

発症間隔分布に基づいた増幅率・再生産数間の関係について (改訂)

S. Yamauchi

2020 年 7 月 29 日

目次

1	はじめに	2
2	発症間隔分布と再生産数の一般的関係	2
3	いくつかの発症間隔分布への適用	4
3.1	SIR モデル (指数分布)	4
3.2	SEIR モデル	5
3.3	ガンマ分布	6
3.4	デルタ分布	7
3.5	正規分布	7
3.6	実測に基づく分布	8

1 はじめに

ある感染症の感染者数観測結果から、その感染症の再生産数を推測することができる。観測結果を適当に指数関数 $y = Ae^{rt}$ で近似すれば、その増加率 r の時間的推移を知ることができる。感染の拡大または収束を定量的に表すにはこれでも十分であろうが、場合によっては、ある感染者が引き起こす二次感染者数の平均値、つまり再生産数 R を知ることが必要な場合もある。もちろん、増加率 r の正負は R の 1 に対する大小に対応しており、したがって感染が拡大するか縮小するかに対応する。しかし、 r と R の関係には、一次感染者が二次感染者を引き起こす時間差（発症間隔、世代間隔^{*1}）の分布が関係する。ここでは、しばしば引用されている Wallinga と Lipsitch の論文 [1] をもとに、わかりやすく解説し、いくつかの具体的な発症間隔分布に対してその関係式を導く。

2 発症間隔分布と再生産数の一般的関係

ある個体が感染症に感染して t 日後（日を単位として時間を表す）において、他の個体を 1 日の間で感染させる数（二次感染の確率）を $n(t)$ と表す。その個体が感染して以降、その病原体は次第に感染力を増して最大値に達し、やがて感染力を失って消滅するであろうから、 $n(t)$ は図 1 のような曲線を描くであろう（この曲線の縦軸は感染力に比例する量であって、必ずしも症状の強さとは一致しないかもしれない）。ある地域内である期間にわたって、この感染力を表す曲線 $n(t)$ は一定不変であると仮定する。

ある集団内では、この確率 $n(t)$ で二次感染者が生まれ、それぞれの二次感染者がその感染時以降で同じ分布で第三次感染者を生み、さらに第三次感染者が同様に第四次感染者を生み出す等々を繰り返して、結果的にある増加率 r の指数関数的な感染拡大を引き起こすことになる。

ここで別の方向から見て、ある集団である時刻 t に 1 日に発生する新規感染者数を $b(t)$ とする。その時刻に感染後 a 日経過した感染者数は a 日前の感染者数 $b(t-a)$ であり、この感染者から生み出される新たな感染者数は $b(t-a)n(a)$ である。したがって、これを全ての感染後日数 a について加算したものは、その時刻の全感染者発生数 $b(t)$ となる。

$$b(t) = \int_0^{\infty} b(t-a)n(a)da \quad (1)$$

次に、感染者数は指数関数的に増加（または減少）していると仮定し、その指数関数的増加率を r とすると、

$$b(t) = b(t-a)e^{ra}$$

と表すことができる。このとき式 (1) は

$$b(t) = \int_0^{\infty} b(t)e^{-ra}n(a)da$$

となり、 $b(t)$ を削除できて

$$1 = \int_0^{\infty} e^{-ra}n(a)da \quad (2)$$

^{*1} (serial interval) このような概念は、数理人口学（人口統計学）の分野に由来しているため、親子の年齢差つまり世代間隔と呼ばれてきた。感染症分野では、親元の一次感染者と二次感染者との感染時間の差（日数）の意味である。さらに、感染から症状が発現するまでには日数（潜伏期間）に差があるであろうから、発症間隔というのも正確な表現ではないと思われるが、疫学の分野では、「発症間隔」「世代間隔」で通されているようである。

