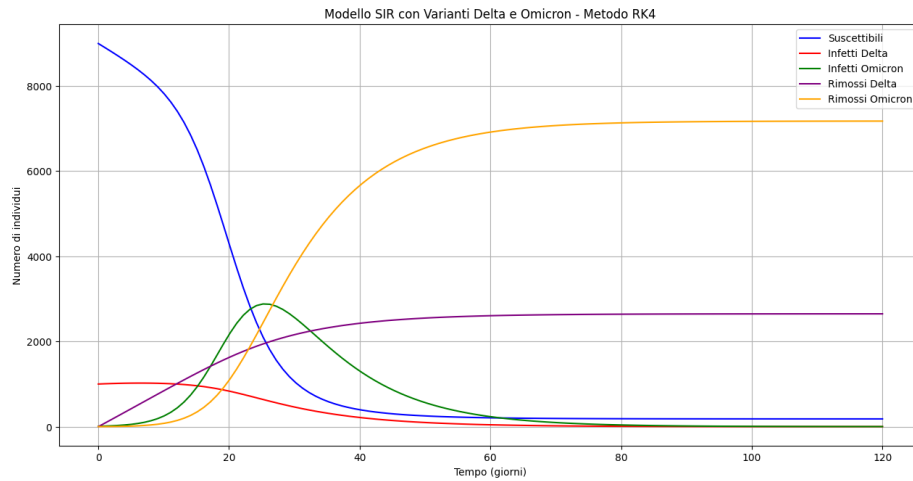


Figura 6.6. Caso $R_0^1 = 1.25$

risulta essere unica per un dato insieme di condizioni iniziali e parametri. Inoltre, la loro accuratezza dipende dal passo temporale scelto e dell'ordine del metodo. Al contrario, il metodo di Gillespie è un metodo stocastico, utilizzato per simulare sistemi che evolvono secondo dinamiche probabilistiche. Sono generate traiettorie diverse ogni volta che la simulazione viene eseguita, a causa della natura casuale del processo. Ogni esecuzione rappresenta una possibile realizzazione del sistema

stocastico.

Nel codice implementato con il metodo di Gillespie (A.3), oltre a rappresentare l'evoluzione temporale del modello SIR nelle classiche variabili S, I, R per due varianti del virus, abbiamo anche calcolato e rappresentato graficamente

- la probabilità **totale** di infezione con la variante 2 come il rapporto tra il numero di individui attualmente infetti con la variante 2 (I_2) e il numero totale di individui infetti ($I_1 + I_2$) in ogni istante temporale;
- la probabilità **giornaliera** di infezione con la variante 2 come il rapporto tra il numero di nuove infezioni giornaliere con 2 e il numero totale di nuove infezioni giornaliere.

Anche in questo caso, le simulazioni sono state eseguite con i parametri presenti nella tabella 6.1.

Nel codice sono stati considerati gli stessi parametri iniziali utilizzati in precedenza, fatta eccezione per T_{max} . Abbiamo infatti scelto un tempo massimo di simulazione di 40 giorni anziché di 120, per evitare di ottenere risultati potenzialmente fuorvianti dovuti al basso numero di individui infetti verso la fine del periodo. Man mano che il numero di infetti diminuisce, le fluttuazioni stocastiche diventano più rilevanti, il che potrebbe portare a risultati meno affidabili e più soggetti a variazioni casuali. Pertanto, abbiamo limitato la durata della simulazione per garantire che i risultati riflettano accuratamente le dinamiche epidemiche principali senza essere influenzati in modo eccessivo dall'aleatorietà associata ai bassi numeri di infetti.

