Inferring generation interval distributions from contact tracing data

Sang Woo Park, David Champredon and Jonathan Dushoff

September 4, 2018

1 Introduction

An epidemic can be characterized by its speed (exponential growth rate, r) and its strength (reproductive number, $\mathcal{R}$). Reproductive number, defined as the average number of secondary cases arising from a primary case, is of particular interest as it provides information about the final size of an epidemic [CITE]. However, directly measuring the reproductive number often requires knowledge of entire disease history and may not be feasible early in an epidemic [CITE]. Instead, the reproductive number can be *indirectly* estimated from exponential grwoth rate, which can easily be estimated from incidence data [CITE]. These two quantities are linked by generation interval distributions (Wallinga and Lipsitch, 2007).

Generation interval is defined as the time between when a person becomes infected and when that person infects another person. As each individual experiences different course of infection, *individual* generation interval distribution varies among infectors [CITE sven]. Hence, the *intrinsic* generation interval distribution, which provides link between r and $\mathcal{R}$, is a pooled distribution across all potential infections.

Due to individual variation in infection time, the observed generation interval distribution can change depending on when and how it is measured (Champredon and Dushoff, 2015). Generation interval can be measured *forward* in time by looking at all infections that were caused by infectors that were infected at the same period of time. Early in the epidemic when depletion of susceptible is negligible, we expect the forward generation interval distribution to be similar to the intrinsic generation interval distribution. As epidemic progresses, an infector is less likely to infect individuals later in time due to decrease in susceptibles and the forward generation interval will be shorter, overall.

On the other hand, generation intervals can be measured *backward* in time by considering all infectees that were infected at the same time period and comparing when their infectors were infected. When epidemic is growing, a susceptible individual is more likely to be infected by a newly infected individual and the backward generation interval will be short. When epidemic is subsiding,

most infections are caused by the remaining infectors, rather than new infectors, and the backward generation interval will be long.

In practice, generation interval is often measured via contact tracing [CITE]. Unlike forward or backward generation interval, which focus on individuals that were infected during a short period of time and their corresponding infectees or infectors, contact tracing is often performed over a long period of time. Ideally, contact tracing will be performed from the beginning of an epidemic to a certain time period, whether the epidemic is still ongoing or not. Therefore, generation interval distribution obtained from contact tracing, which we will refer to as the observed generation interval distribution, can be thought of as weighted average of backward generation interval distribution; the observed generation interval distribution will be affected by the temporal variation of the backward generation interval.

[**SWP:** *Need a paragraph about spatial effect?*]

In this study, we explore the temporal and spatial effect in the observed generation interval obtained from contact tracing. We show that using the observed generation interval distribution directly will always underestimate the reproductive number [**SWP:** *need to confirm this statement when we're done with the ms*]. We provide a statistical framework of recovering the intrinsic generation interval distribution from the observed generation interval distribution.

2 Results

Intrinsic generation interval

This section serves to introduce notation from previous work [CITE]. Let $K(t)$ be the infection kernel. The reproduction number is defined as

$$\mathcal{R} = \int_0^\infty K(t).$$

Then, the intrinsic generation interval distributions is defined as

$$g(t) = \frac{K(t)}{\mathcal{R}}.$$

Intrinsic generation interval distribution can be considered as an intrinsic characteristic of a single average infector in a fully susceptible population.

[**SWP:** *Insert renewal equation approach*]

The observed generation interval through time

Assume that contact tracing is performed from the beginning of an epidemic to time t. The number of infection occuring at time s caused by infectors who were themselves infected at time $s - \tau$ is given by

$$i_{s-\tau}(s) = \mathcal{R} i(s-\tau) g(\tau) S(s) \tag{1}$$

Then, total number of secondary infections that are τ time steps apart and occur before time t:

$$\mathcal{R} \int_{\tau}^{t} i(s-\tau) g(\tau) S(s) ds. \tag{2}$$

Then, the observed generation interval distribution via contact tracing at time t is given by

$$g_t(\tau) = \frac{\mathcal{R} \int_{\tau}^{t} i(s-\tau) g(\tau) S(s) ds}{\mathcal{R} \int_{0}^{t} \int_{x}^{t} i(s-x) g(x) S(s) ds dx}. \tag{3}$$

