Exact solution for the infection dynamics with a gamma distribution of serial intervals

Alexei Vazquez[1,2,*]

[1] *Cancer Research UK Beatson Institute, Glasgow, United Kingdom*
[2] *Institute for Cancer Sciences, University of Glasgow, Glasgow, United Kingdom*

(Dated: October 17, 2020)

The basic reproductive number, the expected number of secondary cases generated by an infected individual, is used to forecast infectious disease outbreaks. Yet, the basic reproductive number does not contain information about the incubation period, which influences the serial interval between infection and transmission. Using gamma distributions of serial intervals, I derive a formula to calculate the population reproductive number as a function of the basic reproductive number and the shape parameter of the serial interval distribution. This formula can be used to estimate the basic reproductive number from the serial interval distribution and the disease doubling time.

Infectious diseases can spread to a significant fraction of the population causing an epidemic. The chance for that to happen is determined by the infectious dynamics within each individual and the disease transmission between individuals. The within individual dynamics determines the serial interval, the time interval between a primary case being infected to the disease transmission to a secondary case. The transmission between individuals is reflected in the reproductive number, the average number of secondary infectious caused by a primary case.

The networks underlying the transmission of infectious diseases have connectivity distributions with heavy tails and exhibit the small-world property, as shown for sexual [1] and proximity [2] contact networks. In these networks the average reproductive number is proportional to the ratio between the second and first moments of the connectivity distribution [3–6], which can be very large. This evidence indicates that the topology of modern contact networks enables the occurrence of epidemic outbreaks. Less attention has been given to the interplay between the contact network topology and the distribution of serial intervals.

Before entering the main calculations, let's have a look at standard models of infectious disease dynamics. In the susceptible, infected and removed (SIR) model, susceptible individuals get infected at rate β and infected individuals get removed at a rate γ. The basic reproductive number, the average number of new infections generated by a single infected individual, is given by

$$R_{SIR} = \frac{\beta}{\gamma} \tag{1}$$

The dynamics of the infected individuals at the population level is given by

$$\dot{I} = \beta \frac{I}{N} S - \gamma I \tag{2}$$

where S and I are the number of susceptible and infected individuals and N is the population size. For short times we can assume that $S \approx N$ and $I \ll N$. In this limit we can integrate equation (7) obtaining

$$I(t) \sim e^{(P_{SIR}-1)\gamma t} \tag{3}$$

where

$$P_{SIR} = R_{SIR} \tag{4}$$

is the population reproductive number of the SIR model. According to the SIR model, the basic and population reproductive numbers are identical.

In the SEIR model susceptible individuals get infected at rate β without becoming infectious (exposed), exposed individuals become infectious at a rate α and infectious individuals get removed at a rate γ. The infected compartment is now split into exposed, with number E, and infectious, with number I. Note that once an individual become infectious it behaves exactly as an infected individual in the SIR model. Therefore we obtain the same basic reproductive number

$$R_{SEIR} = R_{SIR} \tag{5}$$

The dynamics of the exposed and infectious individuals is given by

$$\dot{E} = \beta \frac{I}{N} S - \alpha E \tag{6}$$

$$\dot{I} = \alpha E - \gamma I \tag{7}$$

For short times we can assume that $S \approx N$ and $I \ll N$. In this limit we can integrate equation (7) using an eigenvalue approach. Focusing on the largest eigenvalue, we obtain

$$I(t) \sim e^{(P_{SEIR}-1)\gamma t} \tag{8}$$

where

$$P_{SEIR} = \frac{1}{2}\left[1 - \frac{\alpha}{\gamma} + \sqrt{\left(1 - \frac{\alpha}{\gamma}\right)^2 + 4\frac{\alpha}{\gamma} R_{SEIR}}\right] \tag{9}$$

* avazque1@protonmail.com

is the population reproductive number of the SEIR model. The introduction of the exposed compartment carries as a consequence a discrepancy between the local and population reproductive numbers.