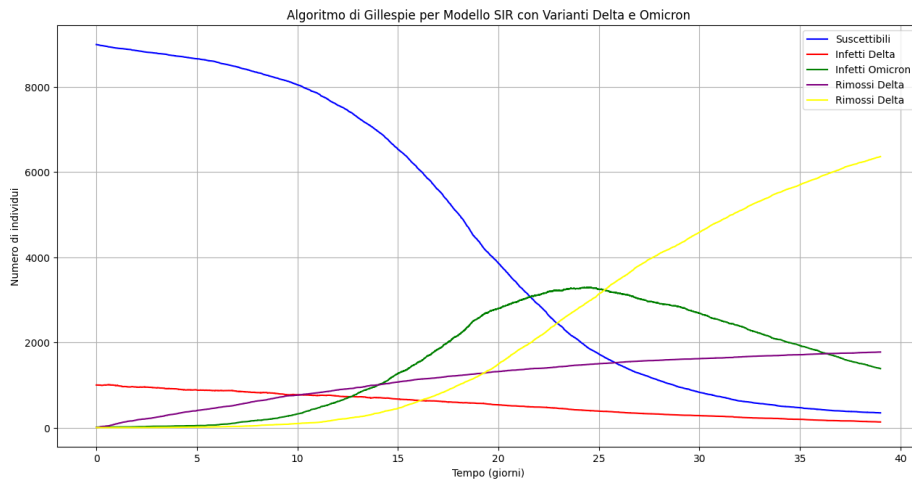


Figura 6.7. Simulazione modello SIR con metodo di Gillespie

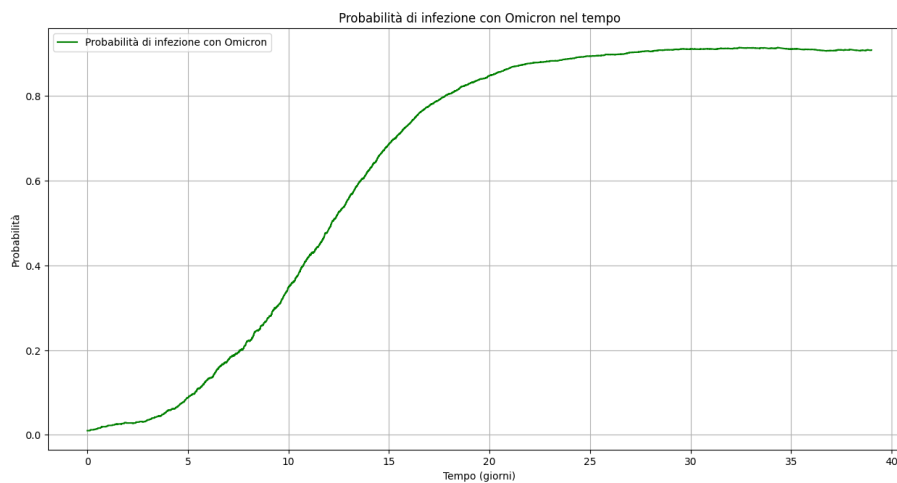


Figura 6.8. Probabilità totale di infezione con la variante 2

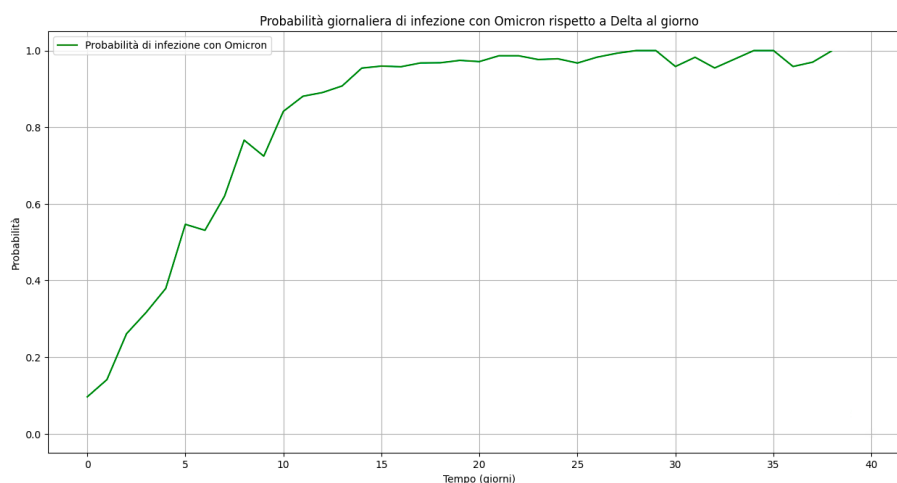


Figura 6.9. Probabilità giornaliera di infezione con la variante 2

A differenza di quanto visto con i metodi utilizzati in precedenza, qui i risultati ottenuti con i 3 casi della tabella 6.1 sono molto simili tra loro dunque, per semplicità, riporteremo solamente quelli ottenuti per $\beta_1 = 1/15$ (caso 1).

