

Understanding serial interval distributions: applications to the
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 and the amount of intervention to prevent an outbreak. Since the basic reproduction number cannot be measured directly, it is often inferred from the observed exponential growth rate using generation or serial intervals.

Generation intervals are defined as the time between when an individual is infected and when an individual infects another person. Similarly, serial intervals are defined as the time between when an infector and an infectee become symptomatic.

There is an apparent paradox between generation and serial intervals. When the epidemic is growing exponentially, the disease dynamics can be described by the “renewal” of incidence based on previous incidence of infection and the associated generation-interval distribution. This renewal formulation allows us to link the exponential growth rate of an epidemic with its reproduction number. Likewise, if infection leads to symptoms, we should be able to describe the renewal of symptomatic cases based on previous incidence of symptomatic cases and the associated serial-interval distribution. The current literature does not support this idea.

Here, we resolve the apparent paradox.

We show that the serial-interval distribution can provide a correct link between the speed

2 Methods

3 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.

4 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 infected individual 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, time between symptom onset of an infector and infection of an infectee can be described by the following distribution:

$$k_\ell(\tau) \propto \int_0^\infty i(\ell - s_0)h_{\ell-s_0}(s_0, \tau + s_0)ds_0 \quad (3)$$

where h is the joint probability distribution of the incubation period and generation interval of an infector.

$$z_\ell(\tau) = \int_0^\infty i(\ell - x_0)h_{\ell-x_0}(x_0, \sigma)k(\tau - \sigma - x_0 - \ell)dx_0d\sigma \quad (4)$$

Then, time between symptom onset of an infector and infection of an infectee can be modeled by the following distributions:

$$k_\ell(\tau) = \int_0^\infty g_{\ell-s_0}(\tau + s_0)b_\ell(s_0)ds_0, \quad (5)$$

where $g_{\ell-s_0}(\tau + s_0)$ represents the forward generation-interval distribution for a primary cohort $\ell-s_0$ and $b_\ell(s_0)$ represents the backward incubation period distribution for a secondary cohort s_0 . Denoting the forward incubation period with h , we have

$$k_\ell(\tau) \propto \int_0^\infty i(\ell - s_0)g_{\ell-s_0}(\tau + s_0)h_{\ell-s_0}(s_0)ds_0, \quad (6)$$

where $i(\ell - s_0)$ represents the incidence of infection. Then, the forward serial interval distribution for a primary cohort ℓ follows:

$$z_\ell(\tau) = \int_{-\infty}^\infty k_\ell(\sigma)h_{\ell+\sigma}(\tau - \sigma)d\sigma, \quad (7)$$

Therefore, the serial-interval distribution depends on incidence of infection and does not have an equal mean as the generation-interval distribution.

During the exponential growth phase, we can assume that the forward generation-interval distributions and incubation period distributions remain constant. Then,

$$\begin{aligned} z(\tau) &= \int k_\ell(\sigma)h(\tau - \sigma)d\sigma \\ &\propto \int_{-\infty}^\infty \int_0^\infty \exp(-rs_0)g(\sigma + s_0)h(s_0)h(\tau - \sigma)ds_0 d\sigma, \end{aligned} \quad (8)$$

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.