The SIR and SEIR models neglect key aspects of real populations. First, there is a variable contact rate across the population, which translates to a variable rate of disease transmission across infectious individuals. Second, while the SEIR model accounts for the incubation period of infectious diseases, it is still too restrictive. Using well stablished mathematics from the theory of age-dependent branching processes, I have previously calculated the expected number of infected individuals of infectious disease outbreaks on heterogeneous populations and any given time interval distribution [7, 8]. However, the applications were limited to an exponential distribution of serial intervals. Here I use this formalism to calculate the population reproductive number for a broader class of serial interval distributions.

The branching process approach maps the contact process mediating the disease transmission into a causal tree of disease transmission. The average reproductive number of patient zero, *i.e.* the expected number of secondary cases causally linked (generated) by patient zero, is given by

$$R_0 = \frac{\langle \beta \rangle}{\gamma} \tag{10}$$

where $\langle \beta \rangle$ is the average contact rate in the population and γ is the rate of recovery from the infection. For infected cases other than patient zero one needs to take into account that the disease spreading biases the disease transmission to individuals with a higher contact rate. The patient zero can be thought as an individual selected at random from the population. Any other infected individual will not be selected at random, but it will be found with a probability proportional to its contact rate: $\beta/N\langle\beta\rangle$, where N is the population size. Once infected, the individual found by contact will engage in new contacts at a rate β. Therefore, the average reproductive number of patients other than patient zero is

$$R = \frac{\langle \beta^2 \rangle}{\langle \beta \rangle \gamma} \tag{11}$$

R_0 gives the average number of infectious at the first generation, those generated by patient zero. $R_0 R$ gives the average number of infections at the second generation and $R_0 R^{d-1}$ gives the average number of infections at the d generation. The actual time when an infected case at generation d becomes infected equals the sum of d serial intervals and it has a probability density function $g^{\star d}(t)$, where the symbol $\star$ denotes convolution. Therefore, the average number of new infected individuals at time t is given by

$$I(t) = N_0 R_0 \sum_{d=1}^{D} R^{d-1} g^{\star d}(t) \tag{12}$$

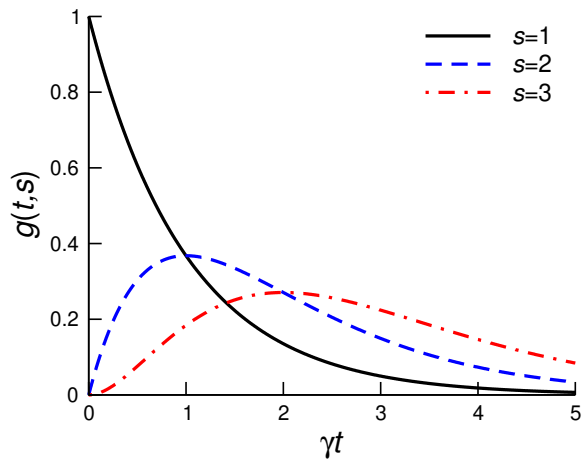

FIG. 1. Gamma distribution with shape parameters s.

where N_0 is the number of patients zero and D is the maximum generation, when the disease transmission ends. When D is small, due to lockdown for example, the spreading dynamics follows a polynomial or power-law growth [7, 9]. Here, I focus on the case $D \to \infty$, or more precisely $t \ll D \int dt g(t) t$. In this limit

$$I(t) = N_0 R_0 \sum_{d=1}^{\infty} R^{d-1} g^{\star d}(t) \tag{13}$$

Now we will focus on the shape of the serial interval distribution. In the SIR model an infected individual is infectious right from the time it becomes infected, until the infected individual is removed. Let t_R be a specific realization of the removal time. Since removal takes place at a rate γ, t_R has the exponential probability density function $\gamma e^{-\gamma t}$. Assuming that each individual has a time-independent contact rate, the serial intervals will be uniformly distributed between 0 and the recovery time, resulting in the same exponential distribution of serial intervals