Come già detto precedentemente, non possiamo fare osservazioni dettagliate sulla simulazione dell'andamento delle tre variabili (figura 6.7), poiché i risultati variano

significativamente da una simulazione all'altra. Tuttavia, è possibile notare che, rispetto ai risultati ottenuti con i metodi di Eulero e Runge-Kutta (figure 6.1 e 6.4), le curve risultano meno lisce a causa dell'approccio stocastico. Le probabilità totali e giornaliere, illustrate rispettivamente nelle figure 6.8 e 6.9, seguono un andamento molto simile: il numero di individui infetti dalla variante 1 diminuisce fino a lasciare completamente il posto a quelli infetti dalla variante 2. Notiamo che la curva rappresentata nella figura 6.9 risulta essere più spezzata ed irregolare rispetto a quella della figura 6.8. Ciò è dovuto al fatto che la probabilità giornaliera di infezione è calcolata sulla base del numero di nuove infezioni registrate in un giorno, il che rende questa misura particolarmente sensibile a piccole variazioni nei dati giornalieri. In particolare, quando il numero di nuovi casi è basso, anche una singola infezione può influenzare significativamente la probabilità giornaliera, generando cambiamenti bruschi nella curva. D'altra parte, la probabilità totale di infezione con la variante 2 tiene conto del numero complessivo di individui da questa variante rispetto al numero totale di infetti. Questo calcolo tende a essere più stabile nel tempo, poiché si basa sull'andamento complessivo delle infezioni, smorzando così le fluttuazioni giornaliere e risultando in una curva più liscia.

6.5 Conclusioni

Le simulazioni condotte utilizzando il metodo di Eulero esplicito e il metodo di Runge-Kutta nel contesto del modello SIR deterministico hanno permesso di trarre diverse conclusioni importanti. In tutte le simulazioni con il metodo di Eulero, condotte con vari valori del tasso di infezione della variante 1 (β_1), si osserva un momento in cui la variante 2 diventa predominante rispetto alla variante 1, indipendentemente dal fatto che R_0^1 sia inferiore o superiore a 1. Con il metodo di Runge-Kutta a 4 stadi, utilizzando un passo temporale maggiore ($dt = 1$), sono stati ottenuti risultati qualitativamente e quantitativamente simili a quelli osservati con il metodo di Eulero con un passo molto più piccolo ($dt = 0.001$). Ciò dimostra che il metodo di Runge-Kutta risulta essere efficace nel ridurre il numero di passi di integrazione necessari, migliorando così l'efficienza computazionale senza intaccare la precisione dei risultati. Quanto presente nei paragrafi 6.2 e 6.3 risponde così al quesito 1 del capitolo 6, mostrando che non è necessario che $R_0^1 < 1$ perchè tutti gli infetti siano contagiati dalla variante dalla variante 2; è sufficiente che R_0^2 sia molto maggiore di R_0^1 .

Le simulazioni condotte utilizzando il metodo di Gillespie per il modello SIR stocastico producono traiettorie diverse ad ogni esecuzione, riflettendo la natura stocastica del processo. I risultati ottenuti mostrano una tendenza generale simile a quella osservata nei modelli deterministici: ancora una volta, la variante 2, con un numero di riproduzione di base maggiore ($R_0^2 \gg R_0^1$), tende a sostituire completamente la variante 1. I grafici riportati nel paragrafo 6.4 evidenziano come, con il trascorrere del tempo, la probabilità che un individuo si infetti con la variante 2 rispetto alla variante 1 aumenti in modo significativo. Questo andamento dimostra chiaramente

che la variante con un numero di riproduzione di base maggiore tende a prevalere, rispondendo in maniera esaustiva e precisa al quesito 2 del capitolo 6, confermando le aspettative teoriche sul comportamento delle due varianti nel modello stocastico.

