

UNIVERSITÀ DEGLI STUDI DI PADOVA

Dipartimento di Fisica e Astronomia “Galileo Galilei”

Corso di Laurea in Fisica

Tesi di Laurea

**Modelli compartimentali per lo studio della decelerazione
della diffusione del COVID-19 dopo un lockdown**

Relatore
Prof. Marco Baiesi

Laureando
Emanuele Giordano

Anno Accademico 2020/2021

Indice

1	Modelli per lo studio della decelerazione	4
1.1	Modello SIR	5
1.2	Modello $SHIR$	5
1.2.1	Approssimazione $\beta I \ll \sigma$	6
1.3	Modello $SPIR$	7
1.3.1	Approssimazione $I \ll 1$	7
1.4	Confronto dei modelli $SHIR$ e $SPIR$ tramite interpolazione dei dati sperimentali	8
1.5	Modello $SS'HIR$	11
1.5.1	Approssimazione $I \ll 1$	12
1.6	Confronto tra i modelli $SPIR$ e $SS'HIR$ tramite interpolazione dei dati sperimentali	12
2	Modelli per lo studio del periodo di latenza	15
2.1	Modello $SEIR$	15
2.1.1	Soluzione approssimata e numero di riproduzione R_0	16
2.2	Modello SE_1E_2IR	17
2.2.1	Tempo di latenza: stati E_1 ed E_2 accoppiati	18
2.2.2	Tempo di latenza: stati E_1 ed E_2 disaccoppiati	19
2.3	Distribuzioni Gamma ed Erlang	19
2.4	Modello $SE_1 \dots E_n IR$	20
2.5	Conclusioni	21

Sommario

In questa tesi si vogliono esaminare diversi modelli compartimentali di diffusione epidemica, con particolare riguardo alla propagazione del COVID-19 e a come la sua diffusione venga modificata dalla messa in atto di misure socio-sanitarie precauzionali quali distanziamento sociale e lockdown. La presenza di tali misure comporterà una modifica al noto modello SIR, con l'introduzione di stati nascosti H o protetti P . Una corretta riproduzione dell'andamento temporale dell'infezione richiederà inoltre l'introduzione di un nuovo compartimento di persone esposte E ed in particolare di stati intermedi di Erlang.

Capitolo 1

Modelli per lo studio della decelerazione

In questa prima parte si vogliono introdurre e studiare alcuni modelli di diffusione epidemica con lo scopo di spiegare la dinamica dell'evoluzione del COVID-19 nelle sue prime fasi e la diminuzione della sua accelerazione a seguito di misure socio-sanitarie quali l'obbligo dell'utilizzo della mascherina, il distanziamento sociale e il lockdown.

Al fine di poter confrontare le simulazioni teoriche con i dati sperimentali, è possibile definire il numero cumulativo di casi $f = I + R$, la velocità (logaritmica) di variazione dei casi $v = \ln f$ e la relativa accelerazione $a = \dot{v}$, dove, per semplicità di notazione, si è omissso di indicare esplicitamente la dipendenza temporale e si è indicata con il puntino ' la derivata temporale; con I si è quindi indicato il compartimento delle persone infette e con R quello dei rimossi, all'interno del quale non è fatta alcuna distinzione tra persone guarite o decedute, trascurando dunque la possibilità di reinfezione.

In tutto il seguito si farà l'ipotesi di modello chiuso, ovvero si assume che la popolazione totale sia costante e si interpretano i vari compartimenti come frazione di persone in tale stato sul totale della popolazione.

1.1 Modello SIR

Il più semplice modello di diffusione epidemica è il famoso modello SIR , nel quale ci sono soltanto 3 compartimenti: quello S delle persone sane ma suscettibili di infezione, quello I delle persone infette e quello R dei rimossi.

La transizione dallo stato S a quello I è parametrizzata dal *contact rate* β , mentre quella da I a R tramite la *recovery rate* μ , che studi precedenti (si veda ad esempio [1]) mostrano essere dell'ordine di $\frac{1}{14days}$. La dinamica del modello SIR può essere efficacemente descritta da uno schema, come in figura 1.1 e le equazioni che governano la dinamica sono dunque:

$$\begin{cases} \dot{S} = -\beta IS \\ \dot{I} = \beta IS - \mu I \\ \dot{R} = \mu I \end{cases} \quad (1.1)$$

con $S + I + R = 1 \forall t$ nell'ipotesi di modello chiuso.

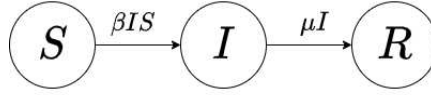


Figura 1.1: Schema del modello SIR .

Nel modello SIR è possibile calcolare le grandezze dinamiche (e in particolare l'accelerazione) ottenendo:

$$v = \ln(\dot{I} + \dot{R}) = \ln(\beta IS) \quad \Rightarrow \quad a = \frac{\dot{I}}{I} + \frac{\dot{S}}{S} = \beta(S - I) - \mu \quad (1.2)$$

Fintanto che l'infezione è contenuta, ovvero $I \ll S$ e $S \approx 1$ si può osservare come un parametro importante sia il numero di riproduzione $R_0 = \frac{\beta}{\mu}$: se $R_0 < 1$, allora l'accelerazione è negativa e l'infezione tende a svanire, mentre se $R_0 > 1$ l'accelerazione è positiva e l'infezione tende a crescere esponenzialmente; per una trattazione più completa in merito al numero di riproduzione si veda ad esempio [2].

Ovviamente questo semplice modello non incorpora nessun meccanismo di protezione degli individui ed è dunque inadatto a studiare il comportamento della diffusione del virus a seguito di un lockdown.

1.2 Modello $SHIR$

Una prima idea per modellizzare l'effetto del lockdown sulla curva pandemica è quello di aggiungere un'ulteriore stato H (accessibile ai suscettibili) di individui nascosti (in inglese *hidden*) che non possono essere contagiati. Il compartimento H è formato da quelle persone che, seppur suscettibili, già dall'inizio dell'epidemia si sono rese conto del potenziale rischio del virus e hanno di conseguenza assunto misure precauzionali forti.

Il conseguente modello prende dunque il nome *SHIR* e viene introdotto e studiato da T. Barnes in [3]. Lo schema della dinamica *SHIR* è presentato in 1.2 e le equazioni che la governano sono:

$$\begin{cases} \dot{S} = -\beta IS - \sigma S \\ \dot{H} = \sigma S \\ \dot{I} = \beta IS - \mu I \\ \dot{R} = \mu I \end{cases} \quad (1.3)$$

dove σ è denominata *hiding rate*.

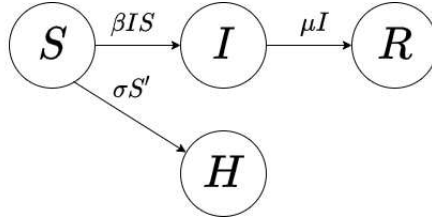


Figura 1.2: Schema del modello *SHIR*.

In questo modello si può calcolare la velocità:

$$\dot{f} = \dot{I} + \dot{R} = \beta IS \quad \Rightarrow \quad v = \ln \beta + \ln I + \ln S \quad (1.4)$$

e da questa l'accelerazione:

$$a = \frac{\dot{I}}{I} + \frac{\dot{S}}{S} = \beta S - \mu - \sigma - \beta I \quad (1.5)$$

1.2.1 Approssimazione $\beta I \ll \sigma$

Le equazioni del modello *SHIR* non sono risolvibili analiticamente, dunque per trovare come S e I variano nel tempo è necessario procedere con delle approssimazioni. Un'ipotesi che è certamente valida nelle prime fasi dell'epidemia è $\beta I \ll \sigma$, dato che nei casi di infezione contenuta si ha $I \ll 1$.