We note that the expression in the denominater is equivalent to cumulative incidence at time t. The intuition behind this is that we are normalizing acrosss all incidence before time t. Then, we have

$$g_t(\tau) = \frac{\mathcal{R} \int_{\tau}^{t} i(s-\tau) g(\tau) S(s) ds}{\int_{0}^{t} i(s) ds}. \tag{4}$$

For convenience, we ignore normalizing constants and write

$$g_t(\tau) \propto g(\tau) \int_{0}^{t} i(s-\tau) S(s) ds. \tag{5}$$

Recovering intrinsic generation interval

The observed generation interval distribution is a weighted intrinsic generation interval distribution (equation 5), Then, the intrinsic generation interval can be recovered by taking the inverse weights:

$$g(\tau) \propto g_t(\tau) \frac{1}{\int_{0}^{t} i(s-\tau) S(s) ds} \tag{6}$$

However, this method requires a knowledge of susceptible dynamics and may not be feasible in practice.

During exponential growth period, we can write $i(\tau) \propto \exp(rt)$, where r is the exponential growth rate. Assuming that $S(t) \approx 1$, the observed generation interval distribution during growth period can be written as follows:

$$g_{\exp}(\tau) \propto g(\tau) \exp(-r\tau), \tag{7}$$

Taking the inverse weight, we obtain the following expression for the intrinsic generation interval distribution:

$$g(\tau) \propto g_{\exp}(\tau) \exp(r\tau) \tag{8}$$

and the reproductive number:

$$\mathcal{R} = \int_{0}^{\infty} g_{\exp}(\tau) \exp(r\tau) d\tau. \tag{9}$$

Therefore, applying the Lotka-Euler equation using the observed generation interval via contact tracing results in underestimation of reproductive number:

$$\int_0^\infty g_{\exp}(\tau)\exp(r\tau)d\tau > \left(\int_0^\infty g_{\exp}(\tau)\exp(-r\tau)d\tau\right)^{-1} \tag{10}$$

This method provides a non-parametric approach for inferring the intrinsic generation interval distribution and the reproductive number from contact tracing data.

[SWP: *Need to rewrite this section. It's merely a place holder for now:*] While the non-parametric method is simple, it does not use all available information from contact tracing data. In particular, it does not take into account who infected whom. Assuming a poisson process, we can obtain a likelihood for observing infections:

$$\mathcal{R}^{n_e}\cdot\prod g(\tau_e)\cdot\exp\left(-\mathcal{R}\int_0^{c-t_{\text{inf}}} g(s)ds\right) \tag{11}$$

This method requires us to make an assumption about the generation interval distributions... See example...

Examples

Note that the observed mean generation interval through contact tracing will always be shorter be shorter than intrinsic mean generation interval (Figure 1). There are two reasons for this phenomenon. First, as contact tracing is performed up to time t, any infection events that occur after time t are not observed. Any individuals that are infected before time t can only complete infection events that are shorter than time t and the distribution of all infection events that occur before time t will be concentrated on shorter intervals. Second, number of susceptibles decrease over the course of an epidemic and any infector is less likely to infect susceptible individuals through long generation intervals than it would have in a fully susceptible population (Champredon and Dushoff, 2015). As a result, even if contact tracing is performed through an entire epidemic, mean generation interval will be underestimated.

2.1 Spatial variation - Effective generation interval

[SWP: *Moved here for now; this text is old*] Intrinsic generation interval distribution implicitly that an infector can exert all infectious contacts without wasting any throughout the infectious period. In other words, it is conditional on the assumption that a contacted individual has not been contacted before. When the population is limited, we must take the probability that a susceptible individual can be found into account.

Let $\beta(t)$ be infectious contact rate per pair. The probability that a susceptible is still susceptible at time t is given by

$$\exp\left(-\int_0^t \beta(s)ds\right).$$

