

感染の発症間隔分布を考慮した SiR モデル

S. Yamauchi

2021 年 2 月 5 日 (誤記訂正)

目次

1	はじめに	1
2	COVID-19 の発症間隔データ	1
3	感染力が感染時間に依存するモデル	2
4	特性曲線による $i(t, a)$ の解	4
5	離散化と数値解	5
6	発症間隔分布関数 $f_{\beta}(a)$ について	6

1 はじめに

COVID-19 の発症間隔は、当初は SARS などの発症間隔をもとに 7~10 日程度と見積もられることが多かったが、最近の報告では、かなり短く、また症状出現前に他者へ感染する割合が多く、また無症状者で感染を広げる等が指摘されている。多く引用されている西浦博氏の調査では、発症間隔は中央値 4.6 日、平均値 4.8 日、標準偏差 2.3 日のワイブル分布をなすとされている [1]。これを SIR モデルまたは SEIR モデルでは分布形が異なるし、また平均値と標準偏差を同時に上記の値に合わせることができない (平均値を合わせると標準偏差が大きくなる) ので、この発症間隔分布を直接用いるモデル [2] を検討する。

2 COVID-19 の発症間隔データ

日本の新型コロナウイルス感染の調査結果 (2 月 12 日まで) をもとに、西浦博氏が 2 月下旬に COVID-19 の発症間隔について、表 1 のような推測値を発表している [1]。1 次-2 次の感染ペアが確認できたものが 28 ペアで、そのうち組み合わせの確実なものが 18 ペアあり、両者それぞれに対してその分布、中央値、平均値、標準偏差を推定している。「CrI」は信頼性区間の意味である。「WAIC」は分布の適合度を表し、小さいほど適合しているようであるが、詳細はわからない。「右側切り捨て補正」(right truncation) とは、採用データは途中で打ち切られているため、発症間隔の長いデータを除外する結果になることの補正のようであり、補正は行う必要があると思える。

表 1 西浦博氏の推測結果

	全 28 データ		确实 18 データ	
	右側切り捨て補正	補正なし	右側切り捨て補正	補正なし
分布 (WAIC)	対数正規 (220)	対数正規 (120)	ワイブル	対数正規
中央値 (95% CrI)	4.0 (3.1, 4.9)	3.9 (3.1, 4.8)	4.6 (3.6, 5.9)	4.1 (3.2, 5.0)
平均値 (95% CrI)	4.7 (3.7, 6.0)		4.8 (3.8, 6.1)	
標準偏差 (95% CrI)	2.9 (1.9, 4.9)		2.3 (1.6, 3.5)	

以上より、确实性の高いデータ 18 ペアを用いて右側切り捨て補正を行った推測結果、中央値 4.6 日、平均値 4.8 日、標準偏差 2.3 日のワイブル分布を発症間隔分布として採用するのが良いと思える。

3 感染力が感染時間に依存するモデル

人口 N の集団を、感受性者 (S)、感染者 (I) および回復者 (R) の 3 区画 (コンパートメント) に分ける。それぞれの人口は時間 (時間の単位を 日 とすれば日数の意味。以下断りなく時間または時刻と表記する) t の関数 $S(t)$ 、 $I(t)$ 、 $R(t)$ であると考え。比較的短い期間を対象とし、新たな人口の発生は考えない。また、この間の死者は $R(t)$ に含めて考えれば、総人口は変わらないとすることができる。

$$N = S(t) + I(t) + R(t) = \text{const.} \quad (1)$$

感染者 $I(t)$ はそれぞれ感染時からの経過時間 a が異なり、経過時間 a から $a + da$ までの間の人数が $i(t, a)da$ となるような感染者分布関数 $i(t, a)$ を考える。このとき、

