

Calcolo del numero di riproduzione effettivo R_t italiano nel tempo con metodo Bayesiano.

Max Pierini*

May 13, 2020

1 Introduzione

Si vuole calcolare il numero di riproduzione effettivo R nel tempo t dato il numero noto di nuovi casi k in t .

Si userà la seguente notazione:

- $X \in (a, b)$: $a < X < b$
- $X \in [a, b]$: $a \leq X \leq b$
- $X \in (a, b]$: $a < X \leq b$
- $X \in [a, b)$: $a \leq X < b$
- X_i : valore della variabile X al tempo i
 - R_t : valore di R in t
 - k_t : valore di k in t
- $P(X)$: probabilità **a priori** che si verifichi X
- $P(X|Y^*)$: probabilità **a posteriori** che si verifichi X noto il valore di Y^*
- $\mathcal{L}(Y^*|X)$: verosimiglianza (likelihood) del verificarsi noto di Y^* data la variabile aleatoria X
- $\mathbf{X}_{n \times m}$: matrice (o vettore) \mathbf{X} di n righe e m colonne
- $|\mathbf{X}|$: modulo del vettore \mathbf{X}
- $\mathcal{N}(\mu, \sigma)$: distribuzione normale (Gaussiana) avente media μ e deviazione standard σ
- $\mathcal{P}(X, Y)$: distribuzione di Poisson di X con parametro Y

1.1 Metodo

Il metodo utilizzato è una versione modificata da Kevin Systrom (2020) [1] di una soluzione presentata da Bettencourt e Ribeiro (2008) [2].

Si calcola la probabilità **a posteriori** del verificarsi di R_t dato k_t^* in t :

*info@maxpierini.it

$$P(R_t|k_t^*)$$

dove con k^* si indica che è l'evento noto (già accaduto).

Per il Teorema di Bayes:

$$P(A|B^*) = \frac{P(B^*|A) \cdot P(A)}{P(B^*)} \quad (1)$$

quindi

$$P(R_t|k_t^*) = \frac{P(k_t^*|R) \cdot P(R_t)}{P(k_t^*)} \quad (2)$$

sapendo che

$$P(k_t^*) = \sum_{i=0}^t \left(P(k_i|R) \cdot P(R_i) \right) \quad (3)$$

Quindi, per ogni t si calcolerà:

- $P(k_t^*|R)$: la probabilità **a posteriori** di aver osservato k_t^* nuovi casi dati i possibili valori di R , primo termine del numeratore nella (2)
- $P(R_t)$: la probabilità **a priori** dei possibili valori di R_t in t in base al valore calcolato del giorno precedente R_{t-1} , secondo termine del numeratore nella (2)
- $P(k_t^*)$: la probabilità **a priori** di osservare k_t^* nuovi casi in generale (pari alla sommatoria dei numeratori precedenti), denominatore (3) della (2)

2 Costanti

2.1 gamma

È noto che R_t dipende da γ , l'inverso dell'intervallo seriale T_{serial} ovvero il tempo medio tra casi successivi in una catena di trasmissione, spesso calcolato come intervallo tra presunta infezione e comparsa dei sintomi (periodo di contagiosità asintomatico). Per Covid-19 è stato stimato intorno ai 7.5 giorni [3]:

$$T_{serial} = 7.5 \text{ days}$$

$$\gamma = \frac{1}{T_{serial}} \quad (4)$$

$$\gamma \simeq 0.1333$$

2.2 vettori \mathbf{R}

Si crea un vettore $\mathbf{R}_{1 \times 1201}$ di tutti i possibili valori di R da R_{min} , pari a 0, ad R_{max} per il quale si può assumere il valore massimo 12 (figura 1).

$$R_{min} = 0, R_{max} = 12$$

$$\mathbf{R} = \{R_{min} \quad R_{min} + 0.01 \quad \dots \quad R_{max}\}_{1 \times 1201}$$

$$\mathbf{R} = \{0.00 \quad 0.01 \quad \dots \quad 11.99 \quad 12.00\}_{1 \times 1201}$$

$$|\mathbf{R}| = 1201$$

$$\mathbf{R}^T = \begin{Bmatrix} R_{min} \\ \vdots \\ R_{max} \end{Bmatrix}_{1201 \times 1}$$

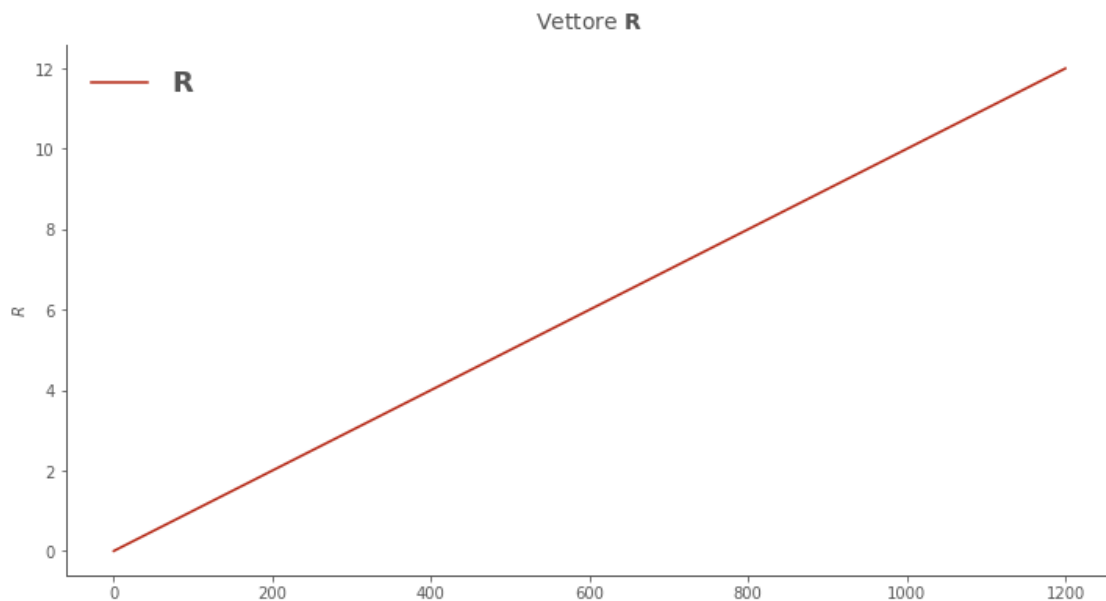


Figure 1: Vettore \mathbf{R} .

3 Valori noti

Si crea un vettore $\mathbf{k}_{t1 \times \omega}$ di tutti i k nuovi casi per giorno da $t = 0$ fino a oggi $t = \omega$ (figura 2). I dati italiani sono estratti dalla pubblicazione del Dipartimento di Protezione Civile [4].

