

Il picco epidemico in un modello SIR

Quadrature 117 (2020)

<https://hal.archives-ouvertes.fr/hal-02518993>

Nicolas Bacaër

Institut de recherche pour le développement
Unità di modellazione matematica e computer di sistemi complessi
Les Cordeliers, Parigi, Francia
nicolas.bacaer@ird.fr

riassunto

Stiamo studiando un'epidemia modellata da un sistema differenziale di tipo SIR. Quando la popolazione N è grande, il picco dell'epidemia è al momento T , con $T \sim (\ln N)/(a - b)$. a è la velocità di contatto effettiva. b è la velocità con cui le persone infette guariscono.

introduzione

Consideriamo ad esempio in [1, 2, 6, 7, 8] il modello SIR di Kermack e McKendrick per un'epidemia

$$\frac{dS}{dt} = -a\frac{SI}{N}, \quad \frac{dI}{dt} = a\frac{SI}{N} - bI, \quad \frac{dR}{dt} = bI.$$

$S(t)$ è il numero di persone sane. $I(t)$ è il numero di persone infette. $R(t)$ è il numero di persone rimosse dalla trasmissione a causa di isolamento, guarigione o morte. $N = S(t) + I(t) + R(t)$ è la popolazione totale. Si presume che questa popolazione totale sia costante e abbastanza grande da rendere ragionevole modellare l'epidemia usando un sistema differenziale piuttosto che un processo stocastico. Considera le condizioni iniziali

$$S(0) = N - 1, \quad I(0) = 1, \quad R(0) = 0,$$

che corrispondono all'inizio dell'epidemia da un singolo "zero paziente".

Il parametro a (con $a > 0$) è in realtà il prodotto di due numeri: il numero di contatti per unità di tempo e la probabilità di trasmissione durante il contatto tra una persona non infetta e una persona infetta. I contatti sono fatti a caso. I/N è la probabilità di una persona che è nel vano io . Questo spiega il termine quadratico in SI/N nel sistema differenziale. $b > 0$ è il tasso a cui le persone lasciano il vano ho . L'ipotesi di base è quindi che il tempo trascorso nel compartimento I sia distribuito secondo una legge esponenziale del parametro b . Assumeremo $a > b$ che si verifichi una vera epidemia.

Un esempio è dettagliato nella figura 1. Abbiamo usato il software Scilab e la sua funzione di risoluzione dei sistemi differenziali. Questa funzione implementa uno schema predittore-correttore Adams per sistemi non ripidi. Noi scegliamo $N = 65 \times 10^6$ (la popolazione della Francia), $a = 1/2$ al giorno e $b = 1/4$ al giorno. All'inizio dell'epidemia, una persona infetta infetta quindi una media di una persona ogni due giorni. Durata media dell'infezione $1/b$ è di 4 giorni. Ciò corrisponde a una riproducibilità $\mathcal{R}_0 = a/b = 2$ (il termine e la notazione provengono da [9, p. 102]). \mathcal{R}_0 è quindi il numero medio di casi secondari che una persona infetta infetta all'inizio dell'epidemia. Senza pretendere di essere realistici, si tratta di valori dello stesso ordine di grandezza di quello che si può trovare per l'epidemia di coronavirus del 2020 nel caso ipotetico in cui non vi è alcun intervento per limitare l'epidemia [5].

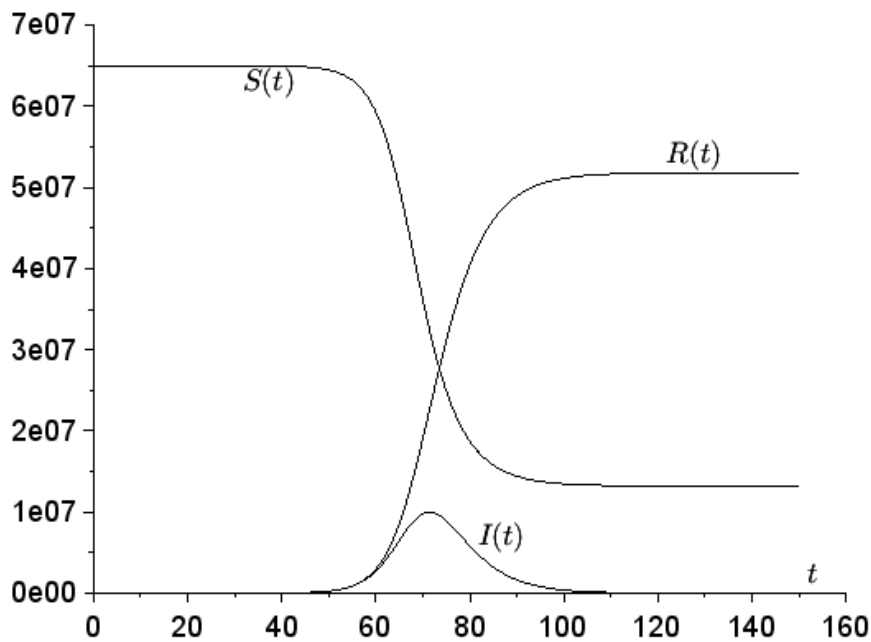


Figura 1. Una simulazione del modello SIR con il tempo t in giorni sull'ascissa .

