Modelli di epidemie con i branching processes

Federico Poloni, Università di Pisa
`http://pages.di.unipi.it/fpoloni/`

Sommario

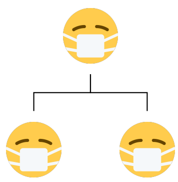
In questo documento presento un modello per descrivere l'evoluzione delle epidemie usando uno strumento matematico chiamato *branching processes* (processi che si ramificano) o *Galton–Watson processes*.

Questo modello è abbastanza semplice da spiegare (se sapete cos'è un'equazione di secondo grado riuscirete a seguirmi), mostra chiaramente cos'è la quantità R_0 di cui si sente parlare in giro, e a differenza di altri modelli è adatto a descrivere anche situazioni con pochi infetti.

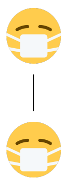
1 Il modello

Data una persona infetta  in un dato giorno,

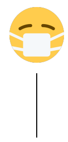
- con una probabilità a, la persona contagia un'altra persona; quindi il giorno successivo avremo due infetti, quello di prima più uno nuovo;



- con una probabilità b, non succede nulla; quindi il giorno successivo la persona sarà di nuovo infetta;



- con una probabilità c, la persona non sarà più infetta il giorno dopo.



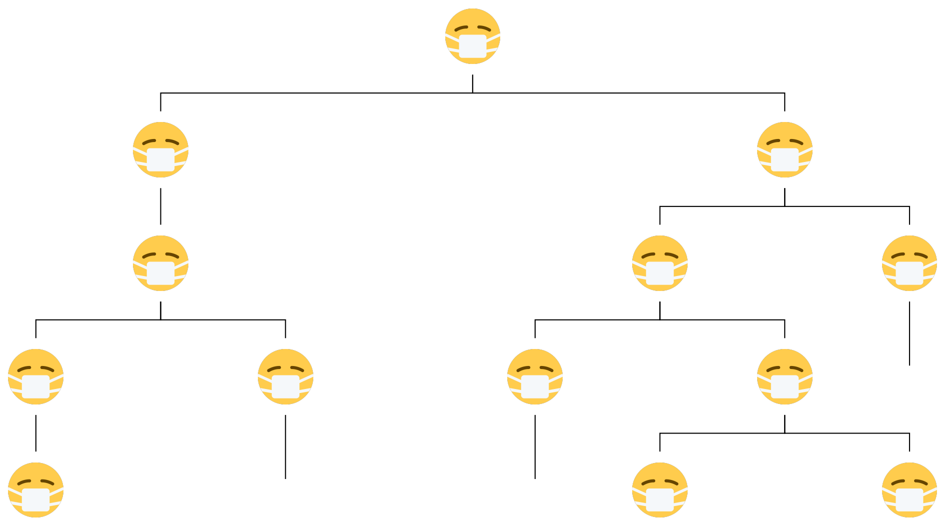
  Perché guarisce, perché muore, perché se ne va: non ci interessa. Chiaramente la differenza tra un guarito e un morto nella vita reale è importante, ma per il nostro modello quello che importa è solo che il giorno dopo abbiamo un infetto in meno. Per semplicità lo chiameremo guarito. Qualche volta si usa il termine *rimosso*.

Poi sceglieremo dei valori per a, b, c, per esempio,

$$a = \frac{1}{2}, \qquad b = \frac{1}{6} \qquad c = \frac{1}{3}. \tag{1}$$

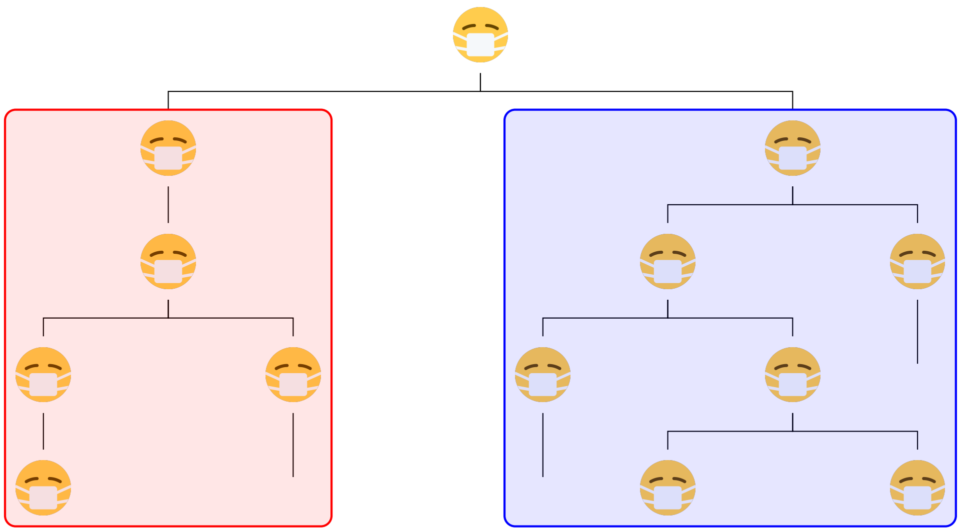
Nel nostro modello, deve succedere per forza una di queste tre cose, e nessun'altra: in particolare, ogni persona non può contagiare due o più persone nello stesso giorno. quindi dovremo scegliere i valori in modo che $a+b+c = 1$.