Appendice A

Codici

A.1 Codice con metodo di Eulero

```
1 import numpy as np
2 import matplotlib.pyplot as plt
3
4 # Definizione dei parametri
5 beta1 = 1/15
6 beta2 = 1/2
7 gamma1 = 1/12
8 gamma2 = 1/10
9 N = 10000 # Popolazione totale
10
11 # Tempo di simulazione
12 T = 120 # giorni
13 dt = 0.1 # passo temporale
14 n_steps = int(T / dt)
15
16 # Inizializzazione delle variabili
17 s = np.zeros(n_steps)
18 i1 = np.zeros(n_steps)
19 i2 = np.zeros(n_steps)
20 r1 = np.zeros(n_steps)
21 r2 = np.zeros(n_steps)
22 t = np.linspace(0, T, n_steps)
23
24 # Condizioni iniziali
25 s[0] = N - 1010
26 i1[0] = 1000
27 i2[0] = 10 # Introdurre una piccola popolazione iniziale infetta da
    variante 2
28 r1[0] = 0
29 r2[0] = 0
30
31 # Metodo di Eulero esplicito per risolvere il sistema continuo
32 for k in range(n_steps - 1):
```

```

33     ds_dt = - (beta1 / N) * s[k] * i1[k] - (beta2 / N) * s[k] * i2[k]
34     ]
35     di1_dt = (beta1 / N) * s[k] * i1[k] - gamma1 * i1[k]
36     di2_dt = (beta2 / N) * s[k] * i2[k] - gamma2 * i2[k]
37     dr1_dt = gamma1 * i1[k]
38     dr2_dt = gamma2 * i2[k]
39
40     s[k + 1] = s[k] + ds_dt * dt
41     i1[k + 1] = i1[k] + di1_dt * dt
42     i2[k + 1] = i2[k] + di2_dt * dt
43     r1[k + 1] = r1[k] + dr1_dt * dt
44     r2[k + 1] = r2[k] + dr2_dt * dt
45
46 # Plot dei risultati
47 plt.figure(figsize=(12, 8))
48 plt.plot(t, s, label='Suscettibili', color='blue')
49 plt.plot(t, i1, label='Infetti variante 1', color='red')
50 plt.plot(t, i2, label='Infetti variante 2', color='green')
51 plt.plot(t, r1, label='Rimossi variante 1', color='purple')
52 plt.plot(t, r2, label='Rimossi variante 2', color='orange')
53
54 plt.xlabel('Tempo (giorni)')
55 plt.ylabel('Numero di individui')
56 plt.title('Modello SIR con Varianti variante 1 e variante 2')
57 plt.legend()
58 plt.grid(True)
59 plt.show()

```

Listing A.1. Codice per la simulazione del modello SIR

A.2 Codice con metodo di Runge-Kutta a 4 stadi

```

1 import numpy as np
2 import matplotlib.pyplot as plt
3
4 # Definizione dei parametri
5 beta1 = 1 / 15
6 beta2 = 1 / 2
7 gamma1 = 1 / 12
8 gamma2 = 1 / 10
9 N = 10000 # Popolazione totale
10
11 # Tempo di simulazione
12 T = 120 # giorni
13 dt = 1.0 # passo temporale
14 n_steps = int(T / dt)
15
16 # Inizializzazione delle variabili
17 s = np.zeros(n_steps)
18 i1 = np.zeros(n_steps)
19 i2 = np.zeros(n_steps)

```

```

20 r1 = np.zeros(n_steps)
21 r2 = np.zeros(n_steps)
22 t = np.linspace(0, T, n_steps)
23
24 # Condizioni iniziali
25 s[0] = N - 1010
26 i1[0] = 1000
27 i2[0] = 10 # Introdurre una piccola popolazione iniziale infetta da
    variante 2
28 r1[0] = 0
29 r2[0] = 0
30
31
32 # Funzione per calcolare i tassi di variazione
33 def derivatives(s, i1, i2, r1, r2):
34     ds_dt = - (beta1 / N) * s * i1 - (beta2 / N) * s * i2
35     di1_dt = (beta1 / N) * s * i1 - gamma1 * i1
36     di2_dt = (beta2 / N) * s * i2 - gamma2 * i2
37     dr1_dt = gamma1 * i1
38     dr2_dt = gamma2 * i2
39     return ds_dt, di1_dt, di2_dt, dr1_dt, dr2_dt
40
41
42 # Metodo di Runge-Kutta a 4 stadi (RK4)
43 for k in range(n_steps - 1):
44     ds_dt1, di1_dt1, di2_dt1, dr1_dt1, dr2_dt1 = derivatives(s[k],
45         i1[k], i2[k], r1[k], r2[k])
46
47     s_temp = s[k] + 0.5 * ds_dt1 * dt
48     i1_temp = i1[k] + 0.5 * di1_dt1 * dt
49     i2_temp = i2[k] + 0.5 * di2_dt1 * dt
50     r1_temp = r1[k] + 0.5 * dr1_dt1 * dt
51     r2_temp = r2[k] + 0.5 * dr2_dt1 * dt
52
53     ds_dt2, di1_dt2, di2_dt2, dr1_dt2, dr2_dt2 = derivatives(s_temp,
54         i1_temp, i2_temp, r1_temp, r2_temp)
55
56     s_temp = s[k] + 0.5 * ds_dt2 * dt
57     i1_temp = i1[k] + 0.5 * di1_dt2 * dt
58     i2_temp = i2[k] + 0.5 * di2_dt2 * dt
59     r1_temp = r1[k] + 0.5 * dr1_dt2 * dt
60     r2_temp = r2[k] + 0.5 * dr2_dt2 * dt
61
62     ds_dt3, di1_dt3, di2_dt3, dr1_dt3, dr2_dt3 = derivatives(s_temp,
63         i1_temp, i2_temp, r1_temp, r2_temp)
64
65     s_temp = s[k] + ds_dt3 * dt
66     i1_temp = i1[k] + di1_dt3 * dt
67     i2_temp = i2[k] + di2_dt3 * dt
68     r1_temp = r1[k] + dr1_dt3 * dt
69     r2_temp = r2[k] + dr2_dt3 * dt