Il picco dell'epidemia

Combinando la prima e la terza equazione del sistema SIR, è ben noto [8, p. 76] che troviamo un primo integrale:

$$\frac{1}{S} \frac{dS}{dt} = -a \frac{I}{N} = -\frac{a}{bN} \frac{dR}{dt}.$$

Quindi abbiamo

$$\ln[S(t)/S(0)] = -\frac{a}{bN} R(t).$$

Facendo riferimento alla prima equazione, troviamo

$$\frac{dS}{dt} = -a \frac{S}{N} (N - S - R) = -a \frac{S}{N} \left(N - S + \frac{bN}{a} \ln[S(t)/S(0)] \right).$$

Il picco epidemico nella Figura 1 corrisponde a $dI/dt = 0$, che avviene secondo la seconda equazione quando $S(t) = Nb/a$. $S(t)$ è una funzione decrescente perché $dS/dt \leq 0$ dalla prima equazione. Ne deduciamo che se T è il tempo del picco dell'epidemia, cioè se

$$I(T) = \max_{t>0} I(t),$$

abbiamo quindi

$$T = \int_0^T dt = \int_{S(0)}^{bN/a} \frac{dS}{-a \frac{S}{N} \left(N - S + \frac{bN}{a} \ln[S/S(0)] \right)}.$$

Definiamo $s = S/N$. Abbiamo quindi

$$T = \frac{1}{a} \int_{b/a}^{1-1/N} \frac{ds}{s \left(1 - s + \frac{b}{a} \ln[s/(1 - 1/N)] \right)}. \quad (1)$$

Cerchiamo un limite superiore di T . Per tutto $b/a < s < 1 - 1/N$, noi abbiamo

$$\frac{1}{s(1-s + \frac{b}{a} \ln[s/(1-1/N)])} \leq \frac{1}{s(1-s + \frac{b}{a} \ln s)}.$$

Nel quartiere di $s = 1$, noi abbiamo

$$\frac{1}{s(1-s + \frac{b}{a} \ln s)} \sim \frac{1}{(1-s)(1-b/a)}.$$

L'integrale

$$\int_{b/a}^1 \frac{ds}{1-s}$$

è divergente. Con $N \rightarrow +\infty$, Quindi abbiamo

$$T \leq \frac{1}{a} \int_{b/a}^{1-1/N} \frac{ds}{s(1-s + \frac{b}{a} \ln s)} \sim \frac{1}{a} \int_{b/a}^{1-1/N} \frac{ds}{(1-s)(1-b/a)} \sim \frac{\ln N}{a-b}.$$

Cerchiamo un limite inferiore di T . abbiamo $\ln s \leq s - 1$ per tutto $s > 0$. Abbiamo anche $\ln(1 - 1/N) \geq -2/N$ per tutti $N \geq 2$. Quindi abbiamo per $s \in]b/a, 1 - 1/N[$ e per qualsiasi numero intero $N \geq 2$,

$$\begin{aligned} 1-s + \frac{b}{a} \ln[s/(1-1/N)] &= 1-s + \frac{b}{a} \ln s - \frac{b}{a} \ln(1-1/N) \\ &\leq 1-s + \frac{b}{a}(s-1) + \frac{2b}{aN}. \end{aligned}$$

Quindi abbiamo

$$T \geq \frac{1}{a} \int_{b/a}^{1-1/N} \frac{ds}{s[1 - \frac{b}{a} + \frac{2b}{aN} - (1-b/a)s]}.$$

