Stima della curva delle infezioni da COVID-19 in Italia e valutazione delle prospettive

Andrea Pugliese, Sara Sottile

Dip. Matematica, Università di Trento

Modelli collettivi, controllo e quantificazione dell'incertezza per malattie infettive e problemi correlati
Workshop Elettronico. 4 Aprile 2020

Outline

Dati disponibili

I dati pubblicati ogni giorno dalla Protezione Civile includono per ogni regione

- il numero di contagiati accertati nel giorno (e cumulativi dall'inizio)

[1] o accertati in questo giorno

Dati disponibili

I dati pubblicati ogni giorno dalla Protezione Civile includono per ogni regione

- il numero di contagiati accertati nel giorno (e cumulativi dall'inizio)
- il numero di tamponi effettuati

[1] o accertati in questo giorno

Dati disponibili

I dati pubblicati ogni giorno dalla Protezione Civile includono per ogni regione

- il numero di contagiati accertati nel giorno (e cumulativi dall'inizio)
- il numero di tamponi effettuati
- il numero di morti infetti da COVID-19 nel giorno[1] (e cumulativi)

[1] o accertati in questo giorno

Dati disponibili

I dati pubblicati ogni giorno dalla Protezione Civile includono per ogni regione

- il numero di contagiati accertati nel giorno (e cumulativi dall'inizio)
- il numero di tamponi effettuati
- il numero di morti infetti da COVID-19 nel giorno[1] (e cumulativi)
- il numero di persone attualmente ospedalizzate per COVID-19

[1] o accertati in questo giorno

Dati disponibili

I dati pubblicati ogni giorno dalla Protezione Civile includono per ogni regione

- il numero di contagiati accertati nel giorno (e cumulativi dall'inizio)
- il numero di tamponi effettuati
- il numero di morti infetti da COVID-19 nel giorno[1] (e cumulativi)
- il numero di persone attualmente ospedalizzate per COVID-19
- il numero di persone attualmente in terapia intensiva per COVID-19

[1] o accertati in questo giorno

Dati disponibili

I dati pubblicati ogni giorno dalla Protezione Civile includono per ogni regione

- il numero di contagiati accertati nel giorno (e cumulativi dall'inizio)
- il numero di tamponi effettuati
- il numero di morti infetti da COVID-19 nel giorno[1] (e cumulativi)
- il numero di persone attualmente ospedalizzate per COVID-19
- il numero di persone attualmente in terapia intensiva per COVID-19
- il numero di persone dimesse perché guarite.

[1] o accertati in questo giorno

Dati disponibili

I dati pubblicati ogni giorno dalla Protezione Civile includono per ogni regione

- il numero di contagiati accertati nel giorno (e cumulativi dall'inizio)
- il numero di tamponi effettuati
- il numero di morti infetti da COVID-19 nel giorno[1] (e cumulativi)
- il numero di persone attualmente ospedalizzate per COVID-19
- il numero di persone attualmente in terapia intensiva per COVID-19
- il numero di persone dimesse perché guarite.

Lo scopo principale di questa analisi è di ricostruire la curva delle infezioni, il dato più importante per comprendere l'andamento dell'epidemia.

[1] o accertati in questo giorno

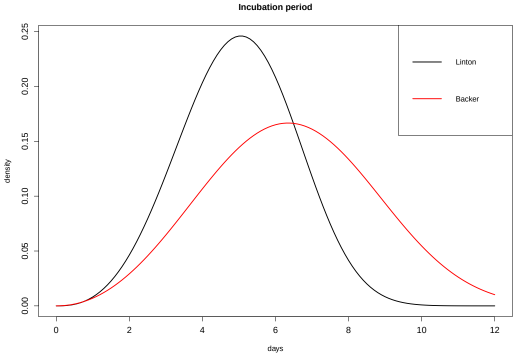
Distribuzioni dei tempi di ritardo

Ho scelto di concentrami sui dati giornalieri dei morti e dei nuovi ospedalizzati come gli indicatori accessibili più affidabili per ricostruire la curva.