$$t \in [0, \omega]$$

Nuovi casi al giorno k_t

$$\mathbf{k}_t = \{k_0 \quad k_1 \quad \dots \quad k_\omega\}_{1 \times \omega}$$

$$\mathbf{k}_t = \{143 \quad \dots \quad 1066\}_{1 \times 79}$$

$$|\mathbf{k}_t| = 79$$

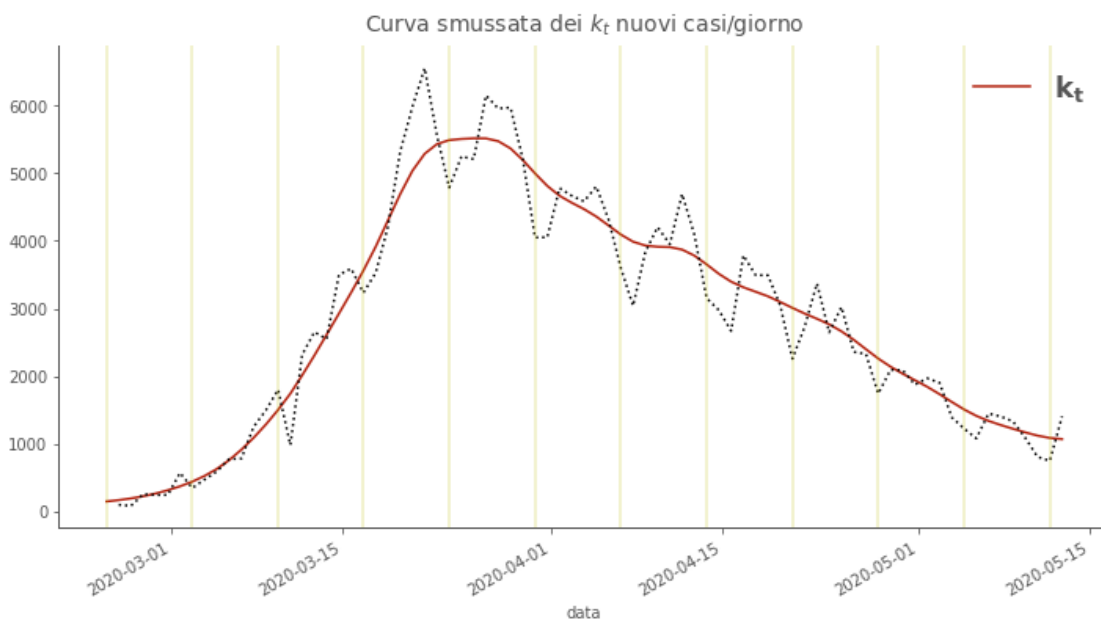


Figure 2: Curva smussata con metodo media mobile gaussiana ($\sigma = 2.5$) dei nuovi casi per giorno in Italia. I valori k_t formano il vettore $\mathbf{k}_{t \times \omega}$

4 $\mathbf{P}(\mathbf{k}|\mathbf{R})$

Per risolvere $P(k_t^*|\mathbf{R})$ si nota anzitutto che, data una λ serie di possibili k nuovi casi

$$P(k|\lambda) = \mathcal{P}(k, \lambda) = \frac{\lambda^k \cdot e^{-\lambda}}{k!} \quad (5)$$

ovvero, la probabilità **a posteriori** di avere k nuovi casi dati λ possibili valori k è una distribuzione di Poisson di k con parametro λ .

Sapendo però che il valore di k è noto (tutti i valori di k^* sono nel vettore \mathbf{k}_t) si può dire che

$$P(k^*|\lambda) \propto \mathcal{L}(\lambda|k^*) \quad (6)$$

Table 1: Estratto di matrice Λ : 1201×78

| | 2020-02-24 | 2020-02-25 | 2020-02-26 | ... | 2020-05-10 | 2020-05-11 | 2020-05-12 |
|-------------|------------|------------|------------|----------|------------|------------|------------|
| $R = 0.00$ | 125.15 | 142.65 | 164.53 | ... | 1019.58 | 977.57 | 947.81 |
| $R = 0.01$ | 125.32 | 142.84 | 164.75 | ... | 1020.94 | 978.87 | 949.08 |
| \vdots | \vdots | \vdots | \vdots | \vdots | \vdots | \vdots | \vdots |
| $R = 11.99$ | 619.04 | 705.62 | 813.85 | ... | 5043.27 | 4835.48 | 4688.29 |
| $R = 12.00$ | 619.87 | 706.57 | 814.94 | ... | 5050.00 | 4841.93 | 4694.55 |

ovvero, la probabilità di *aver avuto* k^* nuovi casi date λ possibilità è proporzionale alla verosimiglianza delle λ possibilità noto k^* .

Dalla letteratura, è nota la relazione tra λ , k_t ed R_t [2] derivante dalle equazioni differenziali del modello **SIR**:

$$\lambda = k_t \cdot \exp [\gamma(R_t - 1)] \quad (7)$$

Grazie a questa relazione è possibile dire che

$$P(k_t^*|R) = P(k_t^*|\lambda) \quad (8)$$

e costruire una matrice Λ di $\omega - 1$ colonne (perché il primo giorno è di outcome) e 1201 righe contenente per ogni giorno t tutti i possibili valori di λ dati i possibili 1201 valori di R (tabella 1 e figura 3).

$$\Lambda = \mathbf{k}_t \cdot \exp [\gamma(\mathbf{R}^T - 1)], \quad t = [0, \omega) \quad (9)$$

$$\Lambda = \left\{ \begin{array}{ccc} k_0 \cdot e^{\gamma(R_{min}-1)} & \dots & k_{\omega-1} \cdot e^{\gamma(R_{min}-1)} \\ \vdots & \vdots & \vdots \\ k_0 \cdot e^{\gamma(R_{max}-1)} & \dots & k_{\omega-1} \cdot e^{\gamma(R_{max}-1)} \end{array} \right\}_{1201 \times \omega-1}$$

Table 2: Estratto di matrice $\mathbf{P}(\mathbf{k}_t^* | \mathbf{R})$ (valori normalizzati): 1201×78

| | 2020-02-25 | 2020-02-26 | 2020-02-27 | ... | 2020-05-10 | 2020-05-11 | 2020-05-12 |
|-------------|------------|------------|------------|----------|------------|------------|------------|
| $R = 0.00$ | 3.67e-05 | 1.05e-05 | 1.69e-06 | ... | 1.95e-04 | 7.22e-05 | 1.46e-05 |
| $R = 0.01$ | 3.86e-05 | 1.11e-05 | 1.82e-06 | ... | 2.22e-04 | 8.30e-05 | 1.71e-05 |
| \vdots | \vdots | \vdots | \vdots | \vdots | \vdots | \vdots | \vdots |
| $R = 11.99$ | 1.73e-106 | 1.13e-119 | 9.51e-136 | ... | 0.00e+00 | 0.00e+00 | 0.00e+00 |
| $R = 12.00$ | 9.42e-107 | 5.64e-120 | 4.30e-136 | ... | 0.00e+00 | 0.00e+00 | 0.00e+00 |

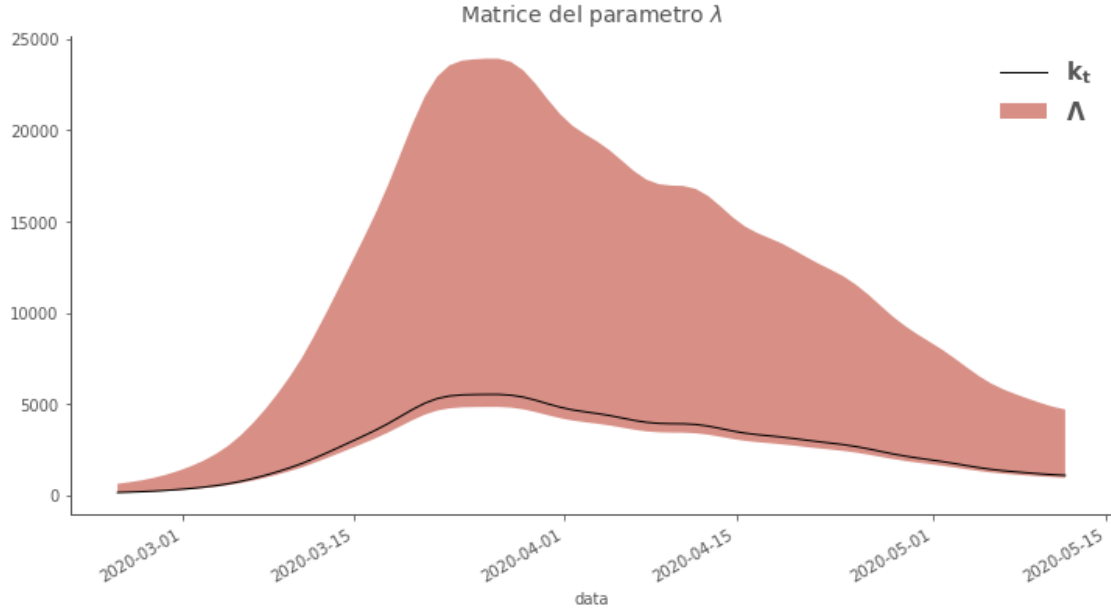


Figure 3: Valori del parametro λ , ovvero matrice $\Lambda_{1201 \times \omega-1}$ sulla serie temporale. Per ogni giorno i valori λ rappresentano tutti i possibili valori di k_t dato il range di R definito dal vettore $\mathbf{R} \in [R_{min}, R_{max}]$.