Ogni nuovo infetto poi procede indipendentemente dagli altri sulla base di queste regole. Partendo da un singolo infetto (il *paziente zero*), mano a mano che i giorni crescono avremo sempre nuovi contagi. Possiamo rappresentarli in un diagramma (ogni riga corrisponde a un giorno).



I matematici e gli informatici chiamano una struttura di questo tipo *albero*. Come dice una barzelletta, si vede che non escono mai di casa, perché credono che gli alberi abbiano il tronco in alto e le foglie in basso.

Comunque, la proprietà interessante è che ogni volta che c'è un contagio questo albero si divide in due 'sotto-epidemie' (qui evidenziate in rosso e in blu), che sono indipendenti tra loro e seguono le stesse regole dell'epidemia iniziale.



Questa osservazione è alla base dei trucchi che useremo per calcolare varie quantità.

2 Probabilità di estinzione

Cominciamo a studiare questo modello. Useremo solo alcuni fatti base di probabilità, come il seguente: se la probabilità che succeda una cosa è p, e la probabilità che ne succeda un'altra è q, allora la probabilità che succedano tutte e due insieme è il loro prodotto, pq (ammesso che non dipendano in nessun modo l'una dall'altra). Per esempio, se lancio una moneta la probabilità che venga testa è $\frac{1}{2}$; se ne lancio due indipendentemente, la probabilità che vengano due teste è $\frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$.

Partiamo da un singolo infetto, un "paziente zero". La domanda interessante è: qual è la probabilità di sconfiggere l'epidemia, cioè arrivare in un certo giorno futuro ad avere zero infetti? Chiamiamo x questa probabilità:

$$x = \text{probabilità di sconfiggere l'infezione.}$$

Chiediamoci cosa succede il primo giorno.

1. Se avviene un contagio (con probabilità a), allora il giorno successivo abbiamo due persone infette, che danno origine a due 'sotto-epidemie',

come quella in rosso e quella in blu nella figura sopra. Ognuna di queste sotto-epidemie si comporta come l'epidemia originale: in particolare, per ognuna delle due c'è una probabilità x che tutti gli infetti guariscano. Quindi la probabilità che *tutte e due* le sotto-epidemie vengano sconfitte è $x \cdot x = x^2$;

2. se il primo giorno non succede nulla (con probabilità b), allora il giorno successivo abbiamo sempre un infetto e un'epidemia da sconfiggere, con probabilità x;

3. se il primo giorno il paziente zero guarisce (con probabilità c), allora abbiamo vinto: sconfiggiamo sempre l'epidemia, con probabilità 1.

Mettendo insieme questi tre casi, abbiamo un'equazione che la probabilità x deve soddisfare:

$$x = \underbrace{a \cdot x^2}_{\text{caso 1}} + \underbrace{b \cdot x}_{\text{caso 2}} + \underbrace{c \cdot 1}_{\text{caso 3}} . \tag{2}$$

Possiamo riscriverla come

$$ax^2 + (b-1)x + c = 0,$$

e questa è un'equazione che probabilmente avete già visto. L'unica differenza rispetto a una classica equazione di secondo grado è che abbiamo scritto $b-1$ al posto di b. La formula risolutiva è

$$x = \frac{-(b-1) \pm \sqrt{(b-1)^2 - 4ac}}{2a};$$

nel nostro esempio (1) quindi le due soluzioni sono

$$x = \frac{-(\frac{1}{6}-1) \pm \sqrt{(\frac{1}{6}-1)^2 - 4 \cdot \frac{1}{2} \cdot \frac{1}{3}}}{2 \cdot \frac{1}{2}} = \frac{\frac{5}{6} \pm \frac{1}{6}}{1} = \begin{cases} \dfrac{2}{3}, \\ 1. \end{cases}$$

Un nostro amico matematico saprebbe dirci, in generale, come trovare le soluzioni facilmente per una qualunque scelta di a, b, c.