In questo caso la prima equazione del modello si semplifica diventando $\dot{S} = -\sigma S$ da cui si ha immediatamente $S(t) = S_0 e^{-\sigma t}$ dove $S_0 = S(t=0)$. Utilizzando nuovamente l'ipotesi di infezione contenuta per trascurare il termine βI in (1.5), si ottiene una formula chiusa per l'accelerazione:

$$a(t) = \beta S_0 e^{-\sigma t} - \mu - \sigma \quad (1.6)$$

A partire da questa formula sarà poi possibile eseguire un'interpolazione dei dati sperimentali. Tuttavia, prima di procedere con l'analisi dei dati, conviene proporre un ulteriore modello con l'intenzione di introdurre una dinamica che meglio descrive quella reale, nel contesto di un virus (il COVID-19) dal quale è difficile pensare di poter essere completamente al sicuro.

1.3 Modello *SPIR*

Il modello *SHIR* parte dall'assunzione piuttosto forte che, tramite opportune misure precauzionali, sia possibile annullare completamente la possibilità di essere infettati. Una situazione più realistica si ottiene sostituendo lo stato *H* del modello *SHIR* con uno stato *P* di individui protetti che hanno una probabilità di essere contagiati inferiore agli altri suscettibili, ovvero una minore *contact rate*.

Il modello può dunque essere chiamato *SPIR*; il suo schema è mostrato in 1.3 e le sue equazioni sono:

$$\begin{cases} \dot{S} = -\beta_s IS - \varphi S \\ \dot{P} = \varphi S - IP\beta_p \\ \dot{I} = \beta_s IS + \beta_p IP - \mu I \\ \dot{R} = \mu I \end{cases} \quad (1.7)$$

dove $\beta_p < \beta_s$ per ipotesi e φ è la *protection rate*.

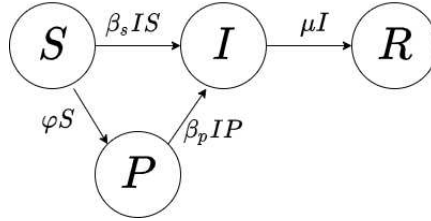


Figura 1.3: Schema del modello *SPIR*.

In questo modello si può calcolare la variazione del numero cumulativo di casi:

$$\dot{f} = \dot{I} + \dot{R} = \beta_s IS + \beta_p IP = \beta_s IS \left(1 + \frac{\beta_p P}{\beta_s S}\right) \quad (1.8)$$

e dunque la velocità:

$$v = \ln \beta_s + \ln I + \ln S + \ln \left(1 + \frac{\beta_p P}{\beta_s S}\right) \quad (1.9)$$

e da questa l'accelerazione:

$$a = \frac{\dot{I}}{I} + \frac{\dot{S}}{S} + \frac{\dot{P}S - P\dot{S}}{S^2} \cdot \frac{1}{\frac{\beta_s}{\beta_p} + \frac{P}{S}} \quad (1.10)$$

1.3.1 Approssimazione $I \ll 1$

Nuovamente le equazioni del modello non sono risolvibili analiticamente, dunque per procedere è necessario agire con delle approssimazioni. Nelle prime fasi dell'epidemia possiamo riprendere l'ipotesi di infezione contenuta $I \ll 1$, che implica tra l'altro che $\beta_i I \ll \varphi$ per $i = s, p$. Il modello in questa approssimazione diventa:

$$\begin{cases} \dot{S} \approx -\varphi S \\ \dot{P} \approx \varphi S \\ \dot{I} = \beta_s IS + \beta_p IP - \mu I \\ \dot{R} = \mu I \end{cases} \quad (1.11)$$

La prima equazione fornisce subito:

$$S(t) = S_0 e^{-\varphi t} \quad (1.12)$$

che sostituita nella seconda restituisce:

$$P(t) = S_0(1 - e^{-\varphi t}) + P_0 \approx 1 - S_0 e^{-\varphi t} \quad (1.13)$$

dove l'ultima approssimazione equivale a $P_0 \approx 1 - S_0$, valida dal momento che la percentuale di infetti I è trascurabile rispetto a S e a P nelle prime fasi dell'epidemia. Con queste approssimazioni si possono calcolare i vari contributi all'accelerazione:

- $\frac{\dot{I}}{I} = \beta_s S + \beta_p P - \mu = \beta_p + S_0 e^{-\varphi t}(\beta_s - \beta_p)$
- $\frac{\dot{S}}{S} = -\varphi$
- $\frac{\dot{P}S - P\dot{S}}{S^2} \cdot \frac{1}{\frac{\beta_s}{\beta_p} + \frac{P}{S}} = \frac{\varphi\beta_p}{\beta_p + S_0 e^{-\varphi t}(\beta_s - \beta_p)}$

Allora conviene definire la quantità:

$$\alpha(t) := \beta_p + S_0 e^{-\varphi t}(\beta_s - \beta_p) \quad (1.14)$$

e in questo modo si può scrivere l'accelerazione come:

$$a(t) = \alpha(t) - \mu - \varphi + \frac{\varphi\beta_p}{\alpha(t)} \quad (1.15)$$

A partire da quest'ultima formula è possibile eseguire un'interpolazione dei dati sperimentali e in particolare valutare se il modello *SPIR* apporti un miglioramento significativo rispetto al modello *SHIR*.

1.4 Confronto dei modelli *SHIR* e *SPIR* tramite interpolazione dei dati sperimentali

Il numero cumulativo di individui infetti in ogni nazione è pubblico e disponibile, ad esempio, in [4]. Dato che si vuole studiare la variazione dell'andamento temporale dell'infezione a seguito di un lockdown, sono stati selezionati solamente i dati che si riferiscono alla prima ondata di COVID-19 nel periodo a cavallo tra 2019 e 2020, evitando di considerare *outbreaks* successivi dovuti a fattori che non vengono descritti dai modelli studiati (ad esempio la formazione di nuove varianti virali).

Inoltre, invece di fittare i dati grezzi, si è deciso di usare una media mobile dei nuovi casi su un periodo di due settimane al fine di evitare grandi fluttuazioni statistiche dovute a fattori secondari non necessariamente legati alla diffusione del virus (un esempio tra tutti è il cosiddetto *effetto weekend* per il quale nel fine

settimana vengono eseguiti mediamente meno tamponi). Esplicitamente, denotando con $f(t)$ il numero cumulativo di casi al tempo t , si è calcolato:

$$\dot{f}(t) = \frac{f(t) - f(t - L)}{L} \quad \text{dove si è posto } L = 14 \text{ days} \quad (1.16)$$

Infine si sono selezionati i dati di 6 nazioni (Italia, Germania, Regno Unito, Russia, India e Brasile) in modo da poter confrontare i comportamenti dei due modelli in varie situazioni differenti. Queste decisioni sono state prese in conformità a quanto fatto da G. Barzon, K.K.H. Manjunatha, W. Rugel, E. Orlandini e M. Baiesi in [5] nel contesto di un progetto del corso magistrale di *Physics of Data* dell'Università di Padova.