$$I(t) = \int_0^{\infty} i(t, a) da \quad (2)$$

である。

ある感染者とそこから二次感染者の組み合わせが特定できれば、それらのデータを集積することにより、発症間隔分布 $f_{\beta}(a)$ が得られる。感染してから時間 a だけ経過した (つまり感染時間 a の) ある感染者が他者へ二次感染する確率は、 $f_{\beta}(a)$ に比例することになる。ただし $f_{\beta}(a)$ は正規化されているものとする。

$$\int_0^{\infty} f_{\beta}(a) da = 1$$

このとき、感染時間 a の感染者が感受性者 $S(t)$ 人を含む集団内で単位時間 (時間の単位を 日 とすれば、1 日あたりの意味。以下断りなく単位時間と表記する) に二次感染させる数は $S(t)$ および $f_{\beta}(a)$ に比例すると考えられるので、これを $\beta_0 S(t) f_{\beta}(a)$ と表すと、全感染者が引き起こす新規感染者は単位時間あたり次式 (人/日) となる。

$$B(t) = \beta_0 S(t) \int_0^{\infty} f_{\beta}(a) i(t, a) da = \beta_0 S(t) p(t) \quad (3)$$

$$p(t) = \int_0^{\infty} f_{\beta}(a) i(t, a) da \quad (4)$$

ある一人の感染者による二次感染数 (再生産数) は

$$(\text{再生産数}) = \beta_0 S(t) \int_0^{\infty} f_{\beta}(a) 1 da = \beta_0 S(t)$$

$S(t) = N$ (総人口) のとき、これは基本再生産数であるので、

$$\beta_0 = \frac{(\text{基本再生産数})}{N} \quad (5)$$

である。

また、感染者が回復、隔離、または死去して感染経路から除外される確率 (隔離率) が、各感染時間に応じて単位時間あたり $\gamma(a)$ で表されるものとする、全隔離者は単位時間あたり

$$q(t) = \int_0^{\infty} \gamma(a) i(t, a) da \quad (6)$$

で表される。これは隔離による感染者 $I(t)$ の減少量であり、回復者 (隔離者) $R(t)$ の単位時間の増加量となる*1。

感染時間 a から $a + \Delta a$ の間の感染者の人数 $i(t, a)\Delta a$ は Δt 時間内に $\gamma(a)i(t, a)\Delta a\Delta t$ だけ減少し、それが Δt 時間後の分布 $i(t + \Delta t, a + \Delta t)\Delta a$ となる。したがって、

$$i(t + \Delta t, a + \Delta t)\Delta a - i(t, a)\Delta a = -\gamma(a) i(t, a)\Delta t\Delta a$$

の関係が成立し、 $i(t + \Delta t, a + \Delta t) = i(t, a) + \frac{\partial i(t, a)}{\partial t}\Delta t + \frac{\partial i(t, a)}{\partial a}\Delta a$ であることを用いると、

$$\frac{\partial i(t, a)}{\partial t} + \frac{\partial i(t, a)}{\partial a} = -\gamma(a) i(t, a)$$

が成立することがわかる。

また、 $i(t, a)$ の $a = 0$ 境界では、 Δt 時間内の新規感染者数は $i(t, 0)\Delta t$ であり、これは S の減少量 $-\frac{dS(t)}{dt}\Delta t$ に等しく、共に $B(t)\Delta t$ である。

$$i(t, 0)\Delta t = -\frac{dS(t)}{dt}\Delta t = B(t)\Delta t$$

以上をまとめると、各区画の人口 $S(t)$ 、 $i(t, a)$ 、 $R(t)$ の変化を表す基礎式は次式となる。

$$\frac{dS(t)}{dt} = -B(t) = -\beta_0 S(t) \int_0^{\infty} f_{\beta}(a) i(t, a) da \quad (7)$$

$$i(t, 0) = B(t) \quad (8)$$

$$\frac{\partial i(t, a)}{\partial t} + \frac{\partial i(t, a)}{\partial a} = -\gamma(a) i(t, a) \quad (9)$$