Distribuzioni dei tempi di ritardo

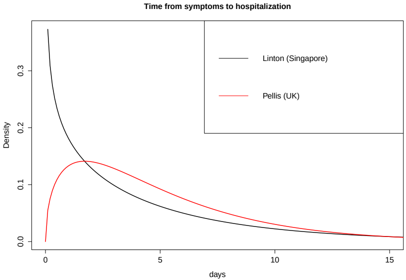
Ho scelto di concentrami sui dati giornalieri dei morti e dei nuovi ospedalizzati come gli indicatori accessibili più affidabili per ricostruire la curva.

Numerosi studi hanno stimato i tempi di incubazione (dall'infezione alla comparsa dei sintomi), dalla comparsa dei sintomi all'ospedalizzazione o alla morte.

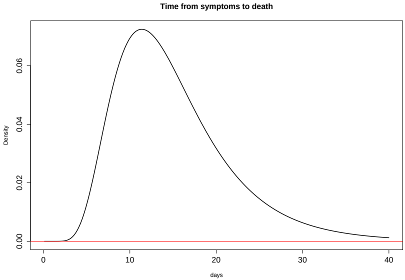


Tempi dalla comparsa dei sintomi

all'ospedalizzazione



alla morte

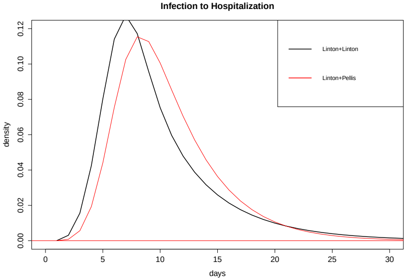


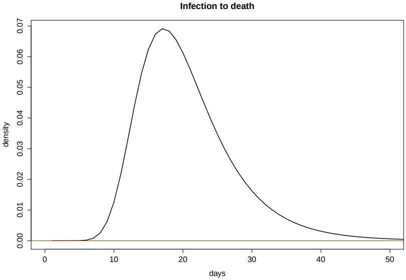
Tempi dall'infezione

Basta fare la convoluzione fra la densità del tempo di incubazione con quelle dalla comparsa dei sintomi. Quindi

$f_1(\tau)$ densità del tempo dall'infezione all'ospedalizzazione

$f_2(\tau)$ densità del tempo dall'infezione alla morte





Obiettivo: dai dati alla curva delle infezioni

Conosciamo[2] gli ospedalizzati H_t e i morti D_t nel giorno t. Date le funzioni (supposte note) ritardo $f_1(\cdot)$ e $f_2(\cdot)$ e la funzione incidenza (tasso di nuove infezioni) $j(\cdot)$ che vogliamo ricostruire, si ha

$$
\begin{aligned}
H_t &= p_1 \int_{t-1}^{t} \int_0^s f_1(s-u) j(u)\, du\, ds \\
&= p_1 \int_0^{t-1} j(u) \int_{t-1}^{t} f_1(s-u)\, ds\, du + p_1 \int_{t-1}^{t} j(u) \int_s^t f_1(s-u)\, ds\, du \\
&\approx p_1 \int_0^{t-1} j(u) f_1(t - \frac{1}{2} - u) du
\end{aligned}
$$

supponendo che le infezioni siano cominciate al tempo $t = 0$, dove p_1 è la probabilità che un infetto sia ospedalizzato.

[2] più o meno

Equazioni integrali

Analogamente per i dati dei morti

$$D_t \approx p_2 \int_0^{t-1} j(u) f_2(t - \frac{1}{2} - u) du$$

dove p_2 è la probabilità che un infetto muoia.

Equazioni integrali

Analogamente per i dati dei morti

$$D_t \approx p_2 \int_0^{t-1} j(u) f_2(t - \frac{1}{2} - u) du$$

dove p_2 è la probabilità che un infetto muoia.