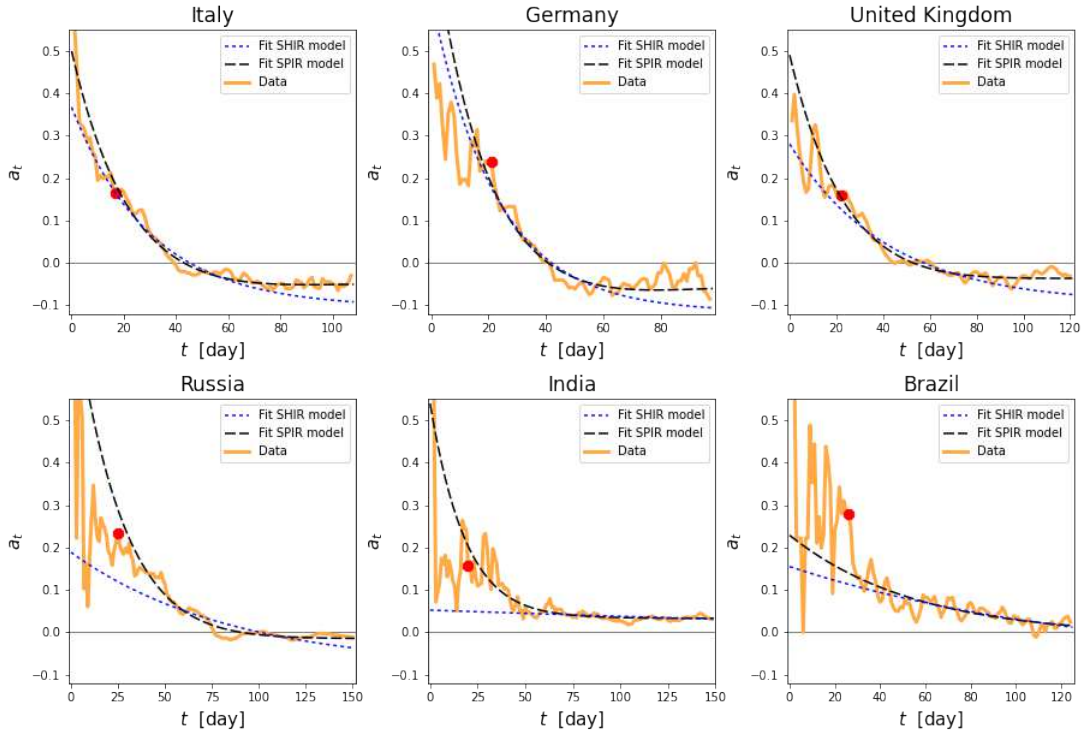


Figura 1.4: Confronto delle interpolazioni dei dati sperimentali tra i modelli *SHIR* e *SPIR*. I grafici riportano l'accelerazione del numero cumulativo di casi confermati di COVID-19 sulle ordinate e i giorni sulle ascisse; il punto rosso indica l'inizio ufficiale del lockdown nazionale.

L'accelerazione dell'infezione è calcolata usando (1.6) per il modello *SHIR* e (1.15) per quello *SPIR*. Si nota come il modello *SHIR* non riesca a riprodurre fedelmente i dati: nei casi di Italia, Germania e Regno Unito il fit si discosta notevolmente dai dati per tempi lunghi, mentre questo avviene per tempi brevi nei casi di Russia, India e Brasile. Contrariamente, il modello *SPIR* riesce a riprodurre meglio l'andamento per tempi sia brevi che lunghi.

Tabella 1.1: Risultati dell'interpolazione dell'accelerazione del numero cumulativo di infetti nel tempo tramite le formule dei modelli *SPIR* e *SHIR*.

	Modello <i>SPIR</i>				Modello <i>SHIR</i>		
	β_s	β_p	φ	S_0	β	σ	S_0
Italia	0.81	0.024	0.046	0.755	0.53	0.033	0.889
Germania	1.06	0.018	0.054	0.868	0.92	0.044	0.806
Regno Unito	0.70	0.035	0.042	0.851	0.62	0.024	0.603
Russia	0.93	0.056	0.038	1.000	1.45	0.012	0.187
India	0.66	0.104	0.050	0.992	0.28	0.001	0.445
Brasile	0.47	0.063	0.017	0.611	0.41	0.008	0.565

Si potrebbe pensare che il miglior comportamento del modello *SPIR* sia dovuto semplicemente al fatto di avere un parametro libero in più. In realtà, osservando la definizione di $\alpha(t)$ in 1.14 e notando che nelle interpolazioni si ottiene sempre $\beta_s \gg \beta_p$, ci si accorge che l'accelerazione non dipende tanto da β_s e S_0 separatamente quanto dal loro prodotto.

Dal momento che nelle interpolazioni si ottiene sempre $S_0 \approx 1$, si potrebbe pensare di ripetere il fit per il modello *SPIR* fissando S_0 a 1. Per quanto appena visto, il valore di β_s dovrebbe cambiare, assumendo il significato di *contact rate* efficace e trasformandosi come $\beta_s^{eff} \approx S_0 \beta_s$.

Tabella 1.2: Risultati dell'interpolazione dell'accelerazione del numero cumulativo di infetti nel tempo tramite la formula del modello *SPIR* in cui si è fissato $S_0 = 1$ e confronto del parametro β_s^{eff} di tale modello con il prodotto $S_0 \beta_s$ del modello *SPIR* standard. Si osservi anche come i valori ottenuti per i parametri β_p e φ coincidano con quelli ottenuti precedentemente fino alle cifre significative considerate.

	β_s^{eff}	β_p	φ	$S_0 \beta_s$
Italia	0.62	0.024	0.046	0.61
Germania	0.92	0.018	0.054	0.92
Regno Unito	0.60	0.035	0.042	0.60
Russia	0.93	0.056	0.038	0.93
India	0.65	0.104	0.050	0.65
Brasile	0.31	0.063	0.017	0.29

I risultati confermano la validità di queste considerazioni, come si può osservare sia dai nuovi valori di β_s^{eff} , molto vicini al prodotto $S_0 \beta_s$ per tutte le nazioni considerate, sia dai grafici comparativi in figura 1.5, in cui si nota come imporre $S_0 = 1$ non abbia modificato sensibilmente le interpolazioni.

Si conclude che, anche fissando $S_0 = 1$, ovvero utilizzando un modello *SPIR* con soli 3 parametri liberi, esso riesce ad offrire una descrizione dell'evoluzione temporale della diffusione del COVID-19 più accurata rispetto a quella fornita dal modello *SHIR*.

