実効再生産数の計算

中村 崇宏 (高校数学科)

1　序

実効再生産数は，1 人の感染者が他の何人に病気を拡げるかを示す数値である。基本再生産数が生物学的要因だけで決まる定数であるのに対し，実効再生産数は感染対策などの社会的要因にも依存する変動量である。実効再生産数が 1 より大きいとき，感染者数が指数関数的に増大するため，実効再生産数を 1 以下に抑えることが感染対策の目標となる。国や自治体が政策を決定する際，実効再生産数は重要な指標となるため，政策の科学的根拠を理解するには，実効再生産数の計算方法を知っていなければならない。

参考文献 [1] のインターネットサイトでは，北海道，京都府，大阪府における新型コロナウィルス感染の実効再生産数を，参考文献 [2] のソフトウェアを用いて計算している。このソフトウェアは Microsoft Excel (VB 言語) で作成されている。本稿はこれを Python 言語に翻訳し，参考文献 [1] の計算結果を再現することを試みる。

2　実効再生産数

計算に用いるデータは，1 日あたりの感染者数 I_t である。ただし添字の t は日付を表す整数 $(t = 0, 1, 2, \cdots)$ である。時刻 t における感染者のうち，s 日前の感染者 1 人あたりが再生産した者の人数を $\beta_{t,s}$ とすると，

$$I_t = \sum_{s=1}^{t} \beta_{t,s}\, I_{t-s} \tag{1}$$

が成り立つ。ここで $\beta_{t,s}$ が

$$\beta_{t,s} = w_s\, R_t \tag{2}$$

のように積で表されると仮定する (参考文献 [3])。式 (1)(2) から

$$R_t = I_t \Big/ \sum_{s=1}^{t} w_s\, I_{t-s} \tag{3}$$

である。式 (2) の w_s が

$$\sum_{s=1}^{\infty} w_s = 1 \tag{4}$$

を満たすように w_s を規格化するとき，式 (3) の R_t は時刻 t における実効再生産数を表す。式 (4) の w_s は，感染から再生産までの日数 s の確率分布を表す。

2.1　感染から再生産までの日数の確率分布

感染者が病気を拡げる過程をベルヌーイ試行とみなせば，単位時間あたりの再生産数はポワソン分布に従う。よって感染から再生産までの時間 s は指数分布に従うことになる。ここで s の平均 $E(s)$ と標準偏差 $\sigma(s)$ を指定できるように，s は指数分布を一般化したガンマ分布

$$s \sim f(s-1; a, b), \quad f(x; a, b) = \frac{x^{a-1} e^{-x/b}}{b^a\, \Gamma(a)} \tag{5}$$

$$a = \left(\frac{E(s) - 1}{\sigma(s)}\right)^2, \quad b = \frac{\sigma(s)^2}{E(s) - 1} \tag{6}$$

に従うものとする。ただし Γ はガンマ関数を表す。参考文献 [4] によると，新型コロナウィルスの場合は $E(s) = 6.3$ 日，$\sigma(s) = 4.2$ 日である。

式 (5) で $s-1$ がガンマ分布に従うとしているのは，感染から再生産までの日数は 1 以上の数値として集計されるからである。実際には，式 (3) の s は自然数の離散変数であるため，式 (5) を離散化して

$$w_s = \int_{s-1}^{s+1} (1 - |x - s|)\, f(x-1; a, b)\, dx \tag{7}$$

を用いる。式 (7) の意図は次の通りである (参考文献 [2] の補遺 §11): 感染者数 I_t は 1 日毎に集計するため，正確な感染時刻には 24 時間の幅がある。よって，感染時刻 t_1 と再生産時刻 t_2 の差 $x = t_2 - t_1$ には 48 時間の幅がある。すなわち，この差が s 日として集計されるような x は区間 $[s-1, s+1]$ の中にある。t_1 と t_2 の分布がともに一様であるとすると，x の分布は平均が s の三角分布 $1 - |x - s|$ であるので，式 (7) は $[s-1, s+1]$ の範囲で f と三角分布を畳み込んでいる。式 (5) を式 (7) に代入して積分を実行すると