$$\frac{dR(t)}{dt} = q(t) \quad (10)$$

*1 この $\gamma(a)$ の考え方は、数理人口学 (人口統計学) において年令 a における年間死亡率の概念が元になっており、人口学に当てはめるとその意味は明快である。感染症伝播問題に当てはめると、(死去を除けば) 検査により感染が確認されて隔離される割合を意味し、やや意味が不明確である。感染者の隔離は人為的な操作であり、政策的な介入の効果を評価するために $\gamma(a)$ を用いることができるかもしれない。計算プログラムでは $\gamma(a) = \gamma_0 = \text{const.}$ とした。

また、一度感染した個人は感染性を失った (回復した) としても、 $\gamma(a)$ に含まれない限り感染者区画 $I(t)$ に含まれるので、感染性を失うのに十分な時間が経過すれば (つまり、 $f_{\beta}(a)$ が十分小さくなれば)、感染力を失ったのに隔離されずに残っている感染者を回復者区画 $R(t)$ へ移すことが必要となる。

$B(t)$ 、 $q(t)$ は式 (3)、(6) であり、また 時刻 $t = 0$ で下記の初期条件が与えられるものとする。

$$S(0) = S_0 \quad (11)$$

$$i(0, a) = i_0(a) \quad (12)$$

$$R(0) = R_0 \quad (13)$$

4 特性曲線による $i(t, a)$ の解

式 (9) の特性曲線は、

$$\frac{dt}{1} = \frac{da}{1} = \frac{di}{-\gamma(a) i(t, a)}$$

となる [3]。つまり、 $dt = da$ で表される $t - a$ 平面上の直線

$$t - a = \text{const.} (= t_0 - a_0)$$

に沿って

$$\frac{di}{da} \left(\text{or } \frac{di}{dt} \right) = -\gamma(a) i(t, a) \quad (14)$$

が成立する。書き直して

$$\frac{di}{i} = -\gamma(a) da$$

積分して

$$\ln i(t, a) - \ln i(t_0, a_0) = - \int_{a_0}^a \gamma(\sigma) d\sigma$$

つまり

$$i(t, a) = i(t_0, a_0) e^{- \int_{a_0}^a \gamma(\sigma) d\sigma} = i(t_0, a_0) \frac{l(a)}{l(a_0)} \quad (15)$$

となる。ただし、 (t_0, a_0) は同じ特性曲線上の点 (したがって、 $t - a = t_0 - a_0$) であり、また $l(a)$ は

$$l(a) = e^{- \int_0^a \gamma(\sigma) d\sigma} \quad (16)$$

である。この $l(a)$ は、新規感染者 $i(t - a, 0)$ のうち感染時間 a まで感染者 $i(t, a)$ であり続ける割合であり、(感染源の)「生残率」とよばれる [2]。

$t_0 = 0$ とした初期条件 式 (12)、または $a_0 = 0$ とした境界条件 式 (8) を満たすように (t_0, a_0) を定めると、解 (15) は次式で表される。

$$i(t, a) = \begin{cases} i(0, a - t) e^{- \int_{a-t}^a \gamma(\sigma) d\sigma} = i_0(a - t) \frac{l(a)}{l(a-t)} & (t < a) \\ i(t - a, 0) e^{- \int_0^a \gamma(\sigma) d\sigma} = B(t - a) \frac{l(a)}{l(0)} & (t > a) \end{cases} \quad (17)$$