$$g_{SIR}(t) = \gamma e^{-\gamma t} \tag{14}$$

In the SEIR model the serial interval can be decomposed into the sum $t_S = t_I + t_R$, where t_I is the time interval from exposed to infectious and t_R is the removal time. Since the transition from exposed to infectious takes place at a constant rate α, then t_I has the exponential probability density function $\alpha e^{-\alpha t}$. Therefore

$$\begin{aligned} g_{SEIR}(t) &= \alpha e^{-\alpha t} \star \gamma e^{-\gamma t} \\ &= \begin{cases} \gamma^2 t e^{-\gamma t} & \alpha = \gamma \\ \frac{\alpha\gamma}{\alpha-\gamma}\left(e^{-\gamma t} - e^{-\alpha t}\right) & \alpha \neq \gamma \end{cases} \end{aligned} \tag{15}$$

There are two key differences between the serial interval distributions for the SIR and SEIR model. First, the mode for $g_{SIR}(t)$ is zero and that for $g_{SEIR}(t)$ is non-zero. Second, around $t = 0$, $g_{SIR}(t) \sim \gamma$, while $g_{SEIR}(t) \sim$

$\alpha\gamma t$. Both differences are a consequence of introducing an incubation state between being infected and becoming infectious.

The case $\alpha \neq \gamma$ in (15) makes difficult the calculation of convolutions. A more suitable functional form is the gamma distribution

$$g(t,s) = \frac{\gamma}{\Gamma(s)}(\gamma t)^{s-1}e^{-\gamma t} \tag{16}$$

where $s \geq 1$ is a shape parameter quantifying the convexity around $t = 0$ (Fig. 1). The case $g(t,1)$ corresponds with the exponential distribution of the SIR model (14). For $s = 2$ we obtain the case $\alpha = \gamma$ of the SEIR model (15). More generally, the values of s and γ could be derived from the fitting to empirical data. For example, for COVID-19, the inspection of inferred serial interval distributions suggest $\gamma \approx 2$ [10, 11] and a fitting to a gamma distribution resulted in $\gamma = 2.5$ [12]. Besides covering many possible scenarios, the gamma distribution is amenable to convolution. Using the Laplace transform method one obtains

$$g^{\star d}(t,s) = \frac{\gamma}{\Gamma(ds)}(\gamma t)^{ds-1}e^{-\gamma t} \tag{17}$$

The convolution of a gamma distribution is itself a gamma distribution, with the exponent scaled by the order of the convolution ($d \to ds$).

Substituting (17) into (13) we obtain

$$I(t,s) = \frac{\gamma N_0 R_0}{R^{1-1/s}}e^{-\gamma t}f(R^{1/s}\gamma t, s) \tag{18}$$

where

$$f(x,s) = \sum_{d=1}^{\infty}\frac{x^{ds-1}}{\Gamma(ds)} \tag{19}$$

Equation (18) provides a series representation for the expected number of new infections. In the following we analyze different cases depending on the value of s.

For $s = 1$ equation (19) is the series expansion of the exponential

$$f(x,1) = \sum_{d=1}^{\infty}\frac{x^{s-1}}{(s-1)!} = e^x \tag{20}$$

For $s = 2$ the series expansion of the hyperbolic sine

$$f(x,2) = \sum_{d=1}^{\infty}\frac{x^{2s-1}}{(2s-1)!} = \sinh(x) \tag{21}$$

For $s = 4$ it decomposes into the series for $\sinh(x)$ and $\sin(x)$

$$\begin{aligned} f(x,4) &= \frac{x^3}{3!} + \frac{x^7}{7!} + \frac{x^{11}}{11!} + \cdots \\ &= \frac{1}{2}\left[\left(\frac{x^1}{1!} + \frac{x^3}{3!} + \frac{x^5}{5!} + \frac{x^7}{7!} + \cdots\right)\right. \\ &\quad \left. - \left(\frac{x^1}{1!} - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \cdots\right)\right] \\ &= \frac{1}{2}[\sinh(x) - \sin(x)] \end{aligned} \tag{22}$$