$$w_s = \Delta^2 \{x\, F(x; a, b) - ab\, F(x; a+1, b)\}|_{x=s-1} \tag{8}$$

$$\Delta^2 g(x) = g(x+1) - 2\, g(x) + g(x-1) \tag{9}$$

$$F(x; a, b) = \int_0^x f(x'; a, b)\, dx' \tag{10}$$

プログラム 1

```
import numpy as np
from scipy.stats import gamma

def si_distr(N, si_mean, si_std):
  """ serial interval distribution """
  a = ((si_mean-1)/si_std)**2
  b = (si_mean-1)/a
  k = np.arange(-2,N)
  return np.diff(k*gamma.cdf(k,a,0,b)
            - a*b*gamma.cdf(k,a+1,0,b), 2)

def Reff(data, si_mean, si_std,
        tau=7, conf=0.95, mu=5):
  """ effective reproduction number """
  N = len(data)
  w = si_distr(N, si_mean, si_std)
  L = np.convolve(data, w)[:N]
  u = np.ones(tau)
  a = 1 + np.convolve(data, u)[:N]
  b = mu/(1 + mu*np.convolve(L, u)[:N])
  return np.vstack([gamma.median(a,0,b),
            gamma.interval(conf,a,0,b)])
```

となる。ただし Δ^2 は 2 階差分演算子，$F(x;a,b)$ は累積ガンマ分布関数である。式 (8) の w_s は式 (4) を満たすことに注意する。

プログラム 1 の関数 si_distr は，式 (8) の w_s を計算する。プログラム 2 は w_s のヒストグラムを描く。プログラム 2 の実行結果を図 1 に示す。図 1 では $E(s) = 6.3$ 日，$\sigma(s) = 4.2$ 日とした (参考文献 [4])。

2.2 実行再生産数の確率分布

本節ではベイズの定理を用いて，時刻 t における実行再生産数 R_t の確率分布を求める (参考文献 [2] の補遺 §1)。R_t の値が与えられているとき，1 日あたりの感染者数 I_t はポワソン分布

$$p(I_t|R_t) = \frac{E(I_t)^{I_t} e^{-E(I_t)}}{I_t!} \tag{11}$$

に従うと仮定する。ただし I_t の平均 $E(I_t)$ は，式 (3) から

$$E(I_t) = R_t L_t, \quad L_t = \sum_{s=1}^{t} w_s I_{t-s} \tag{12}$$

である。ここで R_t の t による変動を滑らかにするため，τ 日間 (例えば $\tau = 7$) の R_t の平均値 $R_{t,\tau}$ を考える。すなわち，時刻 $t-\tau+1$ から時刻 t までの τ 日間に，R_t は一定値 $R_{t,\tau}$ をとるものとみなす。この τ 日間の感染者数

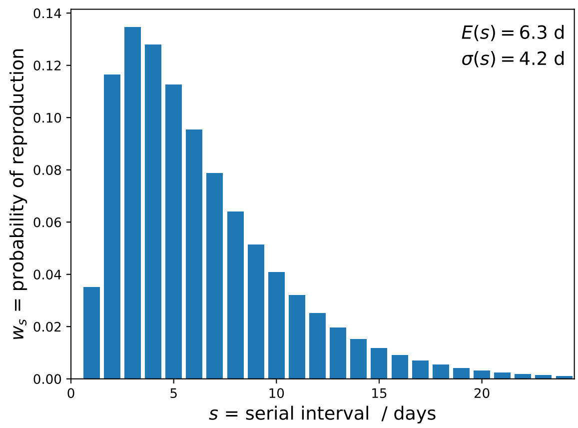
$$I_{t,\tau} = (I_{t-\tau+1}, \cdots, I_t) \tag{13}$$

プログラム 2

```
import matplotlib.pyplot as plt
from program1 import si_distr

w = si_distr(25, 6.3, 4.2)
plt.bar(range(25), w)
plt.show()
```

図 1



の同時分布関数を $p(I_{t,\tau}|R_{t,\tau})$ と書く。$I_{t,\tau}$ の成分は独立な確率変数であると仮定すると，式 (11) から，

$$p(I_{t,\tau}|R_{t,\tau}) = \prod_{s=0}^{\tau-1} \frac{(R_{t,\tau}L_{t-s})^{I_{t-s}} e^{-R_{t,\tau}L_{t-s}}}{I_{t-s}!} \tag{14}$$