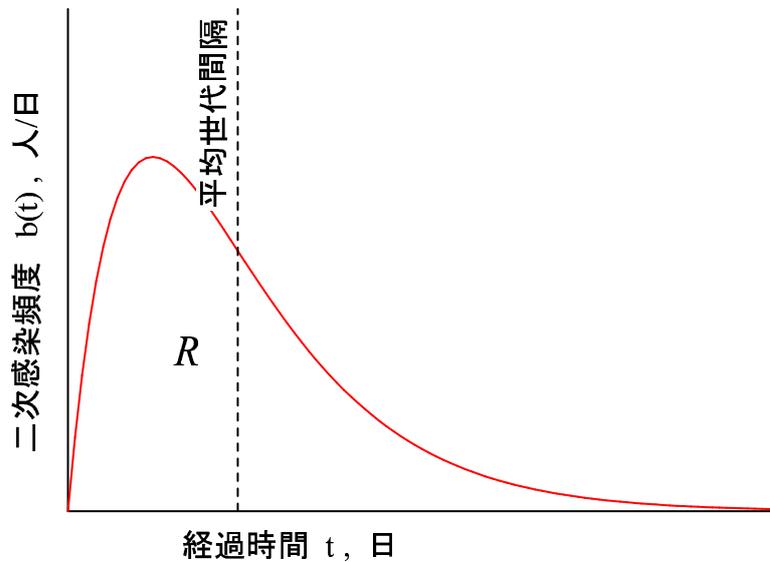


図1 二次感染発生頻度の分布

の関係が成立することがわかる (ロトカ・オイラーの式)。

ところで、図1の曲線 $n(t)$ と t 軸で囲まれる面積は、一人の感染者 (一次感染者) から生じる新たな二次感染者の数、再生産数 R である。

$$R = \int_0^{\infty} n(t) dt \quad (3)$$

図1の曲線の面積が、感染後二次感染を発症するまでの日数 (発症間隔) の確率を表すように (つまり確率密度関数となるように)、 $n(t)$ を正規化して

$$g(t) = \frac{n(t)}{\int_0^{\infty} n(a) da} = \frac{n(t)}{R} \quad (4)$$

を考える。 $g(t)$ は発症間隔分布とよばれる。

このとき、ロトカ・オイラーの式 (2) は

$$\frac{1}{R} = \int_0^{\infty} e^{-rt} g(t) dt \quad (5)$$

となる*2。これより、発症間隔分布 $g(t)$ が与えられれば、増加率 r と再生産数 R の関係が求まる。以下にいくつかの代表的な発症間隔分布に対して、この関係式を求める。

*2 この式 (5) の右辺は関数 $g(t)$ のラプラス変換 $G(s) = \int_0^{\infty} e^{-st} g(t) dt$ そのものであり、また確率論や統計学分野で用いられているモーメント母関数

$$M(z) = \int_0^{\infty} e^{zt} g(t) dt$$

において $z = -r$ と置いたものとなっている。

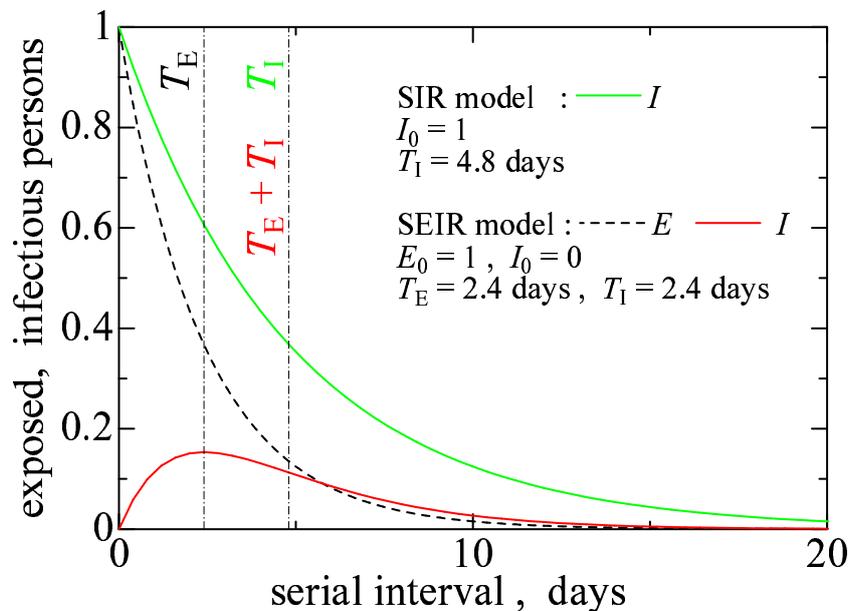


図2 SIRモデル、SEIRモデルの二次感染発生頻度分布

3 いくつかの発症間隔分布への適用

3.1 SIRモデル (指数分布)

SIRモデルでは、対象となる集団を感受性者 (susceptible)、感染者 (infectious)、または回復者 (recovered) のいずれかの状態に分類する。このうち、新たな感染者が加わらない (つまり周囲に感受性者がいない) とき、感染者数 $I(t)$ の時間変化が次式で表されると仮定する。