- Una delle due soluzioni è sempre 1: difatti, se sostituiamo $x = 1$ nell'equazione (2) otteniamo

  $$1 = a + b + c,$$

  che è sempre vera per quello che abbiamo detto sulla scelta di a, b, c.

- L'altra soluzione è sempre $\frac{c}{a}$: difatti, c'è un risultato che dice che in un'equazione di secondo grado il prodotto delle due soluzioni fa sempre $\frac{c}{a}$; quindi se una soluzione è 1 l'altra sarà $\frac{c}{a}$.

Qual è il problema qui? Che le soluzioni sono due! Quale delle due è la *vera* probabilità di sconfiggere l'epidemia? $x = \frac{c}{a}$, oppure $x = 1$?

Osserviamo che $x = 1$ vuol dire che sconfiggiamo sempre l'epidemia, quindi ottime notizie. Nel nostro esempio, anche $x = \frac{2}{3}$ comunque sembra una probabilità alta, ma è perché siamo partiti da un solo infetto iniziale (paziente zero): se partiamo avendo 100 infetti iniziali, la probabilità di guarire tutti è

$$1^{100} = 1 \qquad \text{oppure} \qquad \left(\frac{2}{3}\right)^{100} \approx 0{,}000000000000000002, \tag{3}$$

e la differenza tra i due casi diventa sostanziale!

3 Qual è la soluzione giusta?

La risposta viene da un trucco simile a quello che abbiamo già usato. Invece che x, consideriamo

$$x_k = \text{probabilità di sconfiggere l'epidemia entro i primi } k \text{ giorni.}$$

Cosa possiamo dire sugli x_k? Innanzitutto, $x_1 = c$: l'unico modo di sconfiggere l'epidemia entro un giorno è che il paziente zero guarisca immediatamente.

Poi possiamo trovare una formula che lega x_{k+1} a x_k: guardiamo, di nuovo, cosa succede il primo giorno.

1. Se il primo giorno il paziente zero ne infetta un altro, allora ci ritroviamo con due sotto-epidemie da sconfiggere in k giorni, e questo avviene con probabilità $x_k \cdot x_k$.

2. Se il primo giorno non succede nulla, allora abbiamo sempre un'epidemia da sconfiggere, ma dobbiamo farlo in k giorni: la probabilità è x_k.

3. Se il primo giorno il paziente zero guarisce, allora abbiamo sicuramente sconfitto l'epidemia entro k giorni: l'abbiamo sconfitta addirittura il primo giorno!

I tre casi quindi ci danno

$$x_{k+1} = \underbrace{a \cdot x_k^2}_{\text{caso 1}} + \underbrace{b \cdot x_k}_{\text{caso 2}} + \underbrace{c \cdot 1}_{\text{caso 3}}.$$

Abbiamo quindi queste due formule:

$$x_1 = c, \qquad\qquad x_{k+1} = ax_k^2 + bx_k + c \quad \text{per ogni } k \geq 1. \tag{4}$$

Possiamo ricavare x_2 a partire da x_1, poi x_3 a partire da x_2, e così via.

Se sapete programmare, potete usare un computer per calcolare i primi termini di questa successione: per esempio, potete scrivere questo programma Python 3 che restituisce i primi 50 elementi della successione.

```
a = 1/2
c = 1/3
b = 1 - a - c

x = [None] * 50
x[0] = c
for k in range(49):
        x[k+1] = a*x[k]**2 + b*x[k] + c

print(x)
```

Il risultato è

```
[0.3333333333333333, 0.4444444444444444, 0.5061728395061729,
0.5458009449778998, 0.5732494932657008, 0.5931824063089748,
0.6081297512954154, 0.6195991890878811, 0.6285514424071601,
0.6356303649439208, 0.6412847079100243, 0.6458371562846267,
0.6495256755996777, 0.6525294142315552, 0.6549855539239475,
0.6570006302451881, 0.6586583524454852, 0.6600251380306814,
0.6611541144213235, 0.6620880672450096, 0.662861648934951,
0.663503057636873, 0.6640353300195521, 0.6644773480936796,
0.6648446310804179, 0.6651499635849648, 0.6654038976260341,
0.6656151564256312, 0.665790960969364, 0.6659372953491485,
0.6660591232266608, 0.6661605650211775, 0.6662450433648635,
0.6663154027982683, 0.6663740084694874, 0.6664228276600941,
0.6664634972232535, 0.6664973794360671, 0.6665256083035835,
0.666549127979495, 0.6665687246683618, 0.6665850531310635,
0.666598658717382, 0.6666099956881366, 0.6666194424570249,
0.6666273142736948, 0.6666338737801623, 0.666639339798933,
0.6666438946502674, 0.6666476902456163]
```

Da questi numeri si intuisce che il limite sarà $\frac{2}{3} = 0.666666\ldots$ e non 1. Ma come dimostrarlo? E cosa succede con valori diversi di a, b, c?