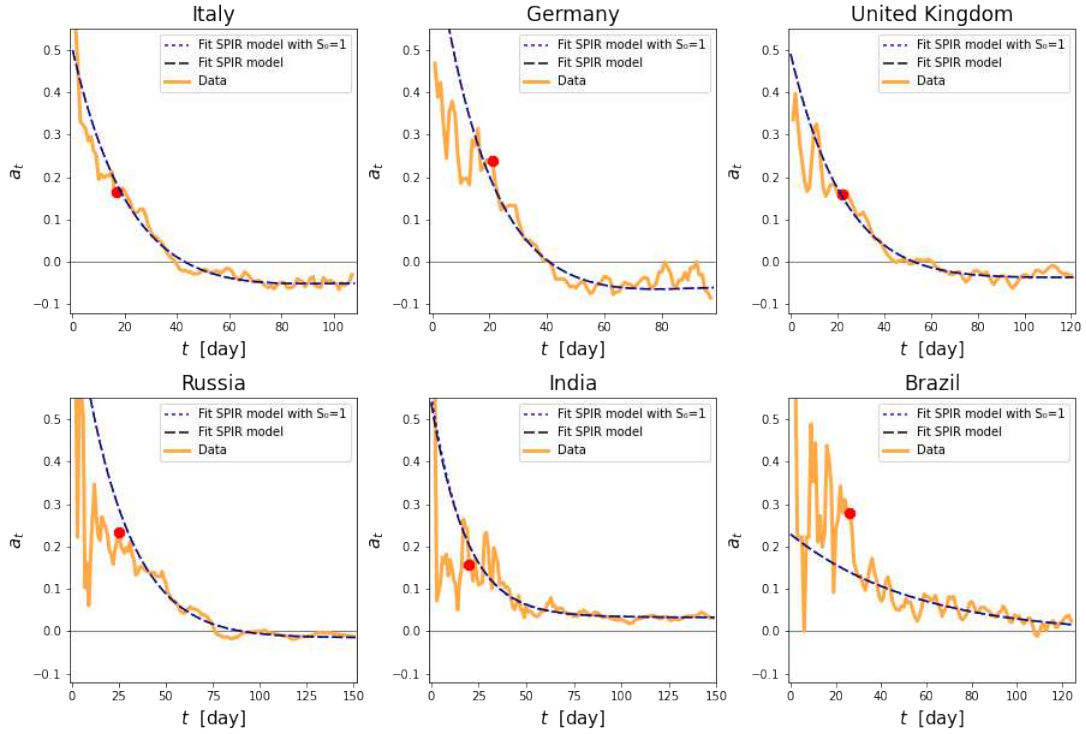


Figura 1.5: Confronto delle interpolazioni tra i modelli *SPIR* standard e *SPIR* con $S_0 = 1$ (dove non sono visibili due curve, i fit risultano sovrapposti). I grafici riportano l'accelerazione del numero cumulativo di casi confermati di COVID-19 sulle ordinate e i giorni sulle ascisse; il punto rosso indica l'inizio ufficiale del lockdown nazionale.

1.5 Modello $SS'HIR$

Oltre alla forte assunzione di potere eliminare completamente il rischio di infezione, il modello *SHIR* presenta un ulteriore problema: il rapido svuotamento del compartimento delle persone suscettibili di infezione. Dato che nelle prime fasi dell'epidemia si ha che $\beta I \ll \sigma$, il compartimento H si riempie molto più velocemente di quello I degli infetti. Il rapido svuotamento di S verso H comporta dunque che in breve tempo la maggior parte della popolazione sarà bloccata nello stato H , invece che transire attraverso I e infine verso R . Questa problematica si riscontra ad esempio in alcuni grafici di figura 1.4 in cui il modello *SHIR* prevede un'accelerazione sensibilmente inferiore ai dati sperimentali per tempi lunghi.

Un modo per risolvere questo problema è suddividere l'insieme dei suscettibili in due compartimenti: uno stato di individui suscettibili S' che hanno la possibilità di accedere ad uno stato nascosto H e uno di suscettibili S che invece non hanno questa possibilità (questo è il caso, ad esempio, di coloro che sono costretti a mantenere un rapporto fisico ravvicinato con altre persone per motivi di lavoro).

Il conseguente modello prende allora il nome $SS'HIR$ ed è stato introdotto per la prima volta in [5]; la dinamica è rappresentata in 1.6 ed è governata dalle equazioni:

$$\begin{cases} \dot{S} = -\beta IS \\ \dot{S}' = -\beta IS' - \sigma S' \\ \dot{H} = \sigma S' \\ \dot{I} = \beta I(S + S') - \mu I \\ \dot{R} = \mu I \end{cases} \quad (1.17)$$

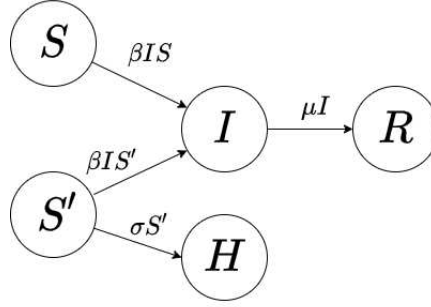


Figura 1.6: Schema del modello $SS'HIR$.

La velocità in questo modello è data da:

$$\dot{f} = \dot{I} + \dot{R} = \beta I(S + S') \Rightarrow v = \ln \beta + \ln I + \ln(S + S') \quad (1.18)$$

e dunque l'accelerazione risulta essere:

$$a = \frac{\dot{I}}{I} + \frac{\dot{S} + \dot{S}'}{S + S'} = \beta(S + S') - \mu - \beta I - \frac{\sigma S'}{S + S'} \quad (1.19)$$

1.5.1 Approssimazione $I \ll 1$

Anche in questo caso è necessario proseguire con l'ipotesi di infezione contenuta $I \ll 1$ tramite la quale si ottiene subito che $S'(t) = S'_0 e^{-\sigma t}$ e $S(t) = \text{cost.}$

Convien definire (seguendo tra l'altro le convenzioni usate in [5]) il parametro $p := S'_0$ e dunque, dato che nelle prime fasi dell'epidemia possiamo assumere come trascurabili le frazioni di individui infetti I e nascosti H , si ha anche che $S_0 = 1 - p$. In questo modo si ottiene la formula per l'accelerazione in funzione del tempo:

$$a(t) = \beta(1 - p + pe^{-\sigma t}) - \mu - \frac{\sigma pe^{-\sigma t}}{1 - p + pe^{-\sigma t}} \quad (1.20)$$

1.6 Confronto tra i modelli $SPIR$ e $SS'HIR$ tramite interpolazione dei dati sperimentali

L'analisi dei dati ripropone quanto fatto precedentemente in 1.4, in particolare vengono utilizzati gli stessi dati con le stesse convenzioni. In questo caso le accelerazioni teoriche sono calcolate mediante le formule (1.15) per il modello $SPIR$, in cui viene fissato $S_0 = 1$ e (1.20) per il modello $SS'HIR$.

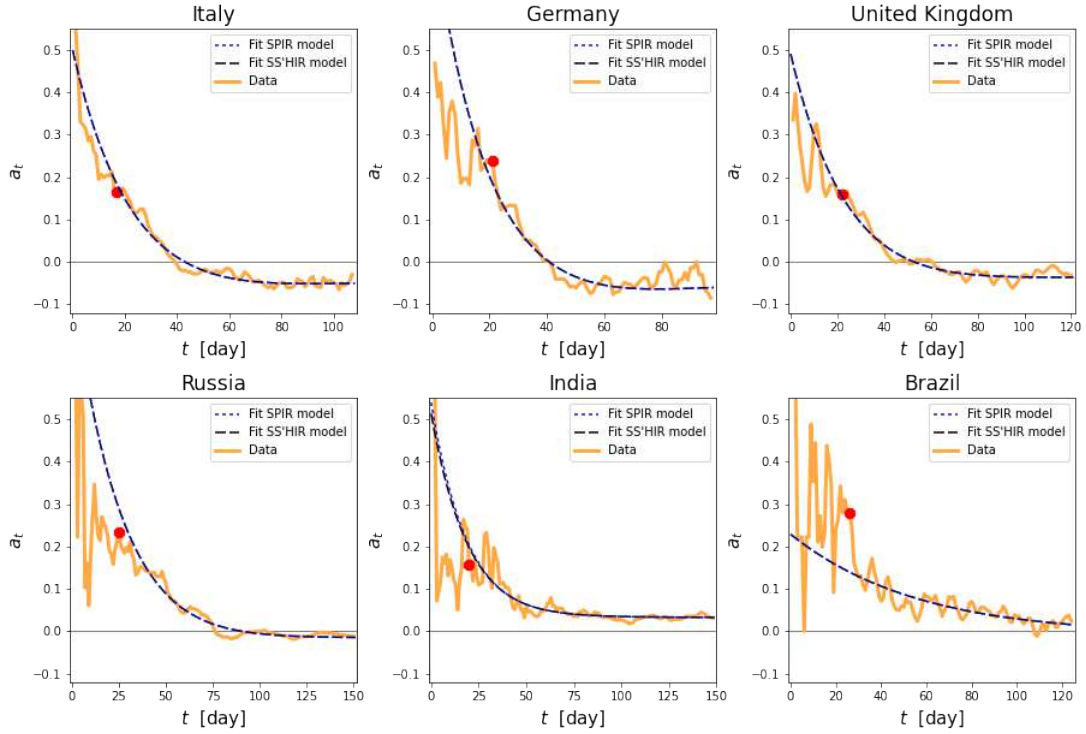


Figura 1.7: Confronto delle interpolazioni dei dati sperimentali tra i modelli $SS'HIR$ e $SPIR$ con $S_0 = 1$ (dove non sono visibili due curve, i fit risultano sovrapposti). I grafici riportano l'accelerazione del numero cumulativo di casi confermati di COVID-19 sulle ordinate e i giorni sulle ascisse; il punto rosso indica l'inizio ufficiale del lockdown nazionale.