$$\frac{dI}{dt} = -\gamma I$$

時刻 $t = 0$ で初期感染者数 $I = I_0$ であったとすると、この解は

$$I(t) = I_0 e^{-\gamma t}$$

となり、図2の緑の実線で示すように、感染者は指数関数的に減少して回復者となる。

SIRモデルでは、この $I(t)$ は感染者数の時間変化であり、集団内での感染力は感染者数に比例しているため、一人の感染者についてみれば、その感染力は一定と考えていることになる。このことは、ある感染者が感染して $I(t)$ に比例する感染力を獲得し、徐々に感染力を変化させ、 $I(t)$ が十分小さくなって感染力を失えば感染者でなくなると考える場合とまったく等価である。したがって、SIRモデル (次のSEIRモデルも同じ) の $I(t)$ は上記の $n(t)$ に比例する量を表していると考えられる。

これより、

$$\int_0^{\infty} I(t) dt = I_0 \int_0^{\infty} e^{-\gamma t} dt = \frac{I_0}{\gamma}$$

となることを用いて $I(t)$ を正規化すると、発症間隔分布の確率密度関数 $g(t)$ が求まる。

$$g(t) = f_I(t) = \frac{I(t)}{\int_0^\infty I(t)dt} = \gamma e^{-\gamma t} \quad (6)$$

これより、式 (5) を用いて R と r の関係を求めると、

$$\frac{1}{R} = \int_0^\infty e^{-rt} f_I(t)dt = \gamma \int_0^\infty e^{(-r-\gamma)t} dt = \frac{\gamma}{r+\gamma}$$

つまり、

$$R = 1 + \frac{r}{\gamma} = 1 + T_I r \quad (7)$$

となる。

発症間隔 t の平均値 (期待値) は

$$\mu = \int_0^\infty t f_I(t)dt = \int_0^\infty \gamma t e^{-\gamma t} dt = \frac{1}{\gamma} = T_I \quad (8)$$

分散は

$$\begin{aligned} \sigma^2 &= \int_0^\infty (t - \mu)^2 f_I(t)dt = \int_0^\infty t^2 f(t)dt - \mu^2 \\ &= \int_0^\infty \gamma t^2 e^{-\gamma t} dt - \frac{1}{\gamma^2} = \frac{2\gamma}{\gamma^3} - \frac{1}{\gamma^2} = T_I^2 \quad (= \mu^2) \end{aligned} \quad (9)$$

となる。

3.2 SEIR モデル

SEIR モデルでは、感受性者と感染者の間に、感染症病原体に暴露されているがまだ他者への感染力を持たない潜伏者 (exposed) を付け加える。そして、新たな潜伏者が加わらないとき、潜伏者 $E(t)$ と感染者数 $I(t)$ の時間変化が次式で表されると仮定する。

$$\begin{aligned} \frac{dE}{dt} &= -\epsilon E \\ \frac{dI}{dt} &= \epsilon E - \gamma I \end{aligned}$$

$t = 0$ で $E = E_0$ 、 $I = 0$ となる解は、

$$\begin{aligned} E(t) &= E_0 e^{-\epsilon t} \\ I(t) &= \frac{\epsilon}{\gamma - \epsilon} E_0 (e^{-\epsilon t} - e^{-\gamma t}) \end{aligned}$$

となる (図 2 の黒破線および赤実線)。SIR モデルと同様に、 $I(t)$ は二次感染の感染力を表していると考えることができるので、

$$\int_0^\infty I(t)dt = \frac{\epsilon}{\gamma - \epsilon} E_0 \int_0^\infty (e^{-\epsilon t} - e^{-\gamma t}) dt = \frac{E_0}{\gamma}$$

を用いて $I(t)$ を正規化すると、

$$g(t) = \frac{I}{\int_0^\infty I(t)dt} = \frac{1}{1/\epsilon - 1/\gamma} (e^{-\epsilon t} - e^{-\gamma t}) \quad (10)$$