Le equazioni ottenute sono equazioni di Volterra di 1a specie (ignoriamo il problema che valgono sono a tempi discreti).
In teoria ognuna delle due sarebbe sufficiente a ricostruire esattamente $j(\cdot)$.

Equazioni integrali

Analogamente per i dati dei morti

$$D_t \approx p_2 \int_0^{t-1} j(u) f_2(t - \frac{1}{2} - u) du$$

dove p_2 è la probabilità che un infetto muoia.

Le equazioni ottenute sono equazioni di Volterra di 1a specie (ignoriamo il problema che valgono sono a tempi discreti).
In teoria ognuna delle due sarebbe sufficiente a ricostruire esattamente $j(\cdot)$.

Però si tratta di problemi notoriamente mal-posti, per i quali sono state spesso proposte tecniche di regolarizzazione, rimanendo in ambito analitico.

Setting probabilistico

Ipotizziamo che

$$H_t \sim \text{Poisson}\left(p_1 \int_0^{t-1} j(u) f_1(t - \frac{1}{2} - u) du\right)$$

$$D_t \sim \text{Poisson}\left(p_2 \int_0^{t-1} j(u) f_2(t - \frac{1}{2} - u) du\right)$$

fra loro indipendenti, e che $j(u)$ sia il risultato di un modello epidemico.

Setting probabilistico

Ipotizziamo che

$$H_t \sim \text{Poisson}\left(p_1 \int_0^{t-1} j(u) f_1(t - \frac{1}{2} - u) du\right)$$

$$D_t \sim \text{Poisson}\left(p_2 \int_0^{t-1} j(u) f_2(t - \frac{1}{2} - u) du\right)$$

fra loro indipendenti, e che $j(u)$ sia il risultato di un modello epidemico.

Più esattamente supponiamo

$$j(t) = \beta(t) \int_0^t \phi(s) j(t-s)\, ds + \beta(t) j_0(t)$$

dove $j_0(t)$ è l'effetto al tempo t di coloro che erano già infetti al tempo 0 ed eventualmente dei casi importati dopo $t = 0$.

$$j(t) = \beta(t) \int_0^t \phi(s) j(t-s)\, ds + \beta(t) j_0(t) \tag{1}$$

(1) è l'equazione originale di Kermack-McKendrick con $S(t) \equiv N$ (*sembra ragionevole supporre che la frazione infettata finora sia sufficientemente bassa da non far variare sensibilmente* $S(t)$ ma $\beta(t)$ variabile.

$$j(t) = \beta(t) \int_0^t \phi(s) j(t-s)\, ds + \beta(t) j_0(t) \tag{1}$$

(1) è l'equazione originale di Kermack-McKendrick con $S(t) \equiv N$ (*sembra ragionevole supporre che la frazione infettata finora sia sufficientemente bassa da non far variare sensibilmente* $S(t)$ ma $\beta(t)$ variabile.

Supporremo che $\beta(t)$ sia una funzione spline parametrizzata dai valori $\{\beta_1, \dots, \beta_n\}$ in certi nodi temporali prefissati e che $j_0(t)$ sia fissato a meno di una costante moltiplicativa C.

$$j(t) = \beta(t) \int_0^t \phi(s) j(t-s)\, ds + \beta(t) j_0(t) \tag{1}$$

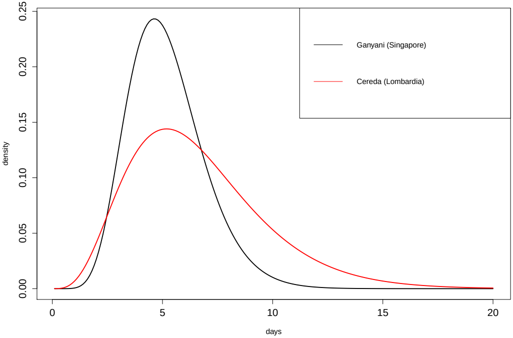
(1) è l'equazione originale di Kermack-McKendrick con $S(t) \equiv N$ (*sembra ragionevole supporre che la frazione infettata finora sia sufficientemente bassa da non far variare sensibilmente* $S(t)$ ma $\beta(t)$ variabile.