$t - a$ 平面上の特性曲線と $t = 0$ および $a = 0$ における初期条件、境界条件のイメージを図 1 に示す。

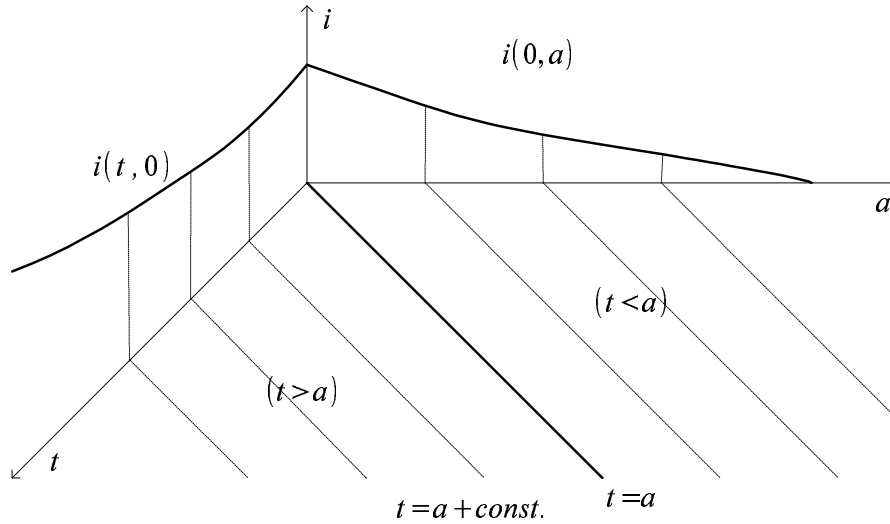


図1 特性曲線による解

5 離散化と数値解

時間刻み Δt 間隔で解 S_k, R_k を求めるものとする。

また $i(t, a)$ についても、 $\Delta t, \Delta a (= \Delta t)$ 間隔で離散化して $i_{k,j}$ とし、 a の関数 $f_\beta(a), \gamma(a)$ も Δa 間隔で離散化して $f_{\beta j}, \gamma_j$ と表されているものとする。

このときある関数の a に関する積分、式 (2)、(4)、(6) は、発症間隔分布関数 $f_\beta(a)$ が所定の微小量 (0.001) 以下となる範囲を無視して、 a の上限値 a_{na} を定め、有限範囲 $[0, a_{na}]$ の積分を、各点を均等な重みで評価する矩形近似により求める。

$$\int_0^{a_{na}} f(a) da = \left(\sum_{j=0}^{na} f_j \right) \Delta a$$

台形公式またはシンプソン公式などの不均等な重みの積分公式を用いた場合は $i(t, a)$ の特性曲線法による解との整合性が悪く、 S, I, R の各区画人数の総和が一定値にならない。

以上の方針で離散化し、時間ステップ k ($t = k\Delta t$) までの値が既に求まっているとし、時間ステップ $k+1$ の値を以下のようにして求める。

式 (9) の離散化には式 (15) の解および境界条件 (8) を用いて、

$$i_{k+1,j+1} = i_{k,j} \frac{l_{j+1}}{l_j} \quad (j = 0 \sim na - 1) \quad (18)$$

$$i_{k+1,0} = \beta_0 S_{k+1} p_{k+1} \quad (19)$$

$i_{k+1,0}$ の計算に際して未知の S_{k+1} の他に、 p_{k+1} にも $i_{k+1,0}$ が必要であるが、これは、下記の S_{k+1} の計算と含めて、反復計算により求める。

S_{k+1} を求めるために、式 (7) を次式のように中心差分で置き換える。

$$\frac{S_{k+1} - S_k}{\Delta t} = -\beta_0 \frac{S_{k+1} p_{k+1} + S_k p_k}{2}$$

これより

$$S_{k+1} = \frac{1 - 0.5\beta_0 p_k \Delta t}{1 + 0.5\beta_0 p_{k+1} \Delta t} S_k \quad (20)$$

となる。 p_{k+1} の中に未知の $i_{k+1,0}$ が含まれているので、下記のような反復計算により $i_{k+1,0}$ および S_{k+1} を求めることができる。

- (1) $i_{k+1,0}$ の値を仮定する (例えば、 $i_{k,0}$ の値等)。
- (2) 式 (4) より p_{k+1} を求める。
- (3) 式 (20) より S_{k+1} を求める。
- (4) 式 (19) より $i_{k+1,0}$ を求める。
- (5) $i_{k+1,0}$ が収束するまで ステップ (2) 以下を繰り返す。