となる。

$f_I(t)$ を式 (5) に用いると、 R と r の関係として

$$\begin{aligned}\frac{1}{R} &= \int_0^{\infty} e^{-rt} g(t) dt = \frac{1}{1/\epsilon - 1/\gamma} \int_0^{\infty} \left\{ e^{-(r+\epsilon)t} - e^{-(r+\gamma)t} \right\} dt \\ &= \frac{1}{1/\epsilon - 1/\gamma} \left(\frac{1}{r+\epsilon} - \frac{1}{r+\gamma} \right) = \left(\frac{\epsilon}{r+\epsilon} \right) \left(\frac{\gamma}{r+\gamma} \right)\end{aligned}$$

が得られる。したがって

$$R = \left(1 + \frac{r}{\epsilon}\right) \left(1 + \frac{r}{\gamma}\right) = (1 + rT_E)(1 + rT_I) \quad (11)$$

となる。

発症間隔 t の平均値は

$$\begin{aligned}\mu &= \int_0^{\infty} t f_I(t) dt = \frac{1}{1/\epsilon - 1/\gamma} \int_0^{\infty} t (e^{-\epsilon t} - e^{-\gamma t}) dt \\ &= \frac{1}{1/\epsilon - 1/\gamma} \left(\frac{1}{\epsilon^2} - \frac{1}{\gamma^2} \right) = \frac{1}{\epsilon} + \frac{1}{\gamma} = T_E + T_I\end{aligned} \quad (12)$$

分散は

$$\sigma^2 = \int_0^{\infty} t^2 f_I(t) dt - \mu^2 = \frac{1}{1/\epsilon - 1/\gamma} \int_0^{\infty} t^2 (e^{-\epsilon t} - e^{-\gamma t}) dt - \left(\frac{1}{\epsilon} + \frac{1}{\gamma} \right)^2 \quad (13)$$

$$= \frac{1}{1/\epsilon - 1/\gamma} \left(\frac{2}{\epsilon^3} - \frac{2}{\gamma^3} \right) - \left(\frac{1}{\epsilon} + \frac{1}{\gamma} \right)^2 = \frac{1}{\epsilon^2} + \frac{1}{\gamma^2} = T_E^2 + T_I^2 \quad (14)$$

となる。

この場合は、平均値と分散を独立に選ぶことができるが、

$$\sigma^2 = (\mu - T_I)^2 + T_I^2 = 2T_I^2 - 2\mu T_I + \mu^2 = 2(T_I - \mu/2)^2 + \mu^2/2$$

より $\mu^2/2 \leq \sigma^2 \leq \mu^2$ 、つまり

$$\frac{1}{\sqrt{2}} \leq \frac{\sigma}{\mu} \leq 1$$

の範囲に制限される。

3.3 ガンマ分布

多くの感染症に対して精度良い発症間隔分布として、次式のガンマ分布がしばしば用いられる。

$$g(t) = \frac{1}{\Gamma(m)\eta} \left(\frac{t}{\eta} \right)^{m-1} e^{-t/\eta} \quad (15)$$

ただし、 $\Gamma(\cdot)$ はガンマ関数であり、 m 、 η は正の定数である。

積分公式

$$\int_0^{\infty} t^n e^{-bt} dt = \frac{\Gamma(n+1)}{b^{n+1}} \quad (\text{ただし } n > -1, \quad b > 0)$$

を用いると、式 (5) は

$$\frac{1}{R} = \int_0^{\infty} e^{-rt} g(t) dt = \frac{1}{\Gamma(m)\eta^m} \int_0^{\infty} t^{m-1} e^{-(r+1/\eta)t} dt = \frac{1}{\Gamma(n)\eta^m} \frac{\Gamma(m-1+1)}{\left(r + \frac{1}{\eta}\right)^{m-1+1}} = \frac{1}{(1+r\eta)^m}$$

となり、これより

$$R = (1 + r\eta)^m \quad (\text{ただし } r > -1/\eta) \quad (16)$$

が求まる。

ガンマ分布 (15) の平均値と分散は、次式である。

$$\mu = m\eta \quad (17)$$

$$\sigma^2 = m\eta^2 \quad (18)$$

3.4 デルタ分布

二次感染が常に一定の時間 $t = \mu$ で一度だけ生じる場合、発症間隔分布は次のデルタ関数で表すことができる。

$$g(t) = \delta(t - \mu) = \begin{cases} \infty & (t = \mu) \\ 0 & (\text{else}) \end{cases} \quad \text{ただし } \int_{\mu-0}^{\mu+0} \delta(t - \mu) dt = 1 \quad (19)$$