Incontriamo una frazione razionale della forma $\frac{1}{s(u-vs)}$ con

$$u = 1 - \frac{b}{a} + \frac{2b}{aN}, \quad v = 1 - b/a.$$

Si scompone in elementi semplici

$$\frac{1}{s(u-vs)} = \frac{1}{u} \left(\frac{1}{s} + \frac{v}{u-vs} \right).$$

Una primitiva è

$$\frac{1}{u} [\ln s - \ln(u-vs)] = \frac{1}{u} \ln \frac{s}{u-vs}.$$

Quindi abbiamo

$$\begin{aligned}
T &\geq \frac{1}{a(1 - \frac{b}{a} + \frac{2b}{aN})} \ln \left\{ \frac{1 - 1/N}{b/a} \frac{1 - \frac{b}{a} + \frac{2b}{aN} - (1 - b/a)b/a}{1 - \frac{b}{a} + \frac{2b}{aN} - (1 - b/a)(1 - 1/N)} \right\} \\
&= \frac{1}{a - b + 2b/N} \ln \left\{ \frac{1 - 1/N}{b/a} \frac{(1 - b/a)^2 + \frac{2b}{aN}}{(1 + b/a)/N} \right\} \\
&= \frac{1}{a - b + 2b/N} \left[\ln N + \ln \left\{ \frac{1 - 1/N}{b/a} \frac{(1 - b/a)^2 + \frac{2b}{aN}}{(1 + b/a)} \right\} \right].
\end{aligned}$$

Il membro a destra è $\sim (\ln N)/(a - b)$ Se $N \rightarrow +\infty$. Dato che è lo stesso equivalente di quello del limite superiore ottenuto sopra, lo concludiamo $T \sim (\ln N)/(a - b)$ Se $N \rightarrow +\infty$.

Nell'esempio in Figura 1, abbiamo $T \simeq 71,4$ giorni mentre $(\ln N)/(a - b) \simeq 72,0$ giorni. L'approssimazione sembra abbastanza buona. Nella Figura 2, variamo la dimensione N della popolazione e il parametro a , con b fisso. Notiamo $(\ln N)/(a - b)$ a volte sopravvalutata, a volte sottovalutata T .

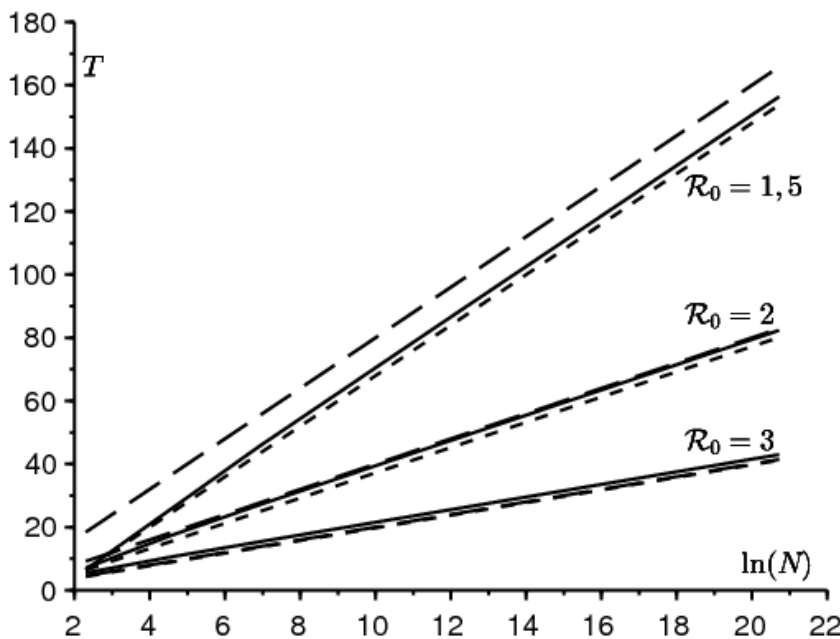


Figura 2. Confronto tra da un lato la data T del picco dell'epidemia (linea continua), calcolata con l'integrale della formula (1), dall'altro l'equivalente $(\ln N)/(a - b)$ (grandi linee tratteggiate) e l'approssimazione $[\ln(2N(b/a - 1)^2)]/(a - b)$ (piccole linee tratteggiate), a seconda di $\ln N$, Se $10 \leq N \leq 10^8$ per tre valori di parametro a con riproducibilità $\mathcal{R}_0 = a/b \in \{1,5 ; 2 ; 3\}$.

$a - b$ è il tasso di crescita iniziale dell'epidemia. Abbiamo quindi per N grande, $S(t) \simeq N$ e $dI/dt \simeq (a - b)I$. Questo tasso può essere stimato da dati epidemiologici. La popolazione N è nota. Possiamo quindi prevedere la data del picco dell'epidemia. In pratica, l'ipotesi alla base del modello di una miscela omogenea della popolazione rimane dubbia per un intero paese, un po' meno per una città.