$i_{k,j}$ が求めれば、それを積分して $I_k = \int_0^\infty i_{k,j} da$ が求まる。

R_{k+1} についても、下記の中心差分を用いて求めることができる。

$$\frac{R_{k+1} - R_k}{\Delta t} = 0.5 (q_{k+1} + q_k)$$

より

$$R_{k+1} = R_k + 0.5 (q_{k+1} + q_k) \Delta t \quad (+ i_{k,na} \Delta a)$$

しかし、この方法では式 (18) の特性曲線による積分解 (厳密解) と差分式による近似解の差のために、全区間の総人口が一定にならない。そこで、式 (18) により特性曲線に沿って $i_{k+1,j+1}$ を計算した際に、隔離した人数を積算することにより R_{k+1} を次式により求めるように改めた。

$$R_{k+1} = R_k + \left[\sum_{j=0}^{na-1} (i_{k,j} - i_{k+1,j+1}) + i_{k,na} \right] \Delta a \quad (21)$$

上式の最後の項 $i_{k,na} \Delta a$ は、感染能力を失った既感染者を回復者の区画へ移す操作である。これにより、全区画人口の総計は一定値となる。

6 発症間隔分布関数 $f_\beta(a)$ について

対象とする発症間隔分布としては、感染症伝播関連で多く取り上げられているワイブル分布、ガンマ分布、対数正規分布、および指数分布 (SIR モデル相当)、SEIR モデル相当分布を対象とする。

ワイブル分布 この分布の確率密度関数は次式で与えられる。

$$f_\beta(a) = \frac{m}{\eta} \left(\frac{a}{\eta} \right)^{m-1} e^{-(a/\eta)^m} \quad (22)$$

この平均値 (期待値) は

$$\mu = E[a] = \int_0^\infty a f_\beta(a) = \eta \Gamma \left(1 + \frac{1}{m} \right)$$

分散は

$$\sigma^2 = E[(a - \mu)^2] = \eta^2 \Gamma \left(1 + \frac{2}{m} \right) - \mu^2$$

となる。平均値 μ と標準偏差 σ が与えられたとき、パラメータ m と η は次式より求まる。

$$F(m) = \frac{\{\Gamma(1 + \frac{1}{m})\}^2}{\Gamma(1 + \frac{2}{m})} = \exp\left\{2\ln\Gamma\left(1 + \frac{1}{m}\right) - \ln\Gamma\left(1 + \frac{2}{m}\right)\right\} = \frac{\mu^2}{\sigma^2 + \mu^2} \quad (23)$$

$$\eta = \frac{\mu}{\Gamma(1 + \frac{1}{m})} \quad (24)$$

ただし $\Gamma(\cdot)$ はガンマ関数*2であり、式 (23) の解 m は二分法等で求めることができる。

ガンマ分布 この分布の確率密度関数は次式で与えられる。

$$f_{\beta}(a) = \frac{a^{m-1}}{\Gamma(m)\eta^m} e^{-a/\eta} \quad (25)$$

この平均値および分散は、次式となる。

$$\begin{aligned} \mu &= m\eta \\ \sigma^2 &= m\eta^2 \end{aligned}$$

平均値 μ と標準偏差 σ が与えられたとき、パラメータ m と η は次式より求まる。

$$m = \frac{\mu^2}{\sigma^2} \quad (26)$$

$$\eta = \frac{\sigma^2}{\mu} \quad (27)$$

対数正規分布 この分布の確率密度関数は次式で与えられる。

$$f_{\beta}(a) = \frac{1}{\sqrt{2\pi}\sigma'a} \exp\left[-\frac{(\ln a - \mu')^2}{2\sigma'^2}\right] \quad (28)$$