Supporremo che $\beta(t)$ sia una funzione spline parametrizzata dai valori $\{\beta_1, \ldots, \beta_n\}$ in certi nodi temporali prefissati e che $j_0(t)$ sia fissato a meno di una costante moltiplicativa C.

Chi è $\phi(s)$? Possiamo considerarla come la densità del 'tempo di generazione', ossia la differenza fra il tempo in cui un individuo è stato infettato e quando lo era stato infettato il suo 'infettore'.

Tempo di generazione o intervallo seriale

Esistono stime per l'intervallo seriale (stessa definizione del tempo di generazione, ma basato sulla comparsa dei sintomi)



Ho scelto quello stimato sulla base dei primi dati della Lombardia.

Funzione di verosimiglianza

$$\text{Da} \qquad j(t) = CF_t(\beta_1, \ldots, \beta_n)$$

$$H_t \sim \text{Poisson}\left(p_1 \int_0^{t-1} j(u) f_1(t - \frac{1}{2} - u) du\right)$$

$$D_t \sim \text{Poisson}\left(p_2 \int_0^{t-1} j(u) f_2(t - \frac{1}{2} - u) du\right)$$

abbiamo

$$\prod_{i=1}^{T} \mathbb{P}(H_t = h_t, D_t = d_t) = L(C_1, C_2, \beta_1, \ldots, \beta_n)$$

dove $C_1 = Cp_1$, $C_2 = Cp_2$, h_t e d_t i dati osservati.

$L(C_1, C_2, \beta_1, \ldots, \beta_n)$ è quindi la funzione di verosimiglianza che si può usare in vari metodi statistici.

Dettagli implementativi

In realtà H_t, il numero dei nuovi ospedalizzati nel giorno t non è noto, ma viene invece fornito CH_t, il numero delle persone ospedalizzate e R_t, quelle dimesse nel giorno t. Supponendo che una proporzione p dei morti D_t fosse ospedalizzata nel giorno precedente, si ricava

$$H_t = CH_t - CH_{t-1} + pD_t + R_t.$$

Ho utilizzato questa formula con $p = 0.8$ (scelta arbitraria).
Inoltre, data la variabilità giornaliera dei valori, ho regolarizzato i dati con una media mobile su 3 giorni.

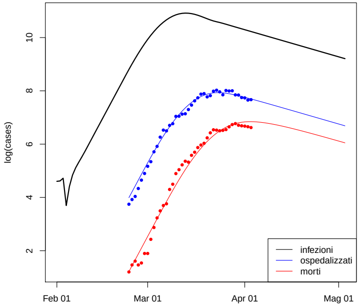
Per quanto riguarda le spline, finora ho usato 3 nodi temporali nei giorni

23 febbraio   7 marzo   23 marzo

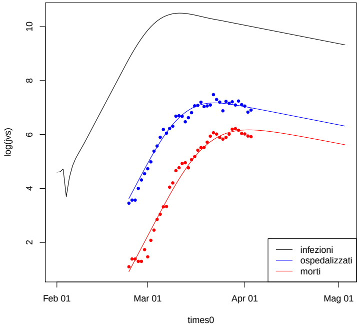
e ho supposto che β sia costante prima del primo nodo e dopo l'ultimo.
Infine ho scelto il 1 febbraio come $t = 0$.
Molte scelte sono arbitrarie e va verificata la sensitività dei risultati.