Storicamente, Kermack e McKendrick avevano approssimativamente risolto il sistema SIR assumendo che la riproducibilità fosse vicina a 1, il che ha portato a una piccola epidemia. Con l'approssimazione $e^x \simeq 1 + x + x^2/2$, erano arrivati a un'equazione di Riccati esplicitamente risolvibile, quindi

$$I(t) \simeq \frac{X}{\text{ch}^2(Yt - Z)}. \quad (2)$$

X , Y e Z sono funzioni complicate dei parametri N , a e b . Vedi ad esempio [3]. Il picco dell'epidemia è $T \simeq Z/Y$. Notiamo per inciso che per nostra conoscenza l'approssimazione (2) non è mai stata

rigorosamente dimostrata, motivo per cui usiamo sopra e sotto il segno informale \simeq . Ora assumiamo inoltre $1/N \ll (\mathcal{R}_0 - 1)^2$. In quel caso,

$$Y \simeq \frac{a-b}{2}, \quad Z \simeq \frac{1}{2} \ln[2N(b/a - 1)^2]$$

(vedi ad esempio [4]), quindi

$$T \simeq \frac{\ln[2N(b/a - 1)^2]}{a-b}.$$

Questa approssimazione è mostrata anche in Figura 2; sembra particolarmente vicino al vero valore di T se \mathcal{R}_0 è vicino a 1, come ci si aspetterebbe. Con N che converge all'infinito, abbiamo comunque $T \simeq \frac{\ln N}{a-b}$. Troviamo lo stesso comportamento asintotico. La nostra analisi basata sulla formula (1) è valida, tuttavia, per tutti $\mathcal{R}_0 > 1$ e non solo per \mathcal{R}_0 vicino a 1. Inoltre, la nostra prova è rigorosa.

Conclusioni

Abbiamo scoperto che il picco dell'epidemia si è verificato alla volta $T \sim (\ln N)/(a-b)$ se N converge all'infinito. Questo risultato sembra nuovo nonostante la sua semplicità. Nella letteratura scientifica, abbiamo trovato solo informazioni sull'altezza del picco, che è facilmente determinata dal primo integrale del sistema. Anzi, con

$$S(T) = \frac{Nb}{a}, \quad \ln \frac{S(T)}{S(0)} = -\frac{a}{bN} R(T), \quad N = S(T) + I(T) + R(T),$$

otteniamo

$$\frac{I(T)}{N} = 1 - \frac{b}{a} + \frac{b}{a} \ln \left(\frac{N}{N-1} \frac{b}{a} \right) \simeq 1 - \frac{b}{a} + \frac{b}{a} \ln \frac{b}{a}$$

(vedi ad esempio [1, equazione 7.26]). In alcuni riferimenti come [7, p. 238], ci sono inversioni in cui il tempo t è una funzione delle variabili monotoniche R o S , che porta a integrali come (1) ad eccezione di un cambiamento di variabile. Ma il sistema differenziale SIR è stato scritto lì nella forma $dS/dt = -\alpha SI$. L'idea probabilmente non è nata dal prendere $\alpha = a/N$, che è in qualche modo più realistico e necessario affinché i confini di una grande popolazione producano risultati interessanti.

Per un'epidemia come quella del coronavirus, il modello SIR manca di un po' di realismo. Ciò porta a considerare modelli con più compartimenti, ad esempio con un compartimento aggiuntivo per le persone in fase latente che sono infette ma non ancora contagiose. Vedi su questo argomento [5], che cerca anche di stimare i parametri dalla curva epidemica. Lo studio analitico del picco epidemico in questo contesto è senza dubbio più difficile.

Grazie

Ringraziamo Hisashi Inaba per i suoi commenti sul testo.

Riferimenti bibliografici

1. B. Anselme, *Biomathématiques*, Dunod, Paris, 2015.
2. N. Bacaër, *Histoires de mathématiques et de populations*, Cassini, Paris, 2009.
3. N. Bacaër (2012) Le modèle de Kermack et McKendrick pour la peste à Bombay et la reproductivité nette d'un type avec saisonnalité. *Journal of Mathematical Biology* 64 (2012) 403 – 422. <https://hal.archives-ouvertes.fr/hal-01340008>

4. N. Bacaër (2019) Sur la taille finale des épidémies dans un environnement périodique. *Comptes Rendus Biologies* 342 (2019) 119 – 123. <https://doi.org/10.1016/j.crvi.2019.07.001>
5. N. Bacaër (2020) Un modèle mathématique de l'épidémie de coronavirus en France. <https://hal.archives-ouvertes.fr/hal-02509142>
6. M. Corlosquet – Habart, J. Janssen, R. Manca, Modélisation stochastique du risque de pandémie : stratégies de couverture et d'assurance, Lavoisier, Cachan, 2012.
7. J. Gani, *Processus stochastiques de population*, P. –L. Hennequin (éd.), École d'été de probabilités de Saint – Flour IV, Springer, Berlin, 1975, p. 188 – 293.
8. A. Hillion (1986) *Les Théories mathématiques des populations*. Presses Universitaires de France, Paris.
9. A. J. Lotka (1939) *Théorie analytique des associations biologiques*, 2e partie. Hermann, Paris.