Noti i valori della matrice Λ si può dunque applicare la distribuzione di Poisson al vettore \mathbf{k}_t parametrizzata su R con $t \in (0, \omega]$ e ottenere così la matrice $\mathbf{P}(\mathbf{k}_t^* | \mathbf{R})$ di tutte le verosimiglianze dei possibili 1201 valori di R dati gli $\omega - 1$ valori di k_t^* (figura 4 e tabella 2):

$$P(\mathbf{k}_t^* | \Lambda) = \mathcal{P}(\mathbf{k}_t^*, \Lambda) = \frac{\Lambda^{\mathbf{k}_t^*} \cdot e^{-\Lambda}}{\mathbf{k}_t^*!} = \mathbf{P}(\mathbf{k}_t^* | \mathbf{R})_{1201 \times \omega-1} \quad (10)$$

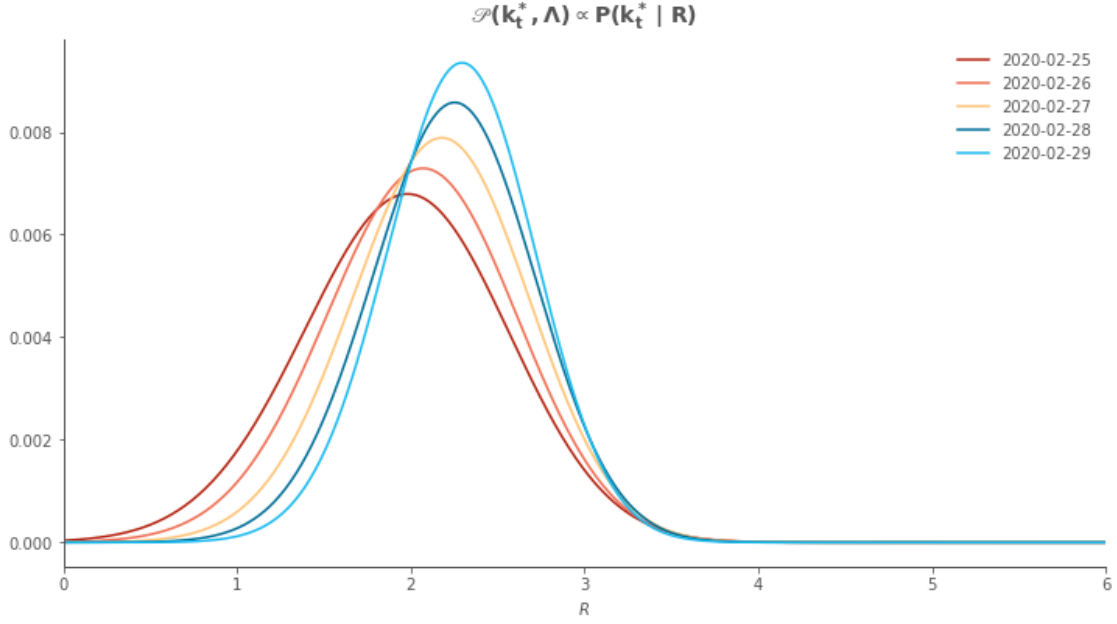


Figure 4: Verosimiglianze $\mathbf{P}(\mathbf{k}_t^* | \mathbf{R})$ dei primi 5 giorni di dati registrati in Italia. Valori normalizzati. Nella (2) non servirà normalizzare le verosimiglianze perché il denominatore stesso ha funzione di normalizzazione. Si nota come ogni giorno diminuisca la deviazione standard e dunque si guadagni progressivamente in confidenza.

5 $\mathbf{P}(\mathbf{R})$

La probabilità **a priori** dei possibili valori R può essere interpretata come una distribuzione normale (Gaussiana) avente media $\mu = R$ e deviazione standard σ che può essere stimata a 0.25.

Si può dunque creare una matrice $\mathbf{P}(\mathbf{R})$ di 1201×1201 i cui elementi $P(n)_m$ di riga n e colonna m sono la probabilità di $R = n$ con $\mu = m$ (tabella 3, figure 5 e 6):

$$\mathbf{P}(\mathbf{R}) = \mathcal{N}(\mathbf{R}, \sigma) = = \left\{ \begin{array}{ccccc} P(0)_0 & \cdots & P(0)_6 & \cdots & P(0)_{12} \\ \vdots & \ddots & \vdots & & \vdots \\ P(6)_0 & \cdots & P(6)_6 & \cdots & P(6)_{12} \\ \vdots & & \vdots & \ddots & \vdots \\ P(12)_0 & \cdots & P(12)_6 & \cdots & P(12)_{12} \end{array} \right\}_{1201 \times 1201} \quad (11)$$

Table 3: Estratto di matrice $\mathbf{P}(\mathbf{R})$: 1201×1201

| | $\mu = 0.00$ | $\mu = 0.01$ | $\mu = 0.02$ | \cdots | $\mu = 6.00$ | \cdots | $\mu = 11.98$ | $\mu = 11.99$ | $\mu = 12.00$ |
|-------------|--------------|--------------|--------------|----------|--------------|----------|---------------|---------------|---------------|
| $R = 0.00$ | 3.14e-02 | 3.04e-02 | 2.95e-02 | \cdots | 1.34e-127 | \cdots | 0.00e+00 | 0.00e+00 | 0.00e+00 |
| $R = 0.01$ | 3.14e-02 | 3.05e-02 | 2.95e-02 | \cdots | 3.49e-127 | \cdots | 0.00e+00 | 0.00e+00 | 0.00e+00 |
| \vdots | \vdots | \vdots | \vdots | \vdots | \vdots | \vdots | \vdots | \vdots | \vdots |
| $R = 6.00$ | 2.63e-127 | 6.66e-127 | 1.68e-126 | \cdots | 1.60e-02 | \cdots | 1.68e-126 | 6.66e-127 | 2.63e-127 |
| \vdots | \vdots | \vdots | \vdots | \vdots | \vdots | \vdots | \vdots | \vdots | \vdots |
| $R = 11.99$ | 0.00e+00 | 0.00e+00 | 0.00e+00 | \cdots | 3.49e-127 | \cdots | 2.95e-02 | 3.05e-02 | 3.14e-02 |
| $R = 12.00$ | 0.00e+00 | 0.00e+00 | 0.00e+00 | \cdots | 1.34e-127 | \cdots | 2.95e-02 | 3.04e-02 | 3.14e-02 |