*2 例えば、文献 [4] では $\ln\Gamma(x)$ を与える次の C 関数が示されている。

```
double gammln(double xx)
//Returns the value ln[ (xx)] for xx ( > 0.0)
{
/* From "Numerical Recipes in C" pp.213-214.
   Internal arithmetic will be done in double precision,
   a nicety that you can omit if five-figure accuracy is good enough.
*/
double x,y,tmp,ser;
static double cof[6]={76.18009172947146, -86.50532032941677,
24.01409824083091, -1.231739572450155,
0.1208650973866179e-2, -0.5395239384953e-5};
int j;
y = x = xx;
tmp = x + 5.5;
tmp -= (x + 0.5)*log(tmp);
ser = 1.000000000190015;
for (j = 0 ; j <= 5 ; j++)
    ser += cof[j]/++y;
return -tmp+log(2.5066282746310005*ser/x);
}
```

この平均値および分散は、次式となる。

$$\begin{aligned}\mu &= e^{\mu' + \sigma'^2/2} \\ \sigma^2 &= e^{2\mu' + \sigma'^2} (e^{\sigma'^2} - 1)\end{aligned}$$

平均値 μ と標準偏差 σ が与えられたとき、パラメータ μ' と σ' は次式より求まる。

$$\sigma' = \sqrt{\ln\left(\frac{\sigma^2}{\mu^2} + 1\right)} \quad (29)$$

$$\mu' = \ln \mu - \frac{1}{2}\sigma'^2 = \ln \mu - \frac{1}{2} \ln\left(\frac{\sigma^2}{\mu^2} + 1\right) \quad (30)$$

指数分布 (SIR モデル相当) この分布の確率密度関数は次式で与えられる。

$$f_{\beta}(a) = \gamma e^{-\gamma a} \quad (31)$$

この平均値および分散は次式となる。

$$\begin{aligned}\mu &= \frac{1}{\gamma} \\ \sigma^2 &= \frac{1}{\gamma^2}\end{aligned}$$

平均値 μ が与えられれば、

$$\gamma = \frac{1}{\mu} \quad (32)$$

この場合は標準偏差を合わせることはできない。

表 2 図 2 の分布のパラメータ等

分布	平均値	標準偏差	パラメータ	備考
ワイブル	4.8	2.3	$m = 2.20, \eta = 5.42$	
ガンマ	4.8	2.3	$m = 4.355, \eta = 1.102$	
対数正規	4.8	2.3	$\mu' = 1.465, \sigma' = 0.4546$	
指数 (SIR 相当)	4.8	4.799	$\gamma = 0.2083 (T_I = 4.8)$	
SEIR モデル相当	4.8	3.3941	$\epsilon = \gamma = 0.4167 (T_E = T_I = 2.4)$	

SEIR モデル相当分布 SEIR モデルでは、一人の感染性者により感染する二次感染者数は次式となる。

$$I(a) = \frac{\epsilon}{\gamma - \epsilon} (e^{-\epsilon a} - e^{-\gamma a})$$

これを正規化すると、確率密度関数は次式で与えられる。

$$f_{\beta}(a) = \frac{\epsilon\gamma}{\gamma - \epsilon} (e^{-\epsilon a} - e^{-\gamma a}) \quad (33)$$