である。τ 日間の開始前における $R_{t,\tau}$ の確率分布関数を $p(R_{t,\tau})$ とすると，ベイズの定理から，

$$p(R_{t,\tau}|I_{t,\tau}) = \frac{p(I_{t,\tau}|R_{t,\tau})\, p(R_{t,\tau})}{p(I_{t,\tau})} \tag{15}$$

が成り立つ。式 (15) 左辺の $p(R_{t,\tau}|I_{t,\tau})$ は，τ 日間のデータ $I_{t,\tau}$ を考慮した後の実行再生産数の確率分布を表す。一方，式 (15) 右辺の $p(R_{t,\tau})$ は，データがまだない状態での実行再生産数の不確実性を表している。$p(R_{t,\tau})$ は平均が μ の指数分布関数

$$p(R_{t,\tau}) = (1/\mu)\, e^{-R_{t,\tau}/\mu} \tag{16}$$

であると仮定して，式 (14)(16) を式 (15) に代入すると，$p(R_{t,\tau}|I_{t,\tau})$ は

$$p(R_{t,\tau}|I_{t,\tau}) \propto {R_{t,\tau}}^{a-1}\, e^{-R_{t,\tau}/b} \tag{17}$$

に比例する。ただし

$$a = 1 + \sum_{s=0}^{\tau-1} I_{t-s}, \quad b = \mu \Big/ \left(1 + \mu \sum_{s=0}^{\tau-1} L_{t-s}\right) \tag{18}$$