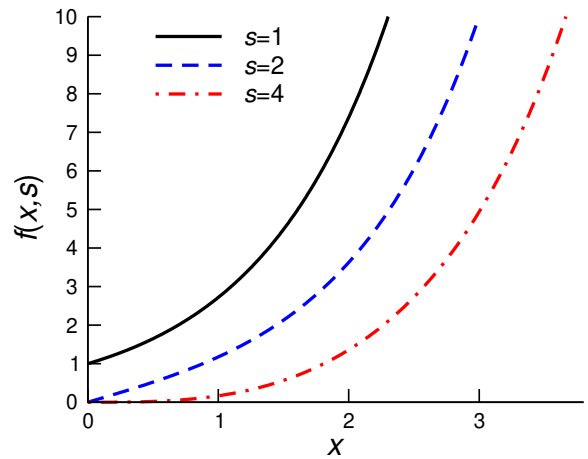

FIG. 2. Shape of $f(x,s)$ for different values of s.

Using this analytical representations we uncover the impact of the time serial distribution on the infection dynamics (Fig. 2). The case $s = 1$ is not representative for $x \ll 1$. Indeed, $f(0,1) = 1$ while $f(0,s) = 0$ for all $s > 1$. We also observe that with increasing s the plot of $f(x,s)$ shifts to the right.

More generally, when s is a natural number

$$f(x,s) = \sum_{d=1}^{\infty}\frac{x^{ds-1}}{(ds-1)!} \tag{23}$$

is a subseries of the exponential function. This series has been calculated [13, 14], resulting in

$$f(x,s) = \frac{1}{s}\sum_{n=0}^{s-1}\frac{e^{x\omega^n}}{\omega^{n(s-1)}} \tag{24}$$

where $\omega = e^{2\pi i/s}$. Since $\Re(\omega^n) < 1$ for all $n = 1, \ldots, s-1$, we obtain the asymptotic behaviour for $x \gg 1$

$$f(x,s) \approx \frac{1}{s}e^x \tag{25}$$

Based on this asymptotic behaviour, for $R^{1/s}\gamma t \gg 1$ the expected number of infected individuals, equation (18), can be approximated by

$$I(t,s) \approx \frac{\gamma N_0 R_0}{R^{1-1/s}}\frac{1}{s}e^{[P-1]\gamma t} \tag{26}$$

where

$$P = R^{1/s} \tag{27}$$

is the population reproductive number.

The equation for the population reproductive number (27) can be used to estimate the basic reproductive number from empirical data for the serial interval distribution and the doubling time. According to equation (26), the disease doubling time is given by

$$t_D = \frac{\ln 2}{(P-1)\gamma} \tag{28}$$

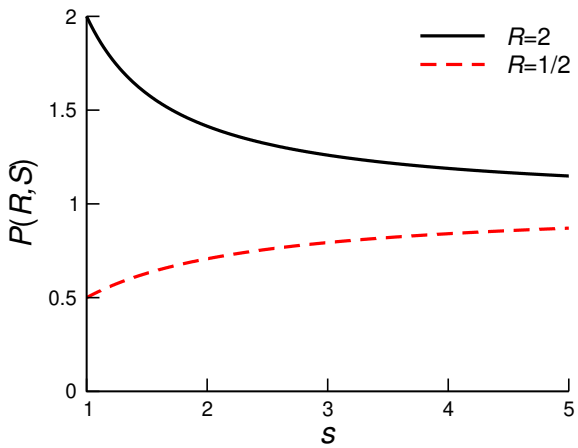

FIG. 3. Population reproductive number P as a function of s for $R > 1$ (solid line) and $R < 1$ (dashed line).

t_D can be estimated from the plot of the number of new cases as a function of time. s and γ can be estimated from a fit of the serial interval distribution to a gamma distribution. Then, using equations (27) and (28) we can estimate P and R.