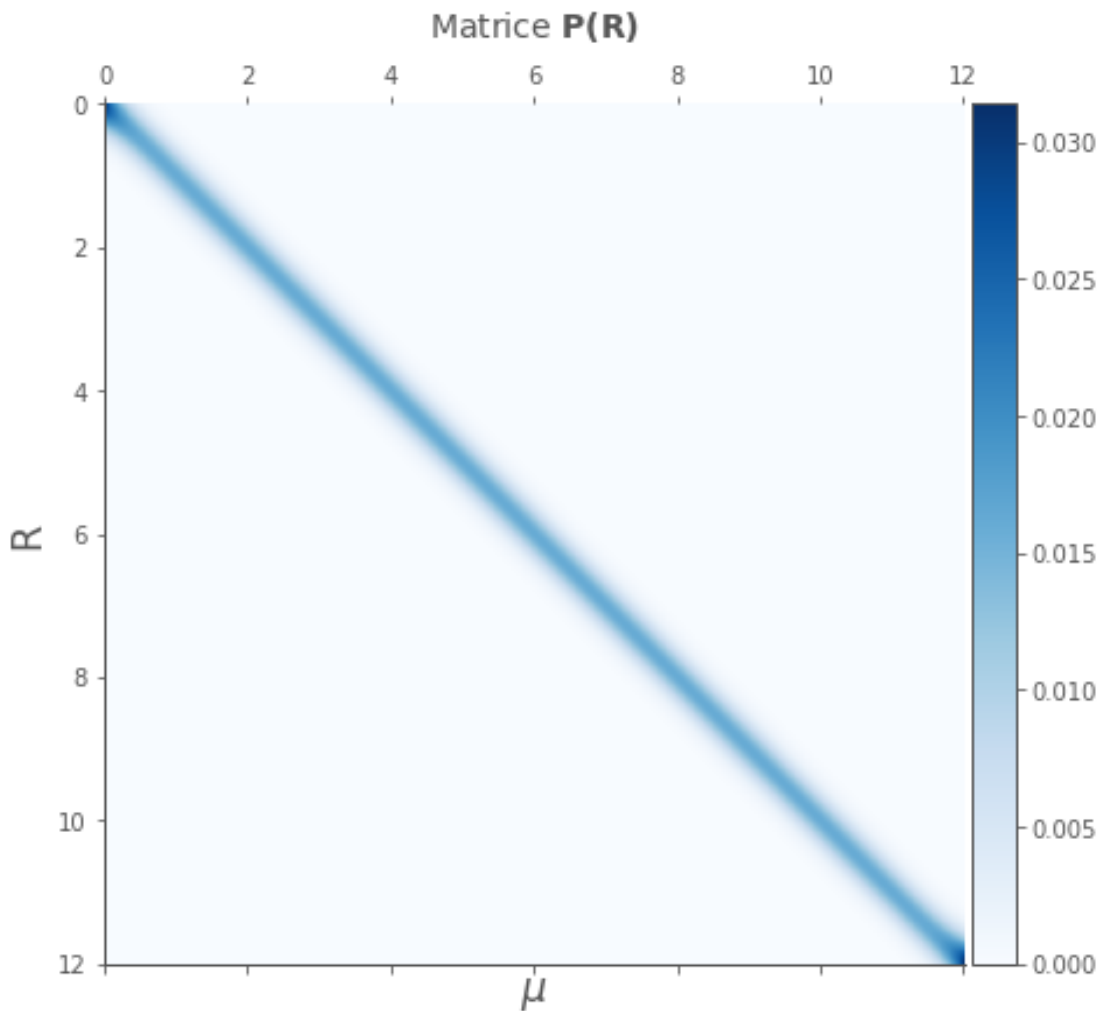


Figure 5: Matrice $\mathbf{P}(\mathbf{R})$. Il colore indica la densità di probabilità. Essendo gaussiana, è concentrata intorno a μ ovvero sulla diagonale della matrice.

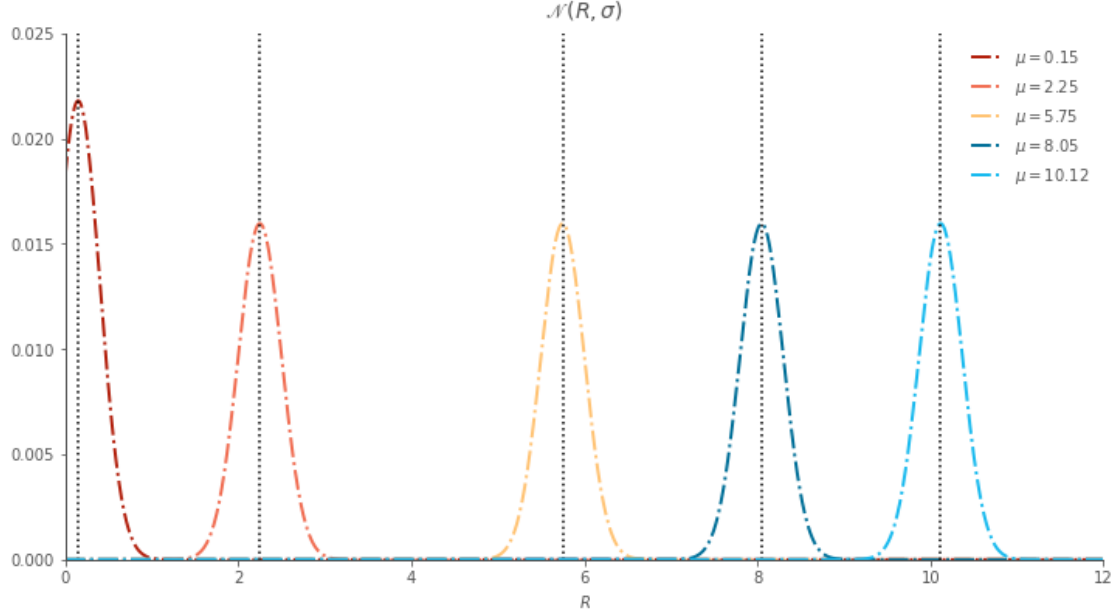


Figure 6: Distribuzione gaussiana di $P(R)$ con $\sigma = 0.25$. Esempio per alcuni valori di μ .

6 $\mathbf{P}(\mathbf{R}|\mathbf{k})$

È dunque possibile ora creare una matrice vuota $\mathbf{P}(\mathbf{R}_t | \mathbf{k}_t^*)$ di $1201 \times \omega$ i cui elementi $P(R_n)_m$ di riga n e colonna m saranno i valori della probabilità di $R = m$ al tempo $t = n$ (tabella 4).

$$\mathbf{P}(\mathbf{R}_t | \mathbf{k}_t^*) = \left\{ \begin{array}{cccc} P(R_0)_0 & P(R_1)_0 & \cdots & P(R_\omega)_0 \\ \vdots & \vdots & & \vdots \\ P(R_0)_6 & P(R_1)_6 & \cdots & P(R_\omega)_6 \\ \vdots & \vdots & & \vdots \\ P(R_0)_{12} & P(R_1)_{12} & \cdots & P(R_\omega)_{12} \end{array} \right\}_{1201 \times \omega} \quad (12)$$

Non avendo alcuna informazione precedente a $t = 0$, si imposta la probabilità iniziale della prima colonna $\mathbf{P}(\mathbf{R}_0 | \mathbf{k}_0^*)$ ovvero degli elementi $P(R_0)_m$ come costante in modo che $\sum P(R_0)_m = 1$:

$$P(R_0)_m = \frac{1}{|\mathbf{R}|} = \frac{1}{1201} = 8.33e - 04, \forall m \in [R_{min}, R_{max}]$$

Per ogni giorno $t \in (0, \omega]$ si calcola la probabilità di R_t basata sulla distribuzione normale del giorno precedente $t - 1$