とおいた。式 (17) では $R_{t,\tau}$ に依存しない因子を省略できるため，式 (15) の分母は計算しなくてよい。

プログラム 3

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from program1 import Reff

d = pd.read_csv('COVID-19.csv', low_memory=0)
d = d[d[' 居住都道府県'] == ' 千葉県']
d = np.array(pd.to_datetime(d[' 確定日']))
t = np.arange('2020-02-27',
              '2020-06-01', dtype='datetime64')
c = np.count_nonzero(d[:,None] == t, axis=0)
R = Reff(c, 6.3, 4.2)
plt.subplot(211); plt.bar(t.astype('float'), c)
plt.subplot(212); plt.semilogy(t, R.T)
plt.show()
```

$p(R_{t,\tau}|I_{t,\tau})$ の全確率が 1 となるように式 (17) を規格化すると,

$$p(R_{t,\tau}|I_{t,\tau}) = f(R_{t,\tau}; a, b) \tag{19}$$

となる。ただし f は式 (5) のガンマ分布関数である。以下では，式 (16) の μ は時刻 t によらない定数であるとする。式 (19) の分布関数の標準偏差 $b\sqrt{a}$ と平均 ab の比 $1/\sqrt{a}$ が十分に小さくなるためには，τ 日間の感染者数の合計 $\sum_{s=0}^{\tau-1} I_{t-s}$ が十分に大きいことが必要である (参考文献 [2] の補遺 §2)。

3　計算例

プログラム 1 の関数 Reff は，式 (19) の $p(R_{t,\tau}|I_{t,\tau})$ を計算し，$R_{t,\tau}$ の中央値と信頼区間を出力する。プログラム 3 は，感染者のデータファイル COVID-19.csv を読み込み，$R_{t,\tau}$ のグラフを描く。このデータファイルは参考文献 [5] からダウンロードできる。プログラム 3 の実行結果を図 2 に示す。プログラム 3 の「千葉県」の部分を様々に変えると図 3〜6 を得る。図 7〜9 では，それぞれ参考文献 [6]〜[8] からダウンロードしたデータを用いた。