Equation (27) is definitive proof that the shape of the serial interval distribution determines the relationship between the basic (R) and population (P) reproductive numbers. For $s = 1$ we recover the SIR model equation (4), when the individual and population reproductive numbers coincide. For $s = 2$ we recover the case $\alpha = \beta$ of the SEIR model in equation (9). Figure 3 shows the population reproductive number as a function of s for two different local reproductive numbers. When $R > 1$, P decreases monotonically with increasing s. When $R < 1$, P increases monotonically with increasing s. In either case, P approaches 1 for large s. Note, however, that the shape parameter s does not change the fact that if $R > 1$ then $I(t)$ grows exponentially, while $I(t)$ decays exponentially when $R < 1$.

The analysis of a gamma distribution of serial intervals can be extended to calculate the population reproductive number when there are heterogenous mixing patterns between individuals according to certain types [15, 16]. The outcome is similar to equation (27), after replacing R by the largest eigenvalue ρ of the missing matrix of reproductive numbers (equation 12 in Ref [16]). After making this substitution, we obtain the population reproductive number for the multi-type generalization

$$P = (R\Lambda_1\Lambda_2\cdots)^{1/s} \tag{29}$$

where Λ_i is the largest eigenvalue of the mixing matrix of the ith population stratification (*e.g.*, age, mask use, etc) [16].

In conclusion, the branching process formalism allows for a flexible description of infectious disease outbreaks that can be fully based on empirical distributions. The outcome is two equations: (i) The population reproductive number as a function of the basic reproductive number and the shape parameter of gamma distribution of serial intervals and (ii) the disease doubling time as a function of the population reproductive number and the scale parameter of the gamma distribution of serial intervals. These equations can be used to make estimate the basic reproductive number from empirical data.

ACKNOWLEDGEMENTS

This work was supported by Cancer Research UK C596/A21140.

[1] F. Liljeros, C. R. Edling, L. A. Amaral, H. E. Stanley, and Y. Aberg, Nature **411**, 907 (2001).
[2] S. Eubank, H. Guclu, V. S. Kumar, M. V. Marathe, A. Srinivasan, Z. Toroczkai, and N. Wang, Nature **429**, 180 (2004).
[3] R. Pastor-Satorras and A. Vespignani, Phys. Rev. Lett. **86**, 3200 (2001).
[4] M. E. J. Newman, Phys. Rev. E **66**, 016128 (2002).
[5] M. Boguñá, R. Pastor-Satorras, and A. Vespignani, Phys. Rev. Lett. **90**, 028701 (2003).
[6] A. Vazquez and Y. Moreno, Phys. Rev. E **67**, 015101 (2003).
[7] A. Vazquez, Phys. Rev. Lett. **96**, 038702 (2006).
[8] A. Vazquez, "Causal tree of disease transmission and the spreading of infectious diseases," in *Discrete Methods in Epidemiology*, DIMACS Series in Discrete Mathematics and Theoretical Computer Science, edited by J. Albello (2006) pp. 161–176.
[9] A. Vazquez, medRxiv 2020.07.23.20160531 (2020).
[10] X. He, E. H. Y. Lau, P. Wu, X. Deng, J. Wang, X. Hao, Y. C. Lau, J. Y. Wong, Y. Guan, X. Tan, X. Mo, Y. Chen, B. Liao, W. Chen, F. Hu, Q. Zhang, M. Zhong, Y. Wu, L. Zhao, F. Zhang, B. J. Cowling, F. Li, and G. M. Leung, Nature Medicine **26**, 672 (2020).
[11] S. Ali and *et al*, Science **369**, 1106 (2020).
[12] Z. Du, X. Xu, Y. Wu, L. Wang, B. Cowling, and L. A. Meyers, **26**, 1341 (2020).
[13] L. Rubel and K. Stolarsky, The American Mathematical Monthly **87**, 371 (1980).
[14] H. Chen, International Journal of Mathematical Education in Science and Technology **41**, 538 (2010).
[15] A. Vazquez, Phys. Rev. E **74**, 066114 (2006).
[16] A. Vazquez, medRxiv 2020.10.09.20210252 (2020).