

Unraveling the paradox between generation and serial intervals:
applications to COVID-19 pandemic
Sang Woo Park

Abstract

1 Introduction

Since the emergence of the novel coronavirus disease (COVID-19), a significant amount of research has focused on estimating its reproduction number \mathcal{R} . The reproduction number, which is defined as the average number of secondary cases caused by a primary case, allows us to predict the extent to which a disease will spread in the population and the amount of intervention to prevent an outbreak. Since the reproduction number cannot be measured directly, it is often estimated from the observed exponential growth rate using generation- and serial-interval distributions.

The generation interval is defined as the time between when an individual (infector) is infected and when an individual infects another person (infectee). On the other hand, the serial interval is defined as the time between when an infector and an infectee become symptomatic. Previous studies have often alluded to differences in their variances, despite having identical expected values, and noted that using serial-interval distribution can give different estimates of \mathcal{R} . Even though these distinctions were first made over a decade ago (Svensson, 2007), the need for a better conceptual and theoretical framework for understanding their differences is becoming clearer as the COVID-19 pandemic unfolds: Researchers continue to rely on both generation and serial intervals to estimate \mathcal{R} for COVID-19 without making a clear distinction, and some even mislabel them.

The lack of clear understanding of their differences can be largely attributed to an apparent paradox. When the epidemic is growing exponentially, the spread of infection can be characterized as a “renewal process” based on previous incidence of infection, the associated generation-interval distribution, and the average infectiousness of an infected individual. This renewal formulation allows us to link the exponential growth rate of an epidemic r with its reproduction number \mathcal{R} . Likewise, since infection typically leads to symptoms, we should be able to describe the renewal process of symptomatic cases using the serial-interval distribution. Therefore, both generation- and serial-interval distributions should give us identical estimates of \mathcal{R} based on the observed epidemic growth rate. Current theory seems to contradict our biological intuition.

Here, we provide an answer to this decade-old paradox.

2 Methods

2.1 Backward and forward delay distributions

In order to model serial intervals, we first explain the differences between backward and forward delays. To do so, we divide epidemiological events as *primary* and *secondary* events. When we measure a within-individual delay, we define primary and secondary events based on their timing; the primary event always occurs before the secondary event. When we measure between-individual delays, we define primary and secondary events based on whether they occurred within an infector or an infectee, respectively; the primary event does not necessarily occur before the secondary event.

We model time delays between a primary and a secondary event from a cohort perspective.

A primary cohort consists of *all* individuals whose primary event occurred at a given time; a secondary cohort can be defined similarly based on the secondary events. For example, when we are measuring serial intervals, a primary cohort s consists of all infectors who became symptomatic at time s . Then, for each primary cohort s , we can define the expected time distribution between primary and secondary events for primary cohort s . We refer to this distribution as the forward delay distribution and denote it as $f_s(\tau)$. We assume that forward delay distributions can vary across primary cohorts.

Likewise, we can define a backward delay distribution $b_s(\tau)$ for a secondary cohort s : The backward delay distribution for secondary host s describes the time delays between a primary and secondary host given that the secondary event occurred at time s . It follows that:

$$b_s(\tau) \propto i(s - \tau)f_{s-\tau}(\tau) \quad (1)$$

where $i(s)$ is the size of the primary cohort s . Therefore, changes in the backward delay distribution depends on the changes in cohort size $i(s)$ (therefore incidence of infection) as well as changes in the forward delay distribution. This concept generalizes the work by (Champredon and Dushoff, 2015) who compared forward and backward generation-interval distributions to describe the realized generation intervals from the perspective of an infector and an infectee, respectively.

2.2 Serial interval distributions

The serial interval is defined as the time between when an infector becomes symptomatic and when an infectee becomes symptomatic. We express realized serial intervals as:

$$-x_0 + \sigma + x_1 \quad (2)$$

where x_0 and x_1 represent the realized time from infection to symptom onset of an infector and an infectee, respectively, and σ represents the realized generation interval. Previous studies have assumed that (i) x_0 and x_1 have the same mean and (ii) therefore the serial and generation intervals have the same mean; however, these results implicitly assume that incidence stays constant.

Using the cohort-based framework provides a clear way of understanding the serial-interval distribution. Given that an infector became symptomatic at time ℓ , we have to first go backward in time by asking when the infector was infected and go forward in time by asking when the infectee became symptomatic. Then, it is clear that x_0 should follow the backward incubation period distribution for secondary cohort ℓ ; σ the forward generation-interval distribution for a primary cohort $\ell - x_0$ conditional on the realized incubation period x_0 ; and s_1 forward incubation period distribution for a primary cohort $\ell - x_0 + \sigma$. Assuming that the forward incubation distribution does not vary across cohorts, the forward serial interval distribution of a primary cohort ℓ can be written as follows:

$$f_\ell(\tau) \propto \int_0^\infty \int_0^\infty i(\ell - x_0)h_{\ell-x_0}(x_0, \sigma)k(\tau - \sigma + x_0)dx_0 d\sigma \quad (3)$$

where i is the incidence of infection, h is the joint probability distribution of the incubation period and generation interval of a primary cohort $\ell - x_0$, and k is a marginal probability distribution of h describing the forward incubation periods.