図 2〜9 では $E(s) = 6.3$ 日，$\sigma(s) = 4.2$ 日，$\tau = 7$ 日，$\mu = 5$ とした。これらの図の横軸は 2020 年の日付である。各図の上段は 1 日あたりの感染者数のヒストグラムであり，下段の赤線と破線はそれぞれ，実行再生産数の確率分布の中央値，信頼度 95%の区間を表す。下段の縦軸は対数目盛であることに注意する。図 2〜9 の上段と下段を比較すると，感染者数の増減に応じて実行再生産数が 1 の上下を変動することが分かる。図 4〜6 は参考文献 [1] の計算結果と一致している。

図 1〜9 の作成に用いたコンピュータープログラムを参考文献 [9] に載せた。(2020 年 6 月 20 日投稿)

図 2: 千葉県

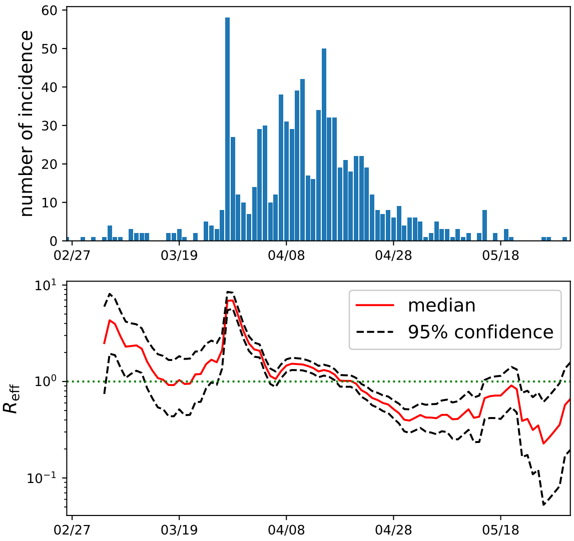
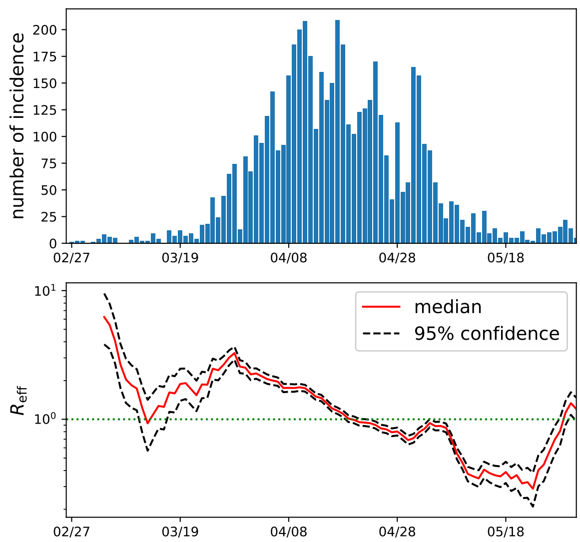


図 3: 東京都



参考文献

[1] www.covid19-yamanaka.com/cont3/16.html

[2] A. Cori, et al., American Journal of Epidemiology 178 (2013) 1505

[3] C. Fraser, PLoS ONE 2(8): e758 (2007)

[4] Q. Bi, et al.
The Lancet Infectious Diseases: April 27 (2020)

[5] gis.jag-japan.com/covid19jp/

[6] github.com/nychealth/coronavirus-data

[7] data.london.gov.uk/dataset

[8] www.rki.de

[9] github.com/tt-nakamura/Reff

図 4: 北海道

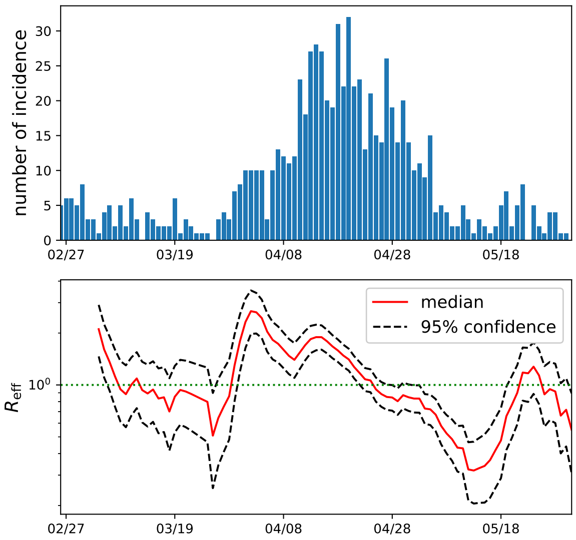


図 5: 京都府

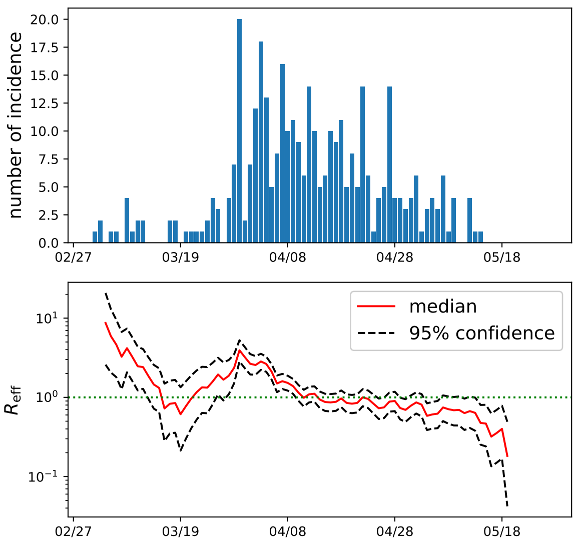


図 6: 大阪府

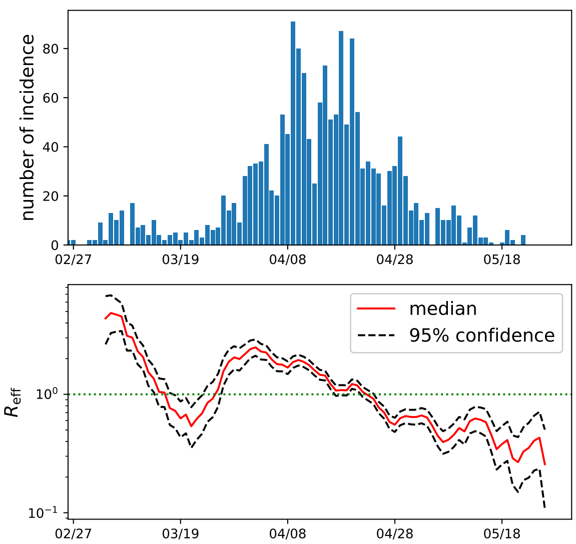


図 7: ニューヨーク市

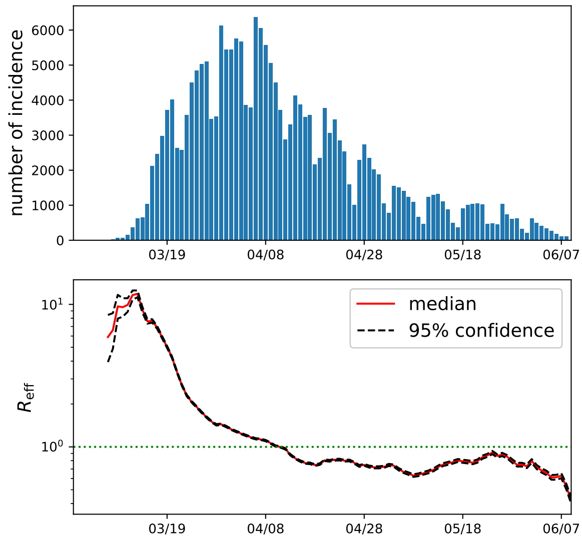


図 8: ロンドン市

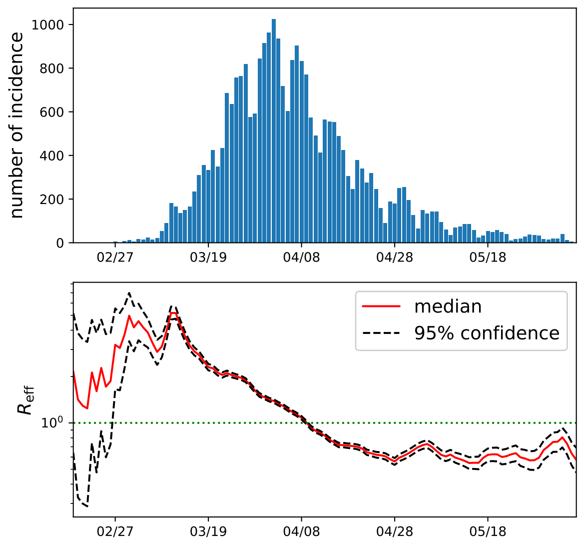


図 9: ベルリン市