Risultati. Curve infezioni, ospedalizzati, morti

Italia

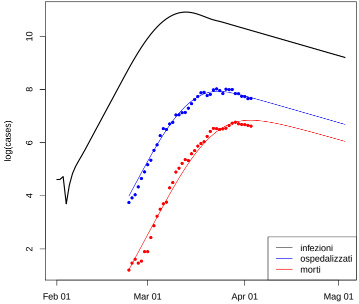


Lombardia

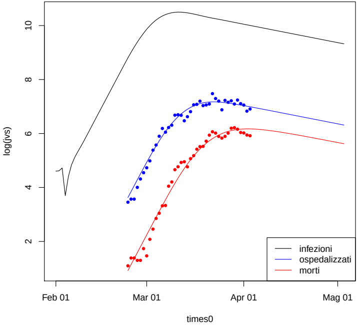


Risultati. Curve infezioni, ospedalizzati, morti

Italia

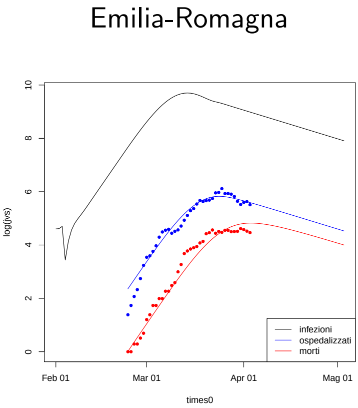


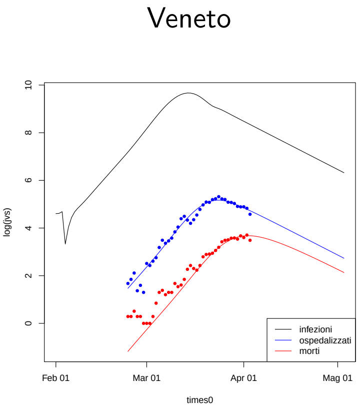
Lombardia



Massimo delle infezioni verso il 10 marzo; dei decessi in questi giorni. Notare che l'altezza della curva delle infezioni è arbitraria. Il metodo non permette di stabilire quale sia la frazione delle infezioni che dà luogo a ospedalizzazioni o morti.

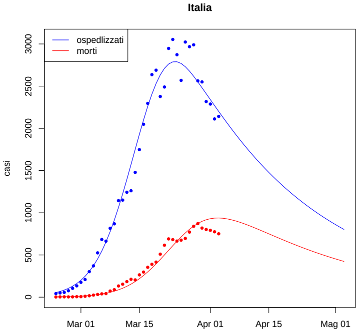
Risultati. Curve infezioni, ospedalizzati, morti. 2