```

```

68     ds_dt4, di1_dt4, di2_dt4, dr1_dt4, dr2_dt4 = derivatives(s_temp,
69         i1_temp, i2_temp, r1_temp, r2_temp)
70     s[k + 1] = s[k] + (dt / 6) * (ds_dt1 + 2 * ds_dt2 + 2 * ds_dt3 +
71         ds_dt4)
72     i1[k + 1] = i1[k] + (dt / 6) * (di1_dt1 + 2 * di1_dt2 + 2 *
73         di1_dt3 + di1_dt4)
74     i2[k + 1] = i2[k] + (dt / 6) * (di2_dt1 + 2 * di2_dt2 + 2 *
75         di2_dt3 + di2_dt4)
76     r1[k + 1] = r1[k] + (dt / 6) * (dr1_dt1 + 2 * dr1_dt2 + 2 *
77         dr1_dt3 + dr1_dt4)
78     r2[k + 1] = r2[k] + (dt / 6) * (dr2_dt1 + 2 * dr2_dt2 + 2 *
79         dr2_dt3 + dr2_dt4)
80
81 # Plot dei risultati
82 plt.figure(figsize=(12, 8))
83 plt.plot(t, s, label='Suscettibili', color='blue')
84 plt.plot(t, i1, label='Infetti variante 1', color='red')
85 plt.plot(t, i2, label='Infetti variante 2', color='green')
86 plt.plot(t, r1, label='Rimossi variante 1', color='purple')
87 plt.plot(t, r2, label='Rimossi variante 2', color='orange')
88
89 plt.xlabel('Tempo (giorni)')
90 plt.ylabel('Numero di individui')
91 plt.title('Modello SIR con Varianti variante 1 e variante 2 - Metodo
92     RK4')
93 plt.legend()
94 plt.grid(True)
95 plt.show()

```

A.3 Codice con metodo di Gillespie

```

1     import numpy as np
2     import matplotlib.pyplot as plt
3
4     # Definizione dei parametri
5     beta1 = 1/15
6     beta2 = 1/2
7     gamma1 = 1/12
8     gamma2 = 1/10
9     N = 10000 # Popolazione totale
10
11     # Condizioni iniziali
12     S = N - 1010
13     I1 = 1000
14     I2 = 10
15     R1 = 0
16     R2 = 0
17
18     # Tempo di simulazione
19     T_max = 40