Table 4: Estratto di matrice vuota $\mathbf{P}(\mathbf{R}_t|\mathbf{k}_t^*)$: 1201×79

| | 2020-02-24 | 2020-02-25 | 2020-02-26 | ... | 2020-05-10 | 2020-05-11 | 2020-05-12 |
|-------------|------------|------------|------------|----------|------------|------------|------------|
| $R = 0.00$ | 8.33e-04 | nan | nan | ... | nan | nan | nan |
| $R = 0.01$ | 8.33e-04 | nan | nan | ... | nan | nan | nan |
| \vdots | \vdots | \vdots | \vdots | \vdots | \vdots | \vdots | \vdots |
| $R = 11.99$ | 8.33e-04 | nan | nan | ... | nan | nan | nan |
| $R = 12.00$ | 8.33e-04 | nan | nan | ... | nan | nan | nan |

$$\begin{aligned}
\mathbf{P}(\mathbf{R}_t)_m &= \mathbf{P}(\mathbf{R})_{1201 \times 1201} \cdot \mathbf{P}(\mathbf{R}_{t-1})_m = \\
&= \begin{Bmatrix} P(0)_0 & \cdots & P(0)_{12} \\ \vdots & \ddots & \vdots \\ P(12)_0 & \cdots & P(12)_{12} \end{Bmatrix}_{1201 \times 1201} \cdot \begin{Bmatrix} P(R_{t-1})_0 \\ \vdots \\ P(R_{t-1})_{12} \end{Bmatrix}_{1201 \times 1} = \\
&= \begin{Bmatrix} P(R_t)_0 \\ \vdots \\ P(R_t)_{12} \end{Bmatrix}_{1201 \times 1}
\end{aligned} \tag{13}$$

e si moltiplica per $\mathbf{P}(\mathbf{k}_t^*|\mathbf{R})$ calcolato in precedenza, ottenendo così il numeratore della funzione (2):

$$\mathbf{P}(\mathbf{k}_t^*|\mathbf{R}) \cdot \mathbf{P}(\mathbf{R}_t)_m \tag{14}$$

Il denominatore $P(k_t^*)$ è la sommatoria dei numeratori precedenti:

$$P(k_t^*) = \sum_{i=0}^t \left(\mathbf{P}(\mathbf{k}_i^*|\mathbf{R}) \cdot \mathbf{P}(\mathbf{R}_i)_m \right) \tag{15}$$

ottenendo dunque per ogni giorno t la probabilità a posteriori di R_t dato k_t (tabella 5 e figure 7 e 8):

$$\mathbf{P}(\mathbf{R}_t|\mathbf{k}_t^*) = \left\{ \frac{\mathbf{P}(\mathbf{k}_t^*|\mathbf{R}) \cdot \mathbf{P}(\mathbf{R}_t)_m}{\sum_{i=0}^t \left(\mathbf{P}(\mathbf{k}_i^*|\mathbf{R}) \cdot \mathbf{P}(\mathbf{R}_i)_m \right)} \right\}_{1201 \times \omega} \tag{16}$$

Table 5: Estratto di matrice finale $\mathbf{P}(\mathbf{R}_t|\mathbf{k}_t^*)$: 1201×79

| | 2020-02-24 | 2020-02-25 | 2020-02-26 | ... | 2020-05-10 | 2020-05-11 | 2020-05-12 |
|-------------|------------|------------|------------|----------|------------|------------|------------|
| $R = 0.00$ | 8.33e-04 | 2.58e-05 | 1.61e-07 | ... | 2.64e-05 | 9.25e-06 | 1.28e-06 |
| $R = 0.01$ | 8.33e-04 | 2.78e-05 | 1.79e-07 | ... | 3.21e-05 | 1.14e-05 | 1.61e-06 |
| \vdots | \vdots | \vdots | \vdots | \vdots | \vdots | \vdots | \vdots |
| $R = 11.99$ | 8.33e-04 | 1.25e-106 | 3.07e-192 | ... | 0.00e+00 | 0.00e+00 | 0.00e+00 |
| $R = 12.00$ | 8.33e-04 | 6.61e-107 | 1.05e-192 | ... | 0.00e+00 | 0.00e+00 | 0.00e+00 |

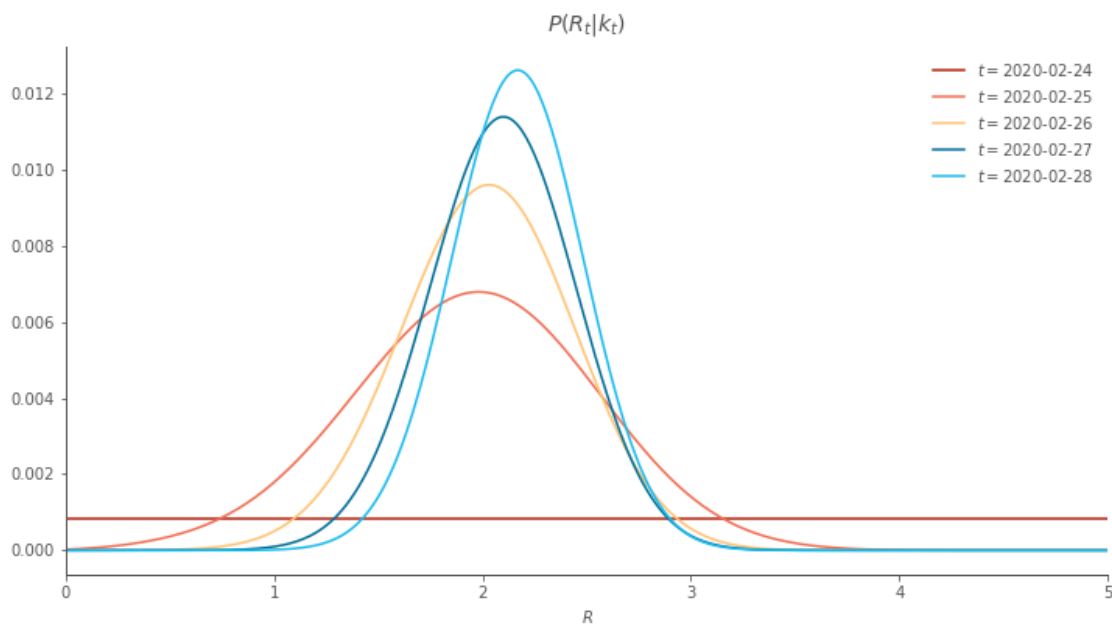


Figure 7: Distribuzione di probabilità $P(R_t|k_t)$ dei primi 5 giorni di dati in Italia. Si nota come nel primo giorno, non avendo alcuna informazione la probabilità è equamente distribuita e successivamente ogni giorno diminuisca la deviazione standard e dunque si guadagni progressivamente in confidenza sul valore di R_t .