I grafici in figura 1.7 mostrano una sovrapposizione quasi perfetta delle interpolazioni dei due modelli (una lievissima differenza si nota nel caso dell'India, che tuttavia si poteva già notare in figura 1.5 ed è dunque imputabile alla successiva assunzione che $S_0 = 1$). Confrontando i risultati nelle tabelle 1.1 e 1.3 si osserva inoltre che il valore del parametro φ nel modello $SPIR$ è sempre uguale al valore del corrispondente parametro σ nel modello $SS'HIR$.

Tabella 1.3: Risultati dell'interpolazione dell'accelerazione del numero cumulativo di infetti nel tempo tramite il modello $SS'HIR$.

	β	σ	p
Italia	0.61	0.046	0.961
Germania	0.92	0.054	0.981
Regno Unito	0.60	0.042	0.941
Russia	0.93	0.038	0.939
India	0.63	0.049	0.833
Brasile	0.31	0.017	0.798

Questa particolarità sembra suggerire che i due modelli possano essere equivalenti nella descrizione dell'accelerazione dei nuovi casi di infezione. Si cerca dunque di

trovare delle sostituzioni che possano uguagliare le due formule delle accelerazioni (1.15) e (1.20):

$$\begin{aligned} a_{SS'HIR}(t) &= \beta(1 - p + pe^{-\sigma t}) - \mu - \frac{\sigma pe^{-\sigma t}}{1 - p + pe^{-\sigma t}} \stackrel{!}{=} \\ &\stackrel{!}{=} a_{SPIR}(t) = \beta_p + S_0 e^{-\varphi t} (\beta_s - \beta_p) - \mu - \varphi + \frac{\varphi \beta_p}{\beta_p + S_0 e^{-\varphi t} (\beta_s - \beta_p)} \end{aligned} \quad (1.21)$$

Notando che il *recovery rate* μ è lo stesso nei due casi e che si deve necessariamente avere $\varphi = \sigma$ per avere uguaglianza $\forall t$, si trova la sostituzione $S_0(\beta_s - \beta_p) = \beta_p$, da cui si ottiene:

$$\begin{aligned} a_{SPIR}(t) &= \beta_p + \beta pe^{-\sigma t} - \mu + \frac{\sigma \beta_p - \sigma(\beta_p + \beta pe^{-\sigma t})}{\beta_p + \beta pe^{-\sigma t}} = \\ &= \beta \left(\frac{\beta_p}{\beta} + pe^{-\sigma t} \right) - \mu - \frac{\sigma pe^{-\sigma t}}{\frac{\beta_p}{\beta} + pe^{-\sigma t}} \end{aligned} \quad (1.22)$$

E a questo punto è chiaro come l'ultima sostituzione da fare sia $\frac{\beta_p}{\beta} = 1 - p$. Ricapitolando e riarrangiando le varie sostituzioni, il passaggio dalla formula per l'accelerazione in *SPIR* a quella in *SS'HIR* è dato da:

$$\begin{cases} \sigma = \varphi \\ \beta = S_0 \beta_s + (1 - S_0) \beta_p \\ p = \frac{S_0(\beta_s - \beta_p)}{S_0(\beta_s - \beta_p) + \beta_p} \end{cases} \quad (1.23)$$

In particolare si nota che questo sistema di equazioni trasforma 4 variabili del modello *SPIR* nelle 3 del modello *SS'HIR*. Questo conferma l'idea che era stata avanzata in 1.4 secondo la quale i parametri davvero significativi per descrivere la dinamica *SPIR* fossero 3 soltanto. In tale sezione si era proposto di sostituire i due parametri S_0 e β_s con il loro prodotto $S_0 \beta_s$; tuttavia dal confronto con la dinamica *SS'HIR* si è riusciti a comprendere che il parametro davvero significativo è $\beta = S_0 \beta_s + (1 - S_0) \beta_p$, che ad ogni modo è molto ben approssimato dal semplice prodotto $S_0 \beta_s$ nel caso in cui si abbia $\beta_p \ll \beta_s$ e $S_0 \approx 1$, che è esattamente la situazione che si aveva in 1.4.

Questa è una situazione molto interessante in cui due modelli dal significato fisico molto diverso risultano essere matematicamente equivalenti, almeno nell'approssimazione di infezione contenuta $I \ll 1$ che è stata necessaria per risolvere le equazioni differenziali che descrivono la loro dinamica. Di fatto in queste ipotesi il compartimento protetto P di *SPIR* si svuota molto più lentamente di quanto si riempie e questo costringe una frazione sempre maggiore di individui a rimanere 'bloccati' in tale compartimento, in analogia a quanto accade nello stato nascosto H in *SS'HIR*. Inoltre la relazione $\beta = S_0 \beta_s + (1 - S_0) \beta_p$ mostra come il flusso di individui dall'intero compartimento dei suscettibili $S + S'$ verso gli infetti in *SS'HIR* sia riprodotto dalla somma dei due flussi da S verso I e da P verso I nel modello *SPIR*.

Capitolo 2

Modelli per lo studio del periodo di latenza

In tutto il primo capitolo si sono studiati modelli in cui non c'è alcun ritardo tra il momento del contagio e il periodo di contagiosità: quando un individuo suscettibile S viene esposto al virus, egli, secondo i modelli finora proposti, diventerà fin da subito infetto e contagioso e potrà immediatamente trasmettere il virus ad altri suscettibili.

Nella realtà, tuttavia, è noto che la dinamica è sostanzialmente differente: quando una persona sana entra in contatto il virus, questo inizia a replicarsi all'interno del suo organismo e la persona diventa contagiosa solamente dopo un tempo caratteristico detto periodo di latenza (*latent period* o *pre-infectious period* in inglese), da non confondersi con il periodo di incubazione, che è invece il tempo che intercorre tra l'esposizione al virus e la comparsa dei primi sintomi. La situazione è schematizzata in figura 2.1.

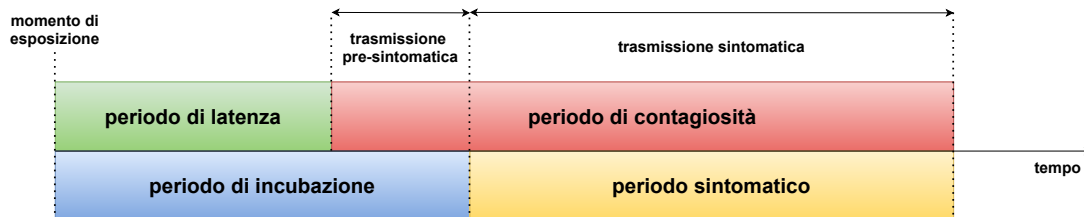


Figura 2.1: Diagramma esplicativo dei vari periodi di interesse di un'infezione: un individuo può trasmettere il virus soltanto se contagioso, ovvero successivamente al periodo di latenza. Il contagio può avvenire anche tramite una persona che non ha ancora manifestato sintomi se il periodo di incubazione è più lungo del periodo di latenza (come avviene, ad esempio, nel caso del COVID-19) e in questo caso si parla di trasmissione pre-sintomatica.