Di nuovo il nostro amico matematico ci viene in aiuto, dimostrando per noi questo teorema.

Teorema 1. *La probabilità di sconfiggere l'epidemia x, cioè il limite degli x_k, è sempre la più piccola delle due soluzioni dell'equazione* (2).

Dimostrazione. Chiamiamo p la più piccola delle due soluzioni di (2). Ci basta dimostrare che i valori x_k non superano mai p. Se $p - x_k \geq 0$, allora

$$\begin{aligned} p - x_{k+1} &= (ap^2 + bp + c) - (ax_k^2 + bx_k + c) \\ &= a(p^2 - x_k^2) + b(p - x_k) \\ &= a(p + x_k)(p - x_k) + b(p - x_k) \geq 0, \end{aligned}$$

perché i termini che compaiono nell'ultima riga (cioè a, b, $p + x_k$, e $p - x_k$) sono tutti maggiori o uguali a zero. □

Questo teorema sembra fatto da un pessimista, perché ci dice che la probabilità di sconfiggere l'epidemia è sempre il più piccolo tra i due possibili candidati. Però quello che ci dice è ragionevole:

- Se $\frac{c}{a} > 1$, cioè $c > a$, è più probabile guarire che non contagiare qualcun altro: in questo caso sconfiggeremo praticamente sempre l'epidemia (probabilità $= 1$).
- Se $\frac{c}{a} < 1$, cioè $c < a$, è più probabile contagiare qualcun altro che non guarire; in questo caso non è per nulla certo che sconfiggeremo l'epidemia (probabilità < 1). La probabilità che questo succeda non è zero, ma serve un colpo di fortuna, perché tutti gli infetti in un dato momento devono guarire contemporaneamente.

4 Ma R_0?

In TV l'esperto di turno spesso parla del "valore R_0"; cosa c'entra con tutto questo? In vari modelli, R_0 è definito come il *numero medio* di persone che un singolo infetto contagia prima di guarire. Come possiamo calcolarlo per questo modello? Con il solito trucco: vediamo cosa succede il primo giorno.

- Se il primo giorno il paziente zero ne contagia un altro (probabilità che succeda a), allora il numero di persone che contagia è pari a 1 (quello appena contagiato), più il numero medio di persone che contagia dal secondo giorno in poi, che è di nuovo R_0 perché la situazione è la stessa di quella di partenza.

- Se il primo giorno non succede nulla (probabilità che succeda b), allora ci interessa solo il numero medio di persone che contagia dal secondo giorno in poi, che è sempre R_0.

- Se il paziente zero guarisce già il primo giorno (probabilità che succeda c), allora non contagia nessuno, e il numero medio è 0.

Quindi, mettendo insieme questi tre casi, abbiamo

$$R_0 = \underbrace{a(1+R_0)}_{\text{caso 1}} + \underbrace{bR_0}_{\text{caso 2}} + \underbrace{c \cdot 0}_{\text{caso 3}}, \tag{5}$$

e riordinando

$$a = (1-a-b)R_0 = cR_0, \quad \text{quindi} \quad R_0 = \frac{a}{c}.$$

In particolare ritroviamo le osservazioni che avevamo già fatto.

- Se $R_0 < 1$ allora vuol dire che $c > a$: è più probabile guarire che contagiare qualcun altro, e sconfiggiamo sempre l'epidemia.

- Se al contrario $R_0 > 1$, cioè $c < a$, è più probabile contagiare che guarire, e l'epidemia non terminerà, a meno di un colpo di fortuna. Come abbiamo visto nell'esempio (3), la probabilità di questo colpo di fortuna potrebbe essere una su un miliardo di miliardi.

Stare a casa riduce la probabilità di contagio, quindi ci aiuta a passare da un caso all'altro.

5 Credits e commenti

Ringrazio Samuele Mongodi (PoliMI) e Gabriele Dalla Torre (UniTN) per alcuni commenti su una prima versione.

Ci sono diversi modi in cui possiamo migliorare questo modello per renderlo più simile alla realtà: per esempio, introduciamo la possibilità di contagiare $2, 3, 4, \ldots$ persone nello stesso giorno; oppure consideriamo tipi diversi di infetti (per esempio: sintomatici e asintomatici); oppure ammettiamo che a, b, c possano variare nel tempo. Tutti questi modelli sono stati considerati dai matematici, ma il loro studio è più complicato.