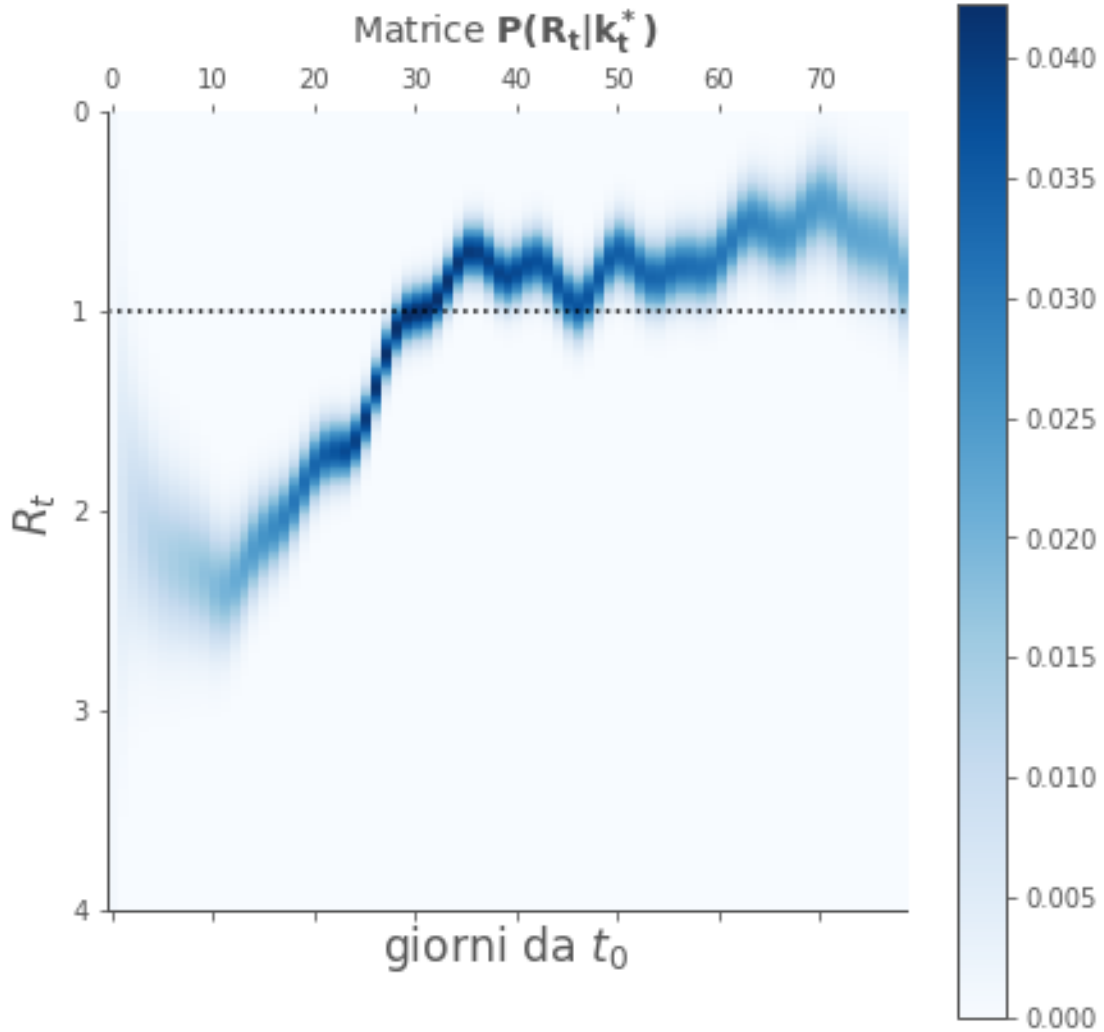


Figure 8: Matrice $\mathbf{P}(R_t | \mathbf{k}_t^*)$. All'asse y il valore di R_t è stato limitato graficamente a 4. Il colore indica la densità di probabilità.

7 Intervallo di credibilità

Data una distribuzione di probabilità $f(x)$ di posteriori Bayesiani della variabile aleatoria discreta x , l'intervallo di credibilità \mathbf{HDI}_p (High Density Interval) corrispondente alla regione a maggior densità di probabilità, scelto il valore di $p \in (0, 1)$ è

$$\mathbf{HDI}_p = [a, b] \in x, a < b : f(a) \simeq f(b), \int_a^b f(x) \simeq p \quad (17)$$

Esistono diversi metodi per il calcolo di \mathbf{HDI}_p ed i vantaggi/svantaggi di ciascuno dipendono dal tipo di studio effettuato [5].

In generale, laddove siano noti solo i valori $\{x, f(x)\}$ ma non la funzione di distribuzione, è possibile calcolare l'intervallo integrando con regola di Simpson a partire dal valore massimo di $f(x)$ fino a che

$$\int_a^b \{x, f(x)\} < p$$

ottenendo così il minor intervallo $[a, b]$ tale per cui l'area sottesa dalla curva sia “il più possibile” vicina a p .

Un metodo equivalente [1] consiste nel calcolare la probabilità cumulativa

$$F = \sum f(x) \quad (18)$$

ottenere la matrice quadrata antisimmetrica

$$\Delta \mathbf{F} = F - F^T = \begin{Bmatrix} F_0 - F_0 & F_1 - F_0 & \cdots & F_{|x|-1} - F_0 & F_{|x|} - F_0 \\ F_0 - F_1 & F_1 - F_1 & \cdots & F_{|x|-1} - F_1 & F_{|x|} - F_1 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ F_0 - F_{|x|-1} & F_1 - F_{|x|-1} & \cdots & F_{|x|-1} - F_{|x|-1} & F_{|x|} - F_{|x|-1} \\ F_0 - F_{|x|} & F_1 - F_{|x|} & \cdots & F_{|x|-1} - F_{|x|} & F_{|x|} - F_{|x|} \end{Bmatrix}_{|x| \times |x|} \quad (19)$$

e gli elementi $\Delta \mathbf{F}_{n,m} > p$ di riga n e colonna m .

Gli indici dell'elemento tale per cui è minore la differenza tra m (limite superiore) ed n (limite inferiore) rappresentano l'intervallo di credibilità, ovvero

$$\mathbf{HDI}_p = [n, m] \in x, n < m : \min(m - n), \forall n, m \in \Delta \mathbf{F}_{n,m} > p \quad (20)$$

Sia data ad esempio una distribuzione di probabilità con asimmetria positiva ($SK = 3, \mu = 7, \sigma = 6$) della variabile aleatoria discreta $x \in [0, 30]$ (figura 9). Fissando $p = .9$ l'intervallo di credibilità Bayesiano è dato dalla curva sottesa tra $x = 5$ e $x = 18$.

Utilizzando il metodo della matrice antisimmetrica, si calcola la probabilità cumulativa F e la matrice $\Delta \mathbf{F}$ (figura 10). Gli elementi $\Delta \mathbf{F}_{n,m} > .9$ sono delimitati dall'area blu in figura 10. L'elemento che soddisfa la (20) è $\Delta \mathbf{F}_{5,18} = 0.91$ (quadrato giallo in figura 10).

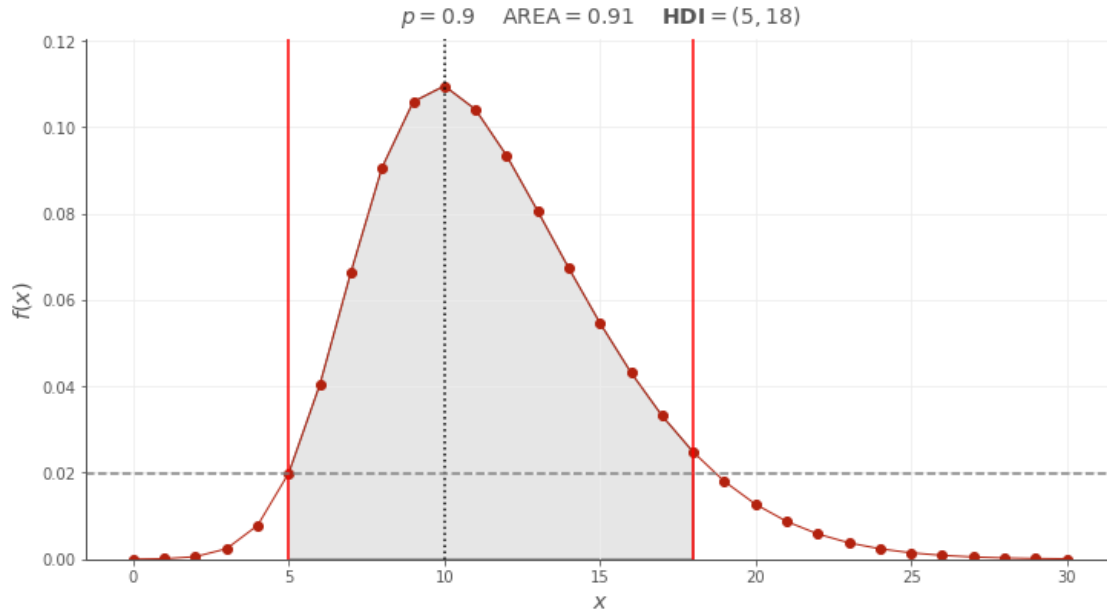


Figure 9: Esempio di distribuzione asimmetrica positiva e intervallo di credibilità $p > .9$.