2.1 Modello $SEIR$

Il più semplice modello che implementa un compartimento E di persone esposte è il modello $SEIR$ la cui dinamica è schematizzata in 2.2 e si può riassumere nel seguente modo: quando un infetto I entra in contatto con un suscettibile S , quest'ultimo può diventare esposto E ; in questo stato non c'è contagiosità, ovvero non c'è la possibilità di trasmettere il virus ad altri suscettibili, tuttavia si è destinati

a diventare infetti con un tasso caratteristico γ .
 Questo modello è descritto dal set di equazioni differenziali:

$$\begin{cases} \dot{S} = -\beta IS \\ \dot{E} = \beta IS - \gamma E \\ \dot{I} = \gamma E - \mu I \\ \dot{R} = \mu I \end{cases} \quad (2.1)$$

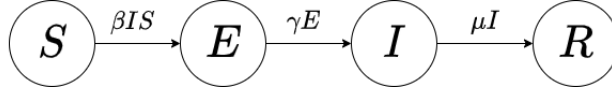


Figura 2.2: Schema del modello *SEIR*.

Se immaginiamo di arrestare le nuove infezioni (ad esempio portando β a 0), il numero di nuovi infetti non diventerà immediatamente nullo a causa della presenza degli esposti; questo compartimento si svuoterà con un tempo caratteristico $\tau = \frac{1}{\gamma}$, che non è altro che il valor medio del tempo di latenza.

Infatti in questo scenario, l'equazione che governa E è semplicemente $\dot{E} = -\gamma E$ da cui si trova immediatamente $E(t) = E_0 e^{-\gamma t} = E_0 e^{-\frac{t}{\tau}}$. Possiamo interpretare questa funzione come una densità di probabilità del tempo di permanenza nello stato E , a patto di normalizzarla tramite $E_0 = \gamma$. In questo modo si ottiene una distribuzione esponenziale, che ha appunto valor medio τ e varianza τ^2 .

2.1.1 Soluzione approssimata e numero di riproduzione R_0

Nelle prime fasi dell'epidemia e fintanto che $S \approx 1$, le equazioni del modello *SEIR* possono essere ben approssimate nel seguente modo:

$$\begin{cases} \dot{S} = -\beta I \\ \dot{E} = \beta I - \gamma E \\ \dot{I} = \gamma E - \mu I \\ \dot{R} = \mu I \end{cases} \quad (2.2)$$

nel quale le equazioni per E ed I formano un sistema lineare di equazioni differenziali ordinarie la cui matrice:

$$A = \begin{pmatrix} -\gamma & \beta \\ \gamma & -\mu \end{pmatrix} \quad (2.3)$$

ha autovalori reali dato che $\Delta = \gamma^2 + \mu^2 + 2\gamma\mu(2\frac{\beta}{\mu} - 1) > 0$ e in particolare sono $\lambda_{\pm} = \frac{1}{2}(-\gamma - \mu \pm \sqrt{\Delta})$. Un sistema fondamentale di soluzioni è dato da $\{e^{\lambda_+ t} v_+, e^{\lambda_- t} v_-\}$ dove v_+ e v_- sono autovettori di A di autovalori λ_+ e λ_- rispettivamente. In particolare, definendo il numero di riproduzione $R_0 = \frac{\beta}{\mu}$, si distinguono 3 casi:

1. $R_0 > 1$, allora $\lambda_+ > 0$ e $\lambda_- < 0$: il ritratto in fase è una sella e c'è una crescita esponenziale di esposti e infetti (per grandi tempi t ciò che domina è la soluzione $e^{\lambda_+ t} v_+$);
2. $R_0 < 1$, allora $\lambda_+ < 0$ e $\lambda_- < 0$: siamo in presenza di un nodo stabile, dunque il numero di esposti e infetti è destinato a scendere esponenzialmente;
3. $R_0 = 1$, allora $\lambda_+ = 0$ e $\lambda_- < 0$: esiste un sottospazio stabile nel quale il numero di esposti e infetti è costante, al di fuori del quale esso scende.

2.2 Modello SE_1E_2IR

Tramite il modello $SEIR$ è dunque possibile calcolare la distribuzione del tempo di latenza, che risulta essere esponenziale; tale distribuzione assume il valore massimo a $t = 0$ indipendentemente dal valor medio, mentre sperimentalmente ¹ ci si aspetterebbe un tipico andamento *a campana* piccato attorno al valore medio e che si annulla per $t \rightarrow 0$ e $t \rightarrow \infty$.

Si costruisce allora il modello SE_1E_2IR nel quale il compartimento degli esposti è suddiviso in due sottostati E_1 ed E_2 , come in figura 2.3, con l'obiettivo di studiare l'impatto di questa suddivisione sulla densità di probabilità del tempo di latenza. Se chiamiamo α il tasso caratteristico di transizione dagli stati E (supposto uguale sia per E_1 che per E_2), le equazioni che governano il modello sono:

$$\begin{cases} \dot{S} = -\beta IS \\ \dot{E}_1 = \beta IS - \alpha E_1 \\ \dot{E}_2 = \alpha E_1 - \alpha E_2 \\ \dot{I} = \alpha E_2 - \mu I \\ \dot{R} = \mu I \end{cases} \quad (2.4)$$

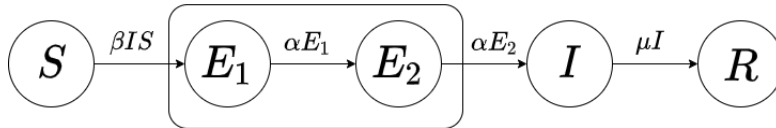


Figura 2.3: Schema del modello SE_1E_2IR .

Siamo qui interessati solo alle equazioni per E_1 ed E_2 in quanto lo scopo è quello di studiare la distribuzione del periodo di latenza analizzando lo svuotamento degli stati esposti, dunque si impone nuovamente $\beta IS = 0$ e si risolve il sistema di equazioni differenziali per E_1 ed E_2 .

¹Non è possibile misurare il tempo di latenza in maniera diretta, dato che è difficile stabilire sia il momento in cui un individuo è stato inizialmente esposto al virus, sia quello in cui egli diventa contagioso; tuttavia è possibile ottenere delle misure indirette, come è stato fatto in [7]. Nello stesso articolo viene calcolato un tempo di latenza medio $\tau = 3.3$ giorni.

2.2.1 Tempo di latenza: stati E_1 ed E_2 accoppiati

Un primo metodo è quello di studiare le due equazioni accoppiate:

$$\begin{cases} \dot{E}_1 = -\alpha E_1 \Rightarrow E_1(t) = E_{1,0} e^{-\alpha t} \\ \dot{E}_2 = \alpha E_1 - \alpha E_2 = \alpha E_{1,0} e^{-\alpha t} - \alpha E_2 \end{cases} \quad (2.5)$$

Si può risolvere l'omogenea associata all'equazione differenziale per E_2 ottenendo $E_2^{omog}(t) = E_{2,0} e^{-\alpha t}$, mentre possiamo cercare una soluzione particolare della forma $E_2^{part}(t) = k t e^{-\alpha t}$ con $k \in \mathbb{R}$, dal momento che α è un autovalore della matrice associata al sistema di equazioni differenziali. Si ottiene infine la soluzione generale, che si può scrivere:

$$E_2(t) = e^{-\alpha t} (E_{2,0} + \alpha E_{1,0} t) \quad (2.6)$$

Volendo interpretare le soluzioni come funzioni di densità di probabilità, bisogna imporre le normalizzazioni $E_{1,0} = \alpha$ e conseguentemente $E_{2,0} = 0$, ovvero $E_2(t) = \alpha^2 t e^{-\alpha t}$.