```



```
20 t = 0
21
22 # Inizializzazione delle liste per tracciare i risultati
23 time_points = [t]
24 S_values = [S]
25 I1_values = [I1]
26 I2_values = [I2]
27 R1_values = [R1]
28 R2_values = [R2]
29
30 # Inizializzazione delle liste per tracciare le nuove infezioni
    giornaliere
31 time_points_daily = np.arange(0, T_max+1)
32 h = len(time_points_daily)
33 new_1_infections = np.zeros(h)
34 new_infections_2 = np.zeros(h)
35 new_removals_1 = np.zeros(h)
36 new_removals_2 = np.zeros(h)
37
38 while t < T_max-1 and (I1 > 0 or I2 > 0):
39     # Calcolo dei tassi di evento
40     rate_infection1 = beta1 * S * I1 / N
41     rate_infection2 = beta2 * S * I2 / N
42     rate_recovery1 = gamma1 * I1
43     rate_recovery2 = gamma2 * I2
44     total_rate = rate_infection1 + rate_infection2 + rate_recovery1
        + rate_recovery2
45
46     if total_rate == 0:
47         break
48
49     # Determinazione del tempo fino al prossimo evento
50     delta_t = np.random.exponential(1 / total_rate)
51     t += delta_t
52     day = int(np.floor(t))
53
54     # Determinazione del tipo di evento
55     r = np.random.uniform(0, total_rate)
56     if r < rate_infection1: # infezione con variante 1
57         S -= 1
58         I1 += 1
59         new_infections_1[day] += 1
60
61     elif r < rate_infection1 + rate_infection2: # infezione con
        variante 2
62         S -= 1
63         I2 += 1
64         new_infections_2[day] += 1
65
66     elif r < rate_infection1 + rate_infection2 + rate_recovery1: #
        rimozione da variante 1
67         I1 -= 1
68         R1 += 1
```

```
69         new_removals_1[day] += 1
70
71     else: # rimozione da variante 2
72         I2 -= 1
73         R2 += 1
74         new_removals_2[day] += 1
75
76
77     # Registrazione dei risultati
78     time_points.append(t)
79     S_values.append(S)
80     I1_values.append(I1)
81     I2_values.append(I2)
82     R1_values.append(R1)
83     R2_values.append(R2)
84
85
86 # operazioni componente per componente
87 new_infections_1_effective = new_infections_1 - new_removals_1
88 new_infections_2_effective = new_infections_2 - new_removals_2
89
90
91
92
93 # Plot dei risultati
94 plt.figure(figsize=(12, 8))
95 plt.plot(time_points, S_values, label='Suscettibili', color='blue')
96 plt.plot(time_points, I1_values, label='Infetti variante 1', color='red')
97 plt.plot(time_points, I2_values, label='Infetti variante 2', color='green')
98 plt.plot(time_points, R1_values, label='Rimossi variante 1', color='purple')
99 plt.plot(time_points, R2_values, label='Rimossi variante 1', color='yellow')
100
101 plt.xlabel('Tempo (giorni)')
102 plt.ylabel('Numero di individui')
103 plt.title('Algoritmo di Gillespie per Modello SIR con Varianti
104           variante 1 e variante 2')
104 plt.legend()
105 plt.grid(True)
106 plt.show()
107
108 # Calcolo della probabilit di infezione con variante 2 rispetto a
109     variante 1
109 P_2 = np.array(I2_values) / (np.array(I1_values) + np.array(
110     I2_values))
110
111 plt.figure(figsize=(12, 8))
112 plt.plot(time_points, P_2, label='Probabilit di infezione con
113     variante 2', color='green')
113 plt.xlabel('Tempo (giorni)')
```

```
114 plt.ylabel('Probabilit ')
115 plt.title('Probabilit di infezione con variante 2 nel tempo')
116 plt.legend()
117 plt.grid(True)
118 plt.show()
119
120 # calcolo della probabilit di infezione con variante 2 rispetto a
    variante 1 AL GIORNO
121 P_daily_2 = np.where(new_infections_1 + new_infections_2 > 0,
122                      new_infections_2 / (new_infections_1 +
123                                         new_infections_2),
124                      0)
125
126 plt.figure(figsize=(12, 8))
127 plt.plot(time_points_daily, P_daily_2, label='Probabilit di
    infezione con variante 2', color='green')
128 plt.xlabel('Tempo (giorni)')
129 plt.ylabel('Probabilit ')
130 plt.title('Probabilit giornaliera di infezione con variante 2
    rispetto a variante 1 al giorno')
131 plt.grid(True)
132 plt.show()
```


Bibliografia

- L. J. S. Allen. An introduction to stochastic epidemic models. In F. Brauer, P. Van Den Driessche, and J. Wu, editors, *Mathematical Epidemiology*, volume 1945 of *Lecture Notes in Mathematics*, pages 81–130. Springer, Berlin, Heidelberg, 2008.
- H. T. Banks, A. Broido, B. Canter, K. Gayvert, S. Hu, M. Joyner, and K. Link. Simulation algorithms for continuous time markov chain models. In *Simulation and Modeling Related to Computational Science and Robotics Technology*, volume 37 of *Studies in Applied Electromagnetics and Mechanics*, pages 3–18, 2012.
- R. L. Burden and J. D. Faires. *Numerical Analysis*. Cengage Learning, Boston, MA, 9th edition, 2011.
- R. Durrett. *Essentials of Stochastic Processes*. 3rd edition, 2016.
- S. M. Ross. *Introduction to Probability Models*. Academic Press, 12th edition, 2019.