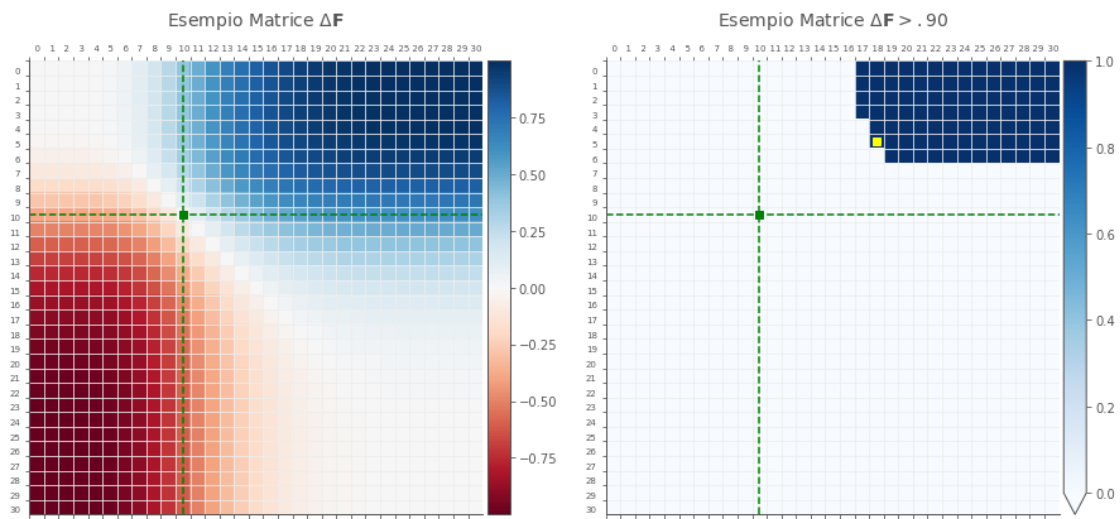


Figure 10: Esempio di matrice ΔF per $x \in [0, 30]$. Le linee verdi indicano la mediana. A sinistra la matrice completa, a destra la matrice per $\Delta F > .90$. Gli estratti di matrice corrispondenti alla regione di interesse sono le tabelle 6 e 7.

8 Grafico

Ottenuti i valori dell'intervallo di credibilità di R_t per ogni t (tabella 8 e figura 11) dalla distribuzione di probabilità della matrice $\mathbf{P}(\mathbf{R}_t | \mathbf{k}_t^*)$ si può quindi infine produrre un grafico dell'andamento temporale di R_t (figura 12).

Table 6: Estratto di matrice $\Delta\mathbf{F}$

| | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 |
|---|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|
| 0 | 0.87 | 0.90 | 0.93 | 0.95 | 0.97 | 0.98 | 0.99 | 0.99 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| 1 | 0.87 | 0.90 | 0.93 | 0.95 | 0.97 | 0.98 | 0.99 | 0.99 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| 2 | 0.87 | 0.90 | 0.93 | 0.95 | 0.97 | 0.98 | 0.99 | 0.99 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| 3 | 0.86 | 0.90 | 0.93 | 0.95 | 0.97 | 0.98 | 0.99 | 0.99 | 0.99 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| 4 | 0.86 | 0.90 | 0.93 | 0.95 | 0.96 | 0.97 | 0.98 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 |
| 5 | 0.85 | 0.88 | 0.91 | 0.93 | 0.95 | 0.96 | 0.97 | 0.97 | 0.98 | 0.98 | 0.98 | 0.98 | 0.98 | 0.98 | 0.98 |
| 6 | 0.82 | 0.86 | 0.88 | 0.91 | 0.92 | 0.93 | 0.94 | 0.94 | 0.95 | 0.95 | 0.95 | 0.95 | 0.95 | 0.95 | 0.95 |
| 7 | 0.76 | 0.80 | 0.83 | 0.85 | 0.87 | 0.88 | 0.89 | 0.89 | 0.89 | 0.89 | 0.90 | 0.90 | 0.90 | 0.90 | 0.90 |

Table 7: Estratto di matrice $\Delta\mathbf{F} > p$

| | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 |
|---|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| 0 | False | True | True | True | True | True | True | True | True | True | True | True | True | True | True |
| 1 | False | True | True | True | True | True | True | True | True | True | True | True | True | True | True |
| 2 | False | True | True | True | True | True | True | True | True | True | True | True | True | True | True |
| 3 | False | True | True | True | True | True | True | True | True | True | True | True | True | True | True |
| 4 | False | False | True | True | True | True | True | True | True | True | True | True | True | True | True |
| 5 | False | False | True | True | True | True | True | True | True | True | True | True | True | True | True |
| 6 | False | False | False | True | True | True | True | True | True | True | True | True | True | True | True |
| 7 | False | False | False | False | False | False | False | False | False | False | False | False | False | False | False |

Table 8: Estratto di tabella dei valori R_t : 79×3

| | Lo 90 | max | Hi 90 |
|------------|----------|----------|----------|
| 2020-02-24 | 0.00 | 0.00 | 10.81 |
| 2020-02-25 | 0.95 | 1.98 | 2.89 |
| 2020-02-26 | 1.32 | 2.03 | 2.69 |
| \vdots | \vdots | \vdots | \vdots |
| 2020-05-10 | 0.35 | 0.67 | 0.95 |
| 2020-05-11 | 0.40 | 0.73 | 1.01 |
| 2020-05-12 | 0.51 | 0.83 | 1.12 |

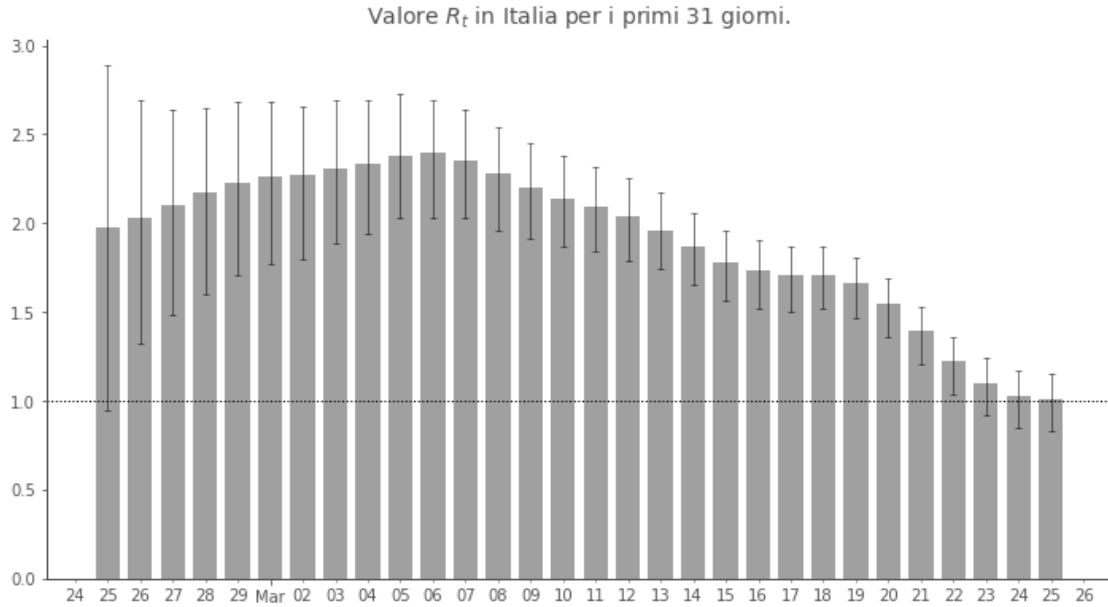


Figure 11: Valore R_t in Italia per i primi 31 giorni con intervallo di confidenza al 90%. Il primo giorno 2020-02-24, non avendo informazioni sui precedenti, serve solo da precedente per il calcolo della probabilità a posteriori del giorno successivo.