La variabile aleatoria t che segue questa distribuzione ammette media $\mathbb{E}(t) = \frac{2}{\alpha}$ e varianza $\text{Var}(t) = \frac{2}{\alpha^2}$, come si può ottenere da calcoli diretti o usando le proprietà delle distribuzioni Gamma, come spiegato in 2.3.

Confrontando questi risultati con quelli trovati con il modello $SEIR$ notiamo che, a parità di valor medio di tempo di latenza τ , il modello SE_1E_2IR ha varianza inferiore, infatti:

$$\mathbb{E}(t) = \frac{2}{\alpha} \stackrel{!}{=} \tau = \frac{1}{\gamma} \quad \Rightarrow \quad \alpha = 2\gamma \quad (2.7)$$

$$\Rightarrow \quad \text{Var}(t) = \frac{2}{\alpha^2} = \frac{1}{2\gamma^2} = \frac{\tau^2}{2} < \tau^2 \quad (2.8)$$

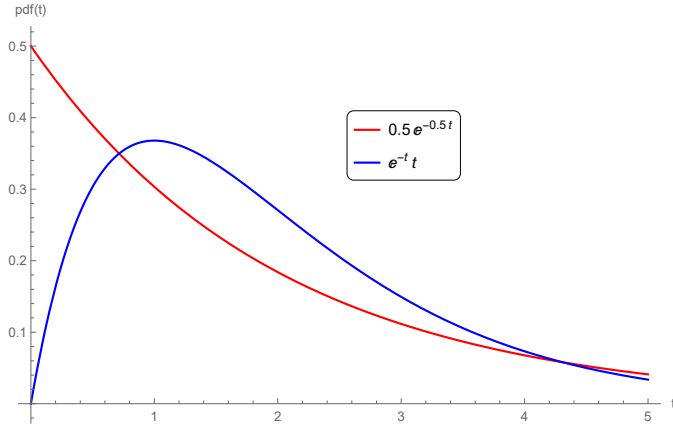


Figura 2.4: Confronto tra le due distribuzioni dei tempi di latenza con parametro $\tau = 2$: quella per il modello $SEIR$ in rosso e quella per il modello SE_1E_2IR in blu.

Oltre ad una diminuzione della varianza, il modello SE_1E_2IR ha altre caratteristiche che lo rendono preferibile al semplice modello $SEIR$: ad esempio in quest'ultimo il massimo valore di probabilità si ottiene per $t = 0$ e decresce poi in modo monotono, mentre nel modello SE_1E_2IR la probabilità è nulla per $t = 0$ e presenta invece un picco poco prima del valor medio τ .

2.2.2 Tempo di latenza: stati E_1 ed E_2 disaccoppiati

Un altro possibile approccio è quello di studiare gli stati E_1 ed E_2 separatamente, ovvero analizzare la distribuzione dei tempi di permanenza in E_1 e poi, indipendentemente, quella in E_2 , per poi inferire la distribuzione del tempo totale di permanenza in uno stato esposto.

In questo caso si ha evidentemente $E_i(t_i) = \alpha e^{-\alpha t_i}$ per $i = 1, 2$ e il tempo di latenza sarà $t = t_1 + t_2$. La funzione di densità di probabilità della somma di due variabili aleatorie indipendenti è data dalla convoluzione delle due densità, ovvero:

$$P(t) := (E_1 * E_2)(t) = \int_{\mathcal{D}} E_1(t-u) E_2(u) du \quad (2.9)$$

Il dominio di integrazione \mathcal{D} va calcolato ricordando che la distribuzione esponenziale è definita solo quando la relativa variabile aleatoria è positiva, da cui:

$$P(t) = \int_0^t \alpha e^{-\alpha(t-u)} \alpha e^{-\alpha u} du = \alpha^2 t e^{-\alpha t} \quad (2.10)$$

I due metodi sono dunque equivalenti e i risultati ottenuti sono conseguenza delle proprietà di una particolare famiglia di distribuzioni, quella delle distribuzioni Gamma.

2.3 Distribuzioni Gamma ed Erlang

Definizione: diciamo che una variabile aleatoria t segue la distribuzione Gamma di parametri n e α , con $n, \alpha \in \mathbb{R}_+$, e scriveremo $t \sim \Gamma(n, \alpha)$, se la relativa funzione di densità di probabilità è:

$$P(t; n, \alpha) = \frac{\alpha^n}{\Gamma(n)} t^{n-1} e^{-\alpha t} \mathbb{I}_{(0, \infty)}(t) \quad (2.11)$$

dove \mathbb{I} è la funzione indicatrice e Γ è la funzione Gamma di Eulero.

In particolare se $n \in \mathbb{N}$ si ha che $\Gamma(n) = (n-1)!$, si dice che la variabile aleatoria t segue una distribuzione di Erlang e si scrive $t \sim \text{Erlang}(n, \alpha)$ con $n \in \mathbb{N}$ e $\alpha \in \mathbb{R}_+$.

Proposizione: siano t_i variabili aleatorie indipendenti con $t_i \sim \Gamma(n_i, \alpha)$ per $i = 1, 2, \dots, m$, allora si ha che:

$$t := \sum_{i=1}^m t_i \sim \Gamma\left(\sum_{i=1}^m n_i, \alpha\right) \quad (2.12)$$

Per la dimostrazione si veda ad esempio [6].

Notiamo poi che per $n = 1$ la densità della distribuzione $\Gamma(n, \alpha)$ si riduce a quella di un'esponenziale di parametro α , ovvero $\Gamma(1, \alpha) = \text{Exp}(\alpha)$. Ne segue immediatamente il seguente corollario.

Corollario: siano t_i variabili aleatorie indipendenti con $t_i \sim \text{Exp}(\alpha)$ per $i = 1, 2, \dots, m$, allora si ha che:

$$t := \sum_{i=1}^m t_i \sim \Gamma(m, \alpha) = \text{Erlang}(m, \alpha) \quad (2.13)$$

Proposizione: sia $t \sim \Gamma(n, \alpha)$, allora t ammette media e varianza e si ha:

$$\mathbb{E}(t) = \frac{n}{\alpha} \quad \text{Var}(t) = \frac{n}{\alpha^2} \quad (2.14)$$

La dimostrazione è un calcolo diretto e si può nuovamente trovare in [6].