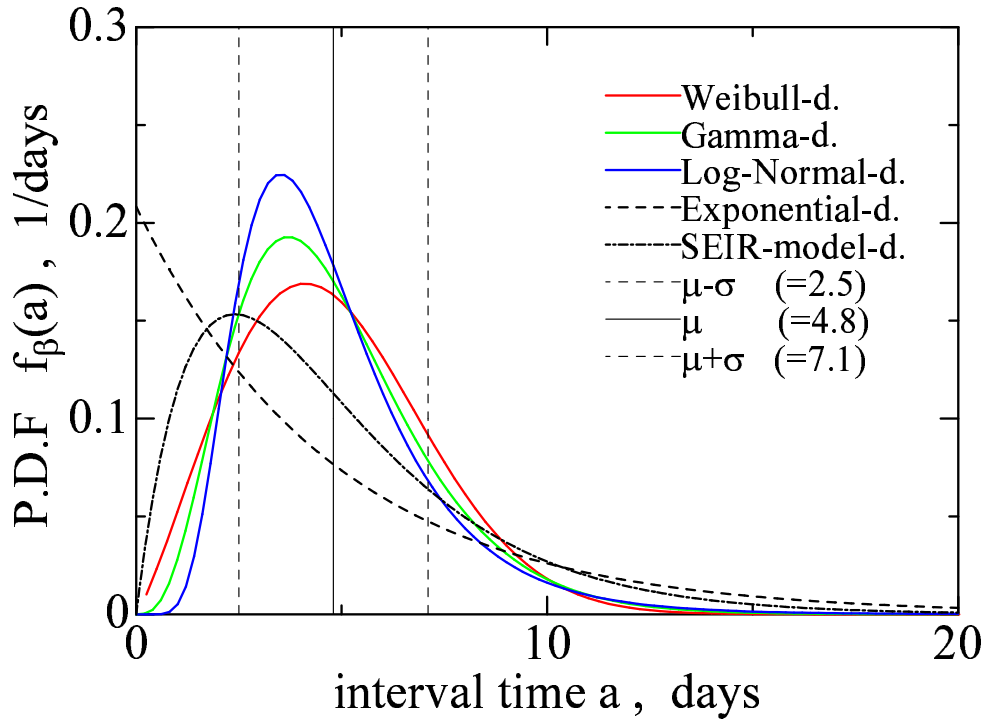


図2 発症間隔分布の確率密度関数

この平均値および分散は次式となる。

$$\mu = \frac{1}{\epsilon} + \frac{1}{\gamma}$$

$$\sigma^2 = \frac{1}{\epsilon^2} + \frac{1}{\gamma^2}$$

平均値と標準偏差が与えられれば、パラメータ ϵ と γ は次式より求まる。

$$\frac{1}{\epsilon} = \frac{1}{2} \left(\mu \mp \sqrt{2\sigma^2 - \mu^2} \right) \quad (34)$$

$$\frac{1}{\gamma} = \frac{1}{2} \left(\mu \pm \sqrt{2\sigma^2 - \mu^2} \right) \quad (35)$$

この場合、可能な分散の範囲は $\frac{1}{2}\mu^2 < \sigma^2 < \mu^2$ であり、この範囲内では指定された平均値と分散を与える分布が2種類存在する。いずれの組み合わせでも、結果は同じである。

$\sigma^2 = \frac{1}{2}\mu^2$ の場合は、 $\epsilon = \gamma$ となり、確率密度関数は

$$f_{\beta}(a) = \epsilon^2 a e^{-\epsilon a} \quad (36)$$

となる。また、 $\sigma^2 = \mu^2$ は SIR モデルに一致する。

以上の5種類の分布について、平均値 $\mu = 4.8$ 、標準偏差 $\sigma = 2.3$ となるようにパラメータを定めたものを図2と表2に示す。ただし、指数分布およびSEIRモデル相当分布は平均値のみ合わせた。

参考文献

- [1] Hiroshi Nishiura, Natalie M. Linton, Andrei R. Akhmetzhanov, "Serial interval of novel coronavirus (COVID-19) infections", *International Journal of Infectious Diseases*, 98 (2020), pp.284-286.
- [2] 稲葉 寿, "数理人口学", 東京大学出版会 (2002), pp. 17-28, 207-209.
- [3] 寺澤寛一, "自然科学者のための数学概論", 岩波書店 (1954), pp. 323-325.
- [4] W. H. Press, S.A. Teukolsky, W.T. Vetterling, B.P. Flannery, "Numerical Recipes in C", Cambridge University Press (1992), pp.213-216.