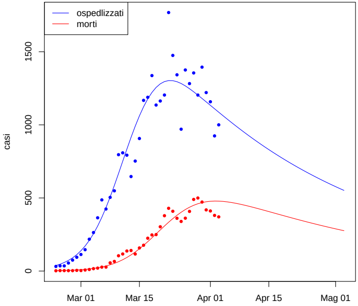


Risultati. Ospedalizzati, morti in scala naturale

Italia

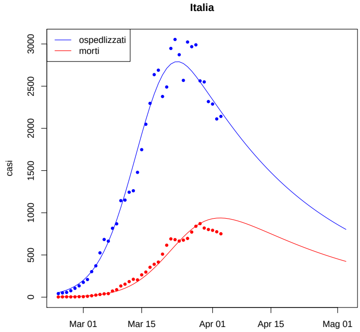


Lombardia

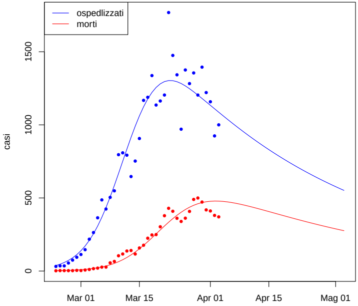


Risultati. Ospedalizzati, morti in scala naturale

Italia



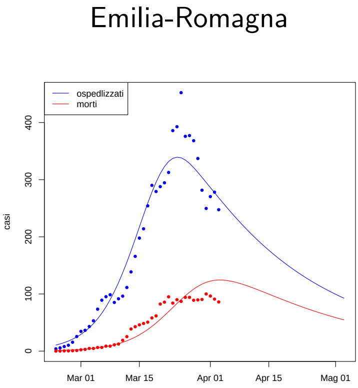
Lombardia

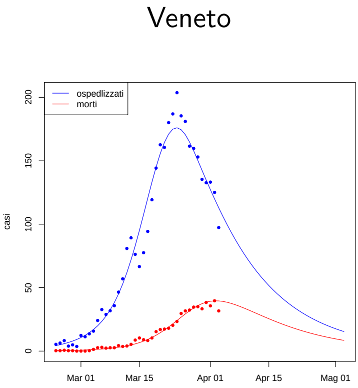


Forse

una tendenza a una diminuzione dei morti negli ultimi giorni più accentuata di quanto previsto dal modello.

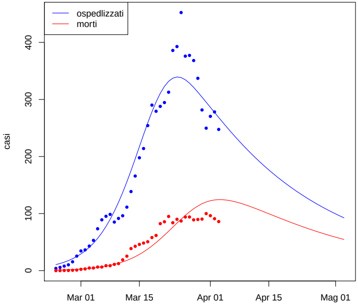
Risultati. Ospedalizzati, morti in scala naturale



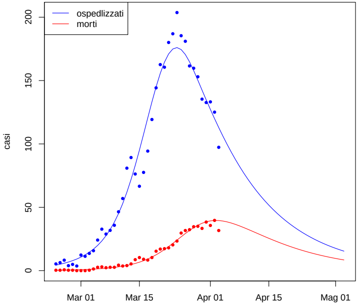


Risultati. Ospedalizzati, morti in scala naturale

Emilia-Romagna



Veneto

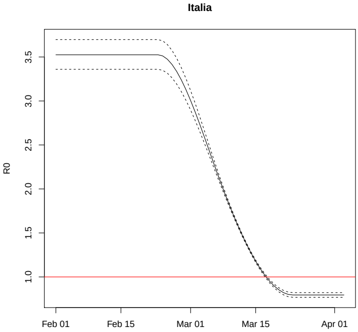


Anche qua il dubbio che il numero dei morti negli ultimi giorni sia minore di quanto previsto dal modello. Inoltre il picco degli ospedalizzati sembra più alto.

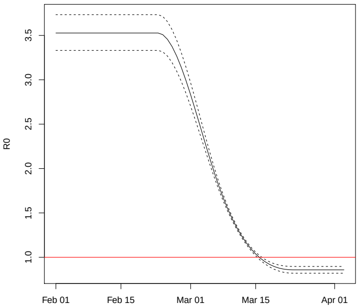
Risultati. Stima di R_0

Mostro il valore di R_0 giorno per giorno, calcolato come se i parameteri fossero fissati a quel valore:

Italia

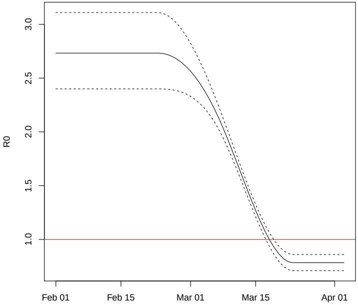


Lombardia

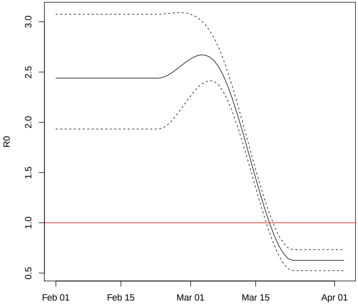


Risultati. Stima di R_0

Emilia-Romagna

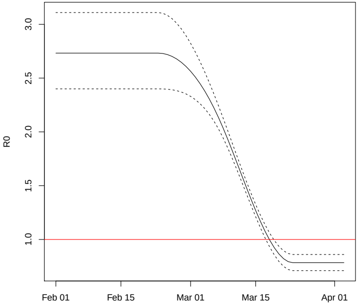


Veneto

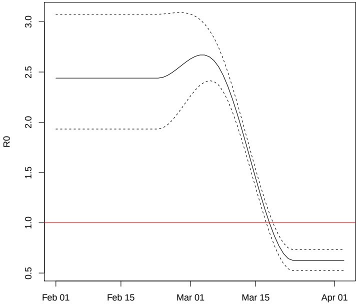


Risultati. Stima di R_0

Emilia-Romagna



Veneto



La crescita di R_0 in Veneto a inizio matrzo potrebbe essere un artefatto dovuto al fatto che il modello è stato fatto partire troppo presto per quella regione.

Intervalli di confidenza sui parametri

	Italia	Lombardia	Emilia-Romagna	Veneto
$R_0(T_1)$	3.36 - 3.70	3.33-3.73	2.40-3.11	1.93-3.07
$R_0(T_2)$	2.10-2.18	1.77-1.86	2.03-2.24	2.37-2.75
$R_0(T_3)$	0.77-0.82	0.82-0.90	0.71-0.86	0.52-0.73
prop. morti fra ospedalizzati[3]	0.37	0.395	0.404	0.258

[3] stima indiretta

Considerazioni finali

Si tratta come si vede di un lavoro provvisorio, che richiede numerosi approfondimenti e soprattutto analisi di sensitività.

Considerazioni finali

Si tratta come si vede di un lavoro provvisorio, che richiede numerosi approfondimenti e soprattutto analisi di sensitività.
La stima finale per R_0 è consistemente sotto 1, ma i valori stimati non sono molto sotto 1. Ciò avrebbe conseguenze poco piacevoli.

Considerazioni finali

Si tratta come si vede di un lavoro provvisorio, che richiede numerosi approfondimenti e soprattutto analisi di sensitività.

La stima finale per R_0 è consistemente sotto 1, ma i valori stimati non sono molto sotto 1. Ciò avrebbe conseguenze poco piacevoli.

Ci sono motivi per cui possa essere una sovrastima? *Forse*.

Considerazioni finali

Si tratta come si vede di un lavoro provvisorio, che richiede numerosi approfondimenti e soprattutto analisi di sensitività.
La stima finale per R_0 è consistemente sotto 1, ma i valori stimati non sono molto sotto 1. Ciò avrebbe conseguenze poco piacevoli.
Ci sono motivi per cui possa essere una sovrastima? *Forse*.

- È possibile che i dati in Lombardia sottostimino il picco (e quindi la velocità della discesa) a causa dello stress del sistema sanitario.

Considerazioni finali

Si tratta come si vede di un lavoro provvisorio, che richiede numerosi approfondimenti e soprattutto analisi di sensitività.

La stima finale per R_0 è consistemente sotto 1, ma i valori stimati non sono molto sotto 1. Ciò avrebbe conseguenze poco piacevoli.

Ci sono motivi per cui possa essere una sovrastima? *Forse*.

- È possibile che i dati in Lombardia sottostimino il picco (e quindi la velocità della discesa) a causa dello stress del sistema sanitario.
- È possibile che il 'lockdown' dia luogo a un transiente di trasmissioni all'interno delle famiglie, ma che poi andranno a esaurirsi.

Considerazioni finali

Si tratta come si vede di un lavoro provvisorio, che richiede numerosi approfondimenti e soprattutto analisi di sensitività.
La stima finale per R_0 è consistemente sotto 1, ma i valori stimati non sono molto sotto 1. Ciò avrebbe conseguenze poco piacevoli.
Ci sono motivi per cui possa essere una sovrastima? *Forse*.

- È possibile che i dati in Lombardia sottostimino il picco (e quindi la velocità della discesa) a causa dello stress del sistema sanitario.
- È possibile che il 'lockdown' dia luogo a un transiente di trasmissioni all'interno delle famiglie, ma che poi andranno a esaurirsi.
- Serve un modello più complesso per fare analisi di scenario. Questo può aiutare a intepretare i trend.