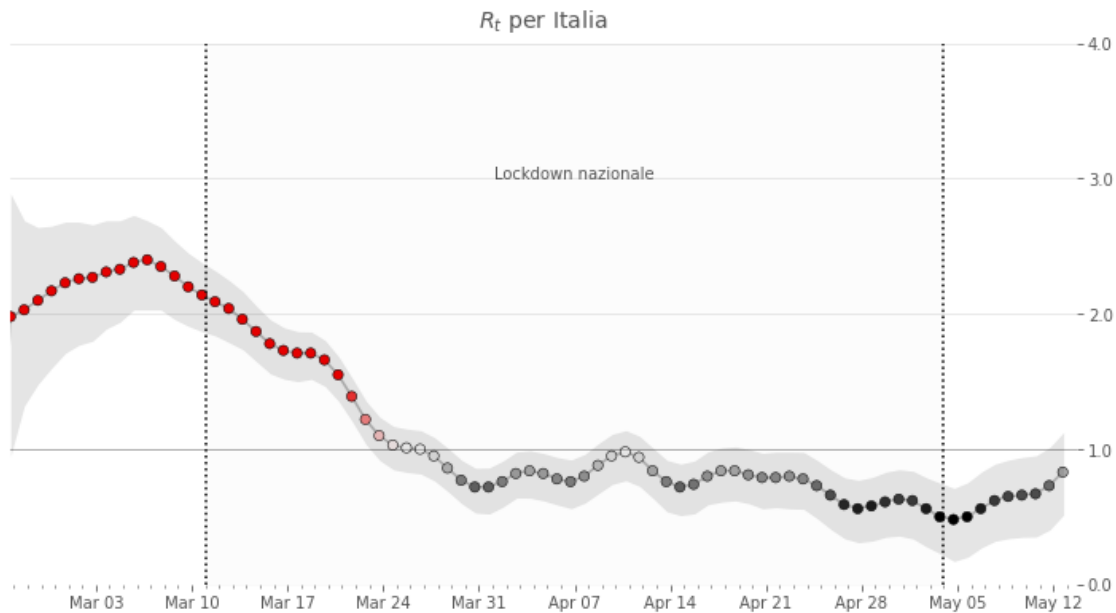


Figure 12: Valori di R_t in Italia con intervallo di credibilità al 90%. L'area ombreggiata delimitata dalle linee verticali indica il periodo di lockdown nazionale completo.

9 Conclusioni

È dunque possibile stimare il valore del numero di riproduzione effettivo R_t in base al numero noto di nuovi casi giornalieri k_t e al tempo seriale T_{serial} con metodo Bayesiano $P(R_t|k_t)$, utilizzando i risultati del giorno precedente come base per il calcolo della probabilità a posteriori, scegliendo l'intervallo di credibilità desiderato. È consigliabile smussare la curva dei nuovi casi (con media

mobile) per diminuire la variabilità dovuta ad errori e correzioni nella collezione giornaliera dei dati.

10 Codice Python

Si fornisce di seguito un esempio di codice Python per l'esecuzione del metodo illustrato per il calcolo di R_t in Italia, adattato dalla versione di Kevin Systrom [1].

```
1 import numpy as np
2 import pandas as pd
3 import scipy.stats as sps
4
5 # DEF: Calcolo nuovi casi e media mobile
6 def smooth(y, std=2):
7     dy = y.diff()
8     dy_smoothed = dy.rolling(7,
9         win_type="gaussian",
10        min_periods=1,
11        center=True).mean(std=std).round()
12     return dy_smoothed, dy
13
14 # DEF: Estrazione intervallo di credibilità
15 def highest_density_interval(pmf, p=.9):
16     if(isinstance(pmf, pd.DataFrame)):
17         return pd.DataFrame(
18             [highest_density_interval(
19                 pmf[col], p=p
20             ) for col in pmf],
21             index=pmf.columns)
22
23     cumsum = np.cumsum(pmf.values)
24     total_p = cumsum - cumsum[:, None]
25     lows, highs = (total_p > p).nonzero()
26     best = (highs - lows).argmin()
27     low = pmf.index[lows[best]]
28     high = pmf.index[highs[best]]
29
30     return pd.Series(
31         [low, high],
32         index=[
33             f"Low_{p*100:.0f}",
34             f"High_{p*100:.0f}"])
35
36 ##### CALCOLO #####
37
38 # Intervallo seriale e gamma
39 T_serial = 7.5
40 gamma = 1 / T_serial
41
42 # Vettori R
43 Rt_min = 0
44 Rt_max = 12
45 Rt_bra = np.linspace(Rt_min, Rt_max, Rt_max*100+1)
```

```

46 Rt_ket = Rt_bra[:, None]
47
48 # Caricamento dati
49 url_base = "https://raw.githubusercontent.com/"
50 url_repo = "pcm-dpc/COVID-19/master/dati-andamento-nazionale/"
51 url_file = "dpc-covid19-ita-andamento-nazionale.csv"
52 url_ita = url_base + url_repo + url_file
53 ita = pd.read_csv(
54     url_ita,
55     usecols=["data", "totale_casi"],
56     parse_dates=["data"],
57     index_col=["data"],
58     squeeze=True).sort_index()
59
60 # Calcolo nuovi casi e media mobile
61 k_bra, orig = smooth(ita, std=2.5)
62
63 # Matrice lambda
64 lam = k_bra[:-1].values * np.exp(gamma * (Rt_ket - 1))
65
66 # Matrice verosimiglianze P(k*/R)
67 likelihoods = pd.DataFrame(
68     data = sps.poisson.pmf(k_bra[1:].values, lam),
69     index = Rt_bra,
70     columns = k_bra.index[1:])
71
72 # Matrice P(R)
73 sigma = .25
74 process_matrix = sps.norm(
75     loc=Rt_bra,
76     scale=sigma
77 ).pdf(Rt_ket)
78 process_matrix /= process_matrix.sum(axis=0)
79
80 # Primo precedente della matrice P(R/k)
81 prior0 = np.ones_like(Rt_bra)/len(Rt_bra)
82 prior0 /= prior0.sum()
83
84 # Matrice vuota P(R/k)
85 posteriors = pd.DataFrame(
86     index=Rt_bra,
87     columns=k_bra.index,
88     data={k_bra.index[0]: prior0}
89 )
90
91 # Calcolo P(R/k*) giorno per giorno
92 for previous_day, current_day in zip(k_bra.index[:-1], k_bra.index[1:]):
93     # Calcolo del precedente P(R_t) = P(R) * P(R_t-1)
94     current_prior = process_matrix @ posteriors[previous_day]
95     # Calcolo numeratore P(k*/R)P(R_t) [equazione (2)]
96     numerator = likelihoods[current_day] * current_prior
97     # Calcolo denominatore P(k*) [equazione (2) e (3)]
98     denominator = np.sum(numerator)
99     # Calcolo della probabilità a posteriori P(k*/R_t) [equazione (2)]

```

```
100     posteriors[current_day] = numerator/denominator
101
102     # Estrazione valori per intervallo di credibilità
103     hdis = highest_density_interval(posteriors, p=.9)
104
105     # Estrazione massima probabilità
106     most_likely = posteriors.idxmax().rename("ML")
107
108     # Tabella valori di R in t
109     result = pd.concat([most_likely, hdis], axis=1)
```

References

- [1] K. System, “Estimating covid-19’s rt in real-time,” *GitHub*. [Online]. Available: <https://github.com/k-sys/covid-19/blob/master/RealtimeR0.ipynb>
- [2] L. M. A. Bettencourt and R. M. Ribeiro, “Real time bayesian estimation of the epidemic potential of emerging infectious diseases,” *PLoS ONE*, vol. 3, no. 5, 2008. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2366072/>
- [3] S. Sanche, Y. T. Lin, C. Xu, E. Romero-Severson, N. Hengartner, and R. Ke, “High contagiousness and rapid spread of severe acute respiratory syndrome coronavirus 2.” *Emerging infectious diseases*, vol. 26, no. 7, 2020.
- [4] D. di Protezione Civile, “Dpc covid-19 github repository,” *GitHub*, 2020. [Online]. Available: <https://github.com/pcm-dpc/COVID-19>
- [5] R. J. Hyndman, “Computing and graphing highest density regions,” *The American Statistician*, vol. 50, no. 2, pp. 120–126, 1996.