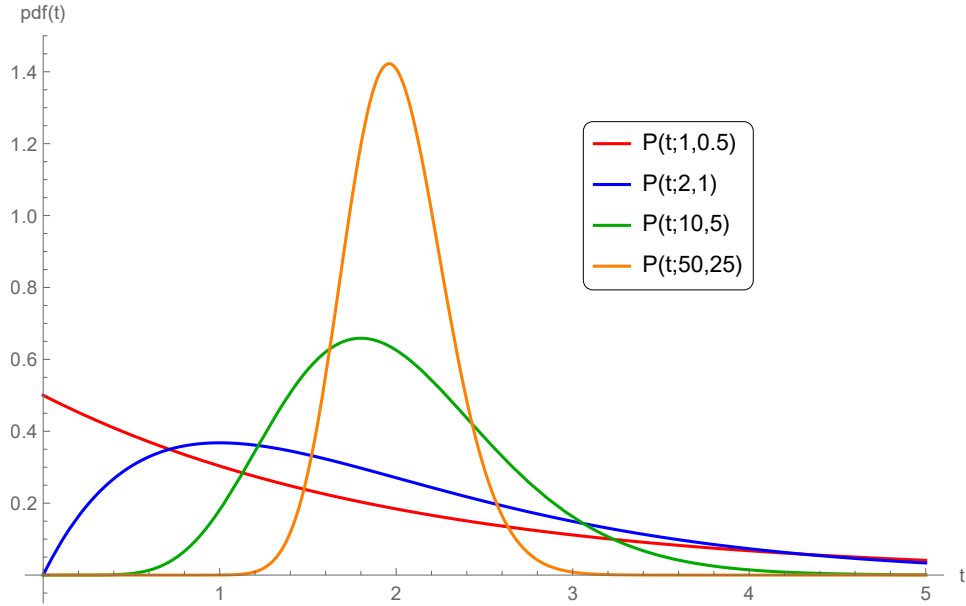


Figura 2.5: Confronto tra varie distribuzioni di Erlang con media fissata $\tau = \frac{n}{\alpha} = 2$.

2.4 Modello $SE_1 \dots E_n IR$

Grazie alle proprietà della distribuzione di Erlang è ora possibile generalizzare quanto fatto in 2.2.2 al caso di n stati esposti successivi E_1, E_2, \dots, E_n con $n \in \mathbb{N}$, ovvero è possibile conoscere la distribuzione del tempo di latenza in un modello $SE_1 \dots E_n IR$, schematizzato in figura 2.6 e descritto dal sistema di equazioni:

$$\begin{cases} \dot{S} = -\beta IS \\ \dot{E}_1 = \beta IS - \alpha E_1 \\ \dot{E}_2 = \alpha E_1 - \alpha E_2 \\ \dots \\ \dot{E}_n = \alpha E_{n-1} - \alpha E_n \\ \dot{I} = \alpha E_n - \mu I \\ \dot{R} = \mu I \end{cases} \quad (2.15)$$

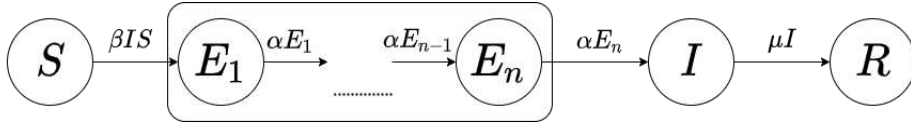


Figura 2.6: Schema del modello $SE_1 \dots E_n IR$.

Ogni stato E_i , infatti, ammette una distribuzione di tempi di permanenza t_i in tale compartimento che segue una legge esponenziale e dunque il tempo di latenza t sarà una variabile aleatoria $t = \sum_{i=1}^n t_i \sim \text{Erlang}(n, \alpha)$.

Imponendo, come in precedenza, che il valor medio di t sia τ si ottiene:

$$\mathbb{E}(t) = \frac{n}{\alpha} \stackrel{!}{=} \tau = \frac{1}{\gamma} \quad \Rightarrow \quad \alpha = n\gamma \quad (2.16)$$

$$\Rightarrow \quad \text{Var}(t) = \frac{n}{\alpha^2} = \frac{1}{n\gamma^2} = \frac{\tau^2}{n} \quad (2.17)$$

Come si vede anche in figura 2.5, le distribuzioni di Erlang con $n \gg 1$ presentano tutte le caratteristiche che si erano richieste inizialmente: si annullano per $t \rightarrow 0$ e per $t \rightarrow \infty$ e hanno un andamento *a campana* nel mezzo, con un picco attorno al valore medio. Di fatto si può calcolare la posizione \bar{t} del massimo:

$$0 \stackrel{!}{=} \frac{d}{dt} \left(\frac{\alpha^n}{\Gamma(n)} t^{n-1} e^{-\alpha t} \mathbb{I}_{(0,\infty)}(t) \right) \Big|_{t=\bar{t}} \quad (2.18)$$

$$\Rightarrow \quad 0 = n - 1 - \alpha \bar{t} \quad \Rightarrow \quad \bar{t} = \tau \left(1 - \frac{1}{n} \right) \xrightarrow{n \rightarrow \infty} \tau \quad (2.19)$$

2.5 Conclusioni

Abbiamo presentato diversi modelli compartimentali al fine di spiegare le caratteristiche salienti della diffusione di virus quali il COVID-19.

Si è studiato come il modello *SHIR* introdotto da T. Barnes in [3] possa essere migliorato in due modi fisicamente differenti ottenendo i modelli *SPIR* e *SS'HIR*, che però risultano matematicamente equivalenti almeno nell'ipotesi di infezione contenuta. A partire dai dati sperimentali riguardanti il numero totale di casi confermati di COVID-19 è stato possibile ricavare i parametri che descrivono la dinamica di questi modelli nel caso di 6 nazioni (tra cui l'Italia) durante la prima ondata di tale virus.

Si è poi analizzato il periodo di latenza osservando come la suddivisione del compartimento E degli esposti in 2 o più sottostati modifichi la sua densità di probabilità, passando da una distribuzione esponenziale a più generiche distribuzioni di Erlang, più adatte a rappresentare il reale andamento di questo parametro.

Si sottolinea che la diffusione di un virus è un fenomeno complesso che non può essere compreso appieno usando un singolo semplice modello compartimentale. Al contrario, modelli diversi danno punti di vista differenti e dunque risultati complementari e tutti necessari per una comprensione più organica del fenomeno.

Bibliografia

- [1] Manash Pratim Barman et al. “COVID-19 pandemic and its recovery time of patients in India: A pilot study”. In: *Diabetes & Metabolic Syndrome: Clinical Research & Reviews* 14.5 (2020), pp. 1205–1211. ISSN: 1871-4021. DOI: <https://doi.org/10.1016/j.dsx.2020.07.004>. URL: <https://www.sciencedirect.com/science/article/pii/S1871402120302502>.
- [2] Sofia Moschin, Enzo Orlandini e Samir Simon Suweis. “Modelli epidemiologici compartimentali: modello SIR e possibili estensioni”. In: *Padua@Thesis* (dic. 2020). URL: <http://tesi.cab.unipd.it/65040/>.
- [3] T. Barnes. *The SHIR Model: Realistic Fits to COVID-19 Case Numbers*. 2020. arXiv: 2007.14804 [q-bio.PE]. URL: <https://arxiv.org/abs/2007.14804>.
- [4] *COVID-19 Data Repository by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University*. URL: <https://github.com/CSSEGISandData/COVID-19>.
- [5] Giacomo Barzon et al. “Modelling the deceleration of COVID-19 spreading”. In: *Journal of Physics A: Mathematical and Theoretical* 54.4 (gen. 2021), p. 044002. DOI: 10.1088/1751-8121/abd59e. URL: <https://doi.org/10.1088/1751-8121/abd59e>.
- [6] Hans-Otto Georgii. *Stochastics*. De Gruyter, 2008. ISBN: 978-3-11-020676-0. DOI: doi:10.1515/9783110206760. URL: <https://doi.org/10.1515/9783110206760>.
- [7] Shi Zhao et al. “Estimating the generation interval and inferring the latent period of COVID-19 from the contact tracing data”. In: *Epidemics* 36 (2021), p. 100482. ISSN: 1755-4365. DOI: <https://doi.org/10.1016/j.epidem.2021.100482>. URL: <https://www.sciencedirect.com/science/article/pii/S1755436521000359>.