Likewise, we can define backward serial interval distribution for a secondary cohort ℓ . Given that an infectee became symptomatic at time ℓ , we have to first go backward in time by asking when the infectee became infected and when the infectee was infected by the infector; then, we have to go forward in time by asking when the infector became symptomatic. In this case, x_1 and σ follow the backward incubation period and generation-interval distributions, respectively, and x_0 follows the forward incubation period distribution. Therefore, the backward serial interval distribution of a secondary cohort ℓ can be written as follows

$$\begin{aligned} b_\ell(\tau) &\propto j(\ell)f_\ell(\tau) \\ &\propto \int_0^\infty i(\ell - x_0)k(x_0)f_\ell(\tau)dx_0 \end{aligned} \quad (4)$$

where j is the incidence of symptomatic cases.

2.3 Serial link between r and \mathcal{R}

Here, we show that the forward serial-interval distribution during the exponential growth phase provides the correct link between the exponential growth rate r of an epidemic. This link is identical to the r - \mathcal{R} link based on the generation-interval distribution. During the exponential growth phase, we have

$$f_{\text{exp}}(\tau) \propto \int_0^\infty \int_0^\infty \exp(-rx_0)h(x_0, \sigma)k(\tau - \sigma + x_0)dx_0 d\sigma \quad (5)$$

3 Results

4 Appendix

Recall that the forward serial interval can be written as:

$$-x_0 + \sigma + x_1 \quad (6)$$

Note that x_1 is independent of x_0 and σ . Then, we get:

$$M_{-x_0+\sigma+x_1}(-r) = M_{-x_0+\sigma}(-r)M_{x_1}(-r). \quad (7)$$

We want to show that

$$M_{-x_0+\sigma}(-r) = \frac{M_\sigma(-r)}{M_{x_1}(-r)} \quad (8)$$

for $r \geq 0$. Note that the time between symptom onset and infection of an infectee follows the following distribution:

$$z(\tau) \propto \int_{\max(0, -\tau)}^\infty i(\ell - x_0)h_{\ell-x_0}(x_0, \tau + x_0)dx_0. \quad (9)$$

During the exponential growth phase, we have

$$z_{\text{exp}}(\tau) = \frac{1}{N} \int_{\max(0, -\tau)}^{\infty} \exp(-rx_0) h(x_0, \tau + x_0) dx_0 \quad (10)$$

where N is the normalization factor:

$$\begin{aligned} N &= \int_{-\infty}^{\infty} \int_{\max(0, -\tau)}^{\infty} \exp(-rx_0) h(x_0, \tau + x_0) dx_0 d\tau \\ &= \int_0^{\infty} \int_{-x_0}^{\infty} \exp(-rx_0) h(x_0, \tau + x_0) d\tau dx_0 \\ &= \int_0^{\infty} \exp(-rx_0) k(x_0) dx_0 \\ &= M_{x_1}(-r) \end{aligned} \quad (11)$$

Therefore,

$$\begin{aligned} &\int_{-\infty}^{\infty} \exp(-r\tau) z_{\text{exp}}(\tau) d\tau \\ &= \frac{1}{M_{x_1}(-r)} \int_{-\infty}^{\infty} \exp(-r\tau) \int_{\max(0, -\tau)}^{\infty} \exp(-rx_0) h(x_0, \tau + x_0) dx_0 d\tau \end{aligned} \quad (12)$$

Finally, we are left to prove that

$$\int_0^{\infty} \exp(-r\tau) g(\tau) d\tau = \int_{-\infty}^{\infty} \exp(-r\tau) \int_{\max(0, -\tau)}^{\infty} \exp(-rx_0) h(x_0, \tau + x_0) dx_0 d\tau \quad (13)$$

where g is a marginal probability distribution of h describing the forward generation intervals:

$$g(\tau) = \int_0^{\infty} h(x_0, \tau) dx_0. \quad (14)$$

Let $a = x_0 + \tau$. Then, by change of variables, it immediately follows that

$$\begin{aligned} &\int_{-\infty}^{\infty} \exp(-r\tau) \int_{\max(0, -\tau)}^{\infty} \exp(-rx_0) h(x_0, \tau + x_0) dx_0 d\tau \\ &= \int_0^{\infty} \int_0^{\infty} \exp(-ra) h(x_0, a) dx_0 da \\ &= \int_0^{\infty} \exp(-r\tau) g(\tau) d\tau \end{aligned} \quad (15)$$

Therefore, the forward serial- and generation-interval distributions give the same link between r and \mathcal{R} .

References

- Champredon, D. and J. Dushoff (2015). Intrinsic and realized generation intervals in infectious-disease transmission. *Proceedings of the Royal Society B: Biological Sciences* 282(1821), 20152026.
- Svensson, Å. (2007). A note on generation times in epidemic models. *Mathematical biosciences* 208(1), 300–311.