この場合 R と r の関係は、式 (5) より

$$R = \frac{1}{\int_0^{\infty} e^{-rt} g(t) dt} = e^{r\mu} \quad (20)$$

となる。一般に増加率 r が与えられたとき、この値が、任意の発症間隔分布に対する再生産数 R の最大値となる。

この平均値は μ そのものであり、分散は 0 である。

3.5 正規分布

もし、発症間隔分布が平均値 μ で標準偏差 σ の正規分布

$$g(t) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(t-\mu)^2}{2\sigma^2}} \quad (21)$$

となる場合、式 (5) より R と r の関係は次式となる。

$$R = \frac{1}{\int_0^{\infty} e^{-rt} g(t) dt} = e^{r\mu - (1/2)r^2\sigma^2} \quad (22)$$

同じ増加率 r に対して、この正規分布で分散 σ^2 を 0 に近づけると、 R の値が増加し、デルタ分布の値 $e^{r\mu}$ に近づく。このことは、他の分布形に対しても言える。

逆に、再生産数 R が同じ値である場合、平均発症間隔が同じとして分散 σ^2 だけが大きくなれば、増加率 r が大きくなって、感染がより速く拡大することがわかる。平均発症間隔が短くなれば、二次感染が平均的に早く生じるので、増加率の増加につながるの容易に理解することができるであろう。一方、分散が増加することは、個々の二次感染の発症間隔が平均値の両側に均等に広がることに対応するが、発症間隔の長い方の二次感染より短い方の二次感染の方が、その後の感染拡大への影響が大きいと解釈することができるであろう。

これらの結果を、まとめて表 1 に示す。

表1 いくつかの発症間隔分布に対する 増加率 r と再生産数 R の関係

分布の種類	$g(t)$	平均値	分散	R	備考
指数分布 (SIR モデル)	$\frac{1}{T_I} e^{-t/T_I}$	T_I	T_I^2	$1 + rT_I$	$r > -1/T_I$
SEIR モデル	$\frac{1}{T_I - T_E} (e^{-t/T_I} - e^{-t/T_E})$	$T_E + T_I$	$T_E^2 + T_I^2$	$(1 + rT_E)(1 + rT_I)$	T_E : 平均潜伏期間 T_I : 平均感染性期間 $r > -1/T_E, r > -1/T_I$
ガンマ分布	$\frac{1}{\Gamma(m)\eta} \left(\frac{t}{\eta}\right)^{m-1} e^{-t/\eta}$	$m\eta$	$m\eta^2$	$(1 + r\eta)^m$	$r > -1/\eta$
正規分布	$\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(t-\mu)^2}{2\sigma^2}}$	μ	σ^2	$e^{r\mu - (1/2)r^2\sigma^2}$	
デルタ分布	$\delta(t - \mu)$	μ	0	$e^{r\mu}$	

3.6 実測に基づく分布

エビデミック初期段階等で感染経路を追跡できた場合は、発症間隔の観測結果をもとにそのヒストグラムを得ることができる可能性がある。そのとき、ヒストグラムの時間区分 $t_0, t_1, t_2, t_3, \dots, t_n$ 間の観測相対頻度が $y_1, y_2, y_3, \dots, y_n$ (ただし、 $\sum_{i=1}^n y_i = 1$) であるとすると、 $g_i = y_i / (t_i - t_{i-1})$ であることより、式 (5) は

$$\frac{1}{R} = \int_0^\infty e^{-rt} g(t) dt \simeq \sum_{i=1}^n \frac{y_i}{t_i - t_{i-1}} \int_0^\infty e^{-rt} dt = \frac{1}{r} \sum_{i=1}^n y_i \frac{e^{-rt_{i-1}} - e^{-rt_i}}{t_i - t_{i-1}}$$

従って次式より再生産数 R が求まる。

$$R \simeq \frac{r}{\sum_{i=1}^n y_i \frac{e^{-rt_{i-1}} - e^{-rt_i}}{t_i - t_{i-1}}} \quad (23)$$

参考文献

- [1] J. Wallinga and M. Lipsitch, "How generation intervals shape the relationship between growth rates and reproductive numbers", *Proceedings of Royal Society B* (2007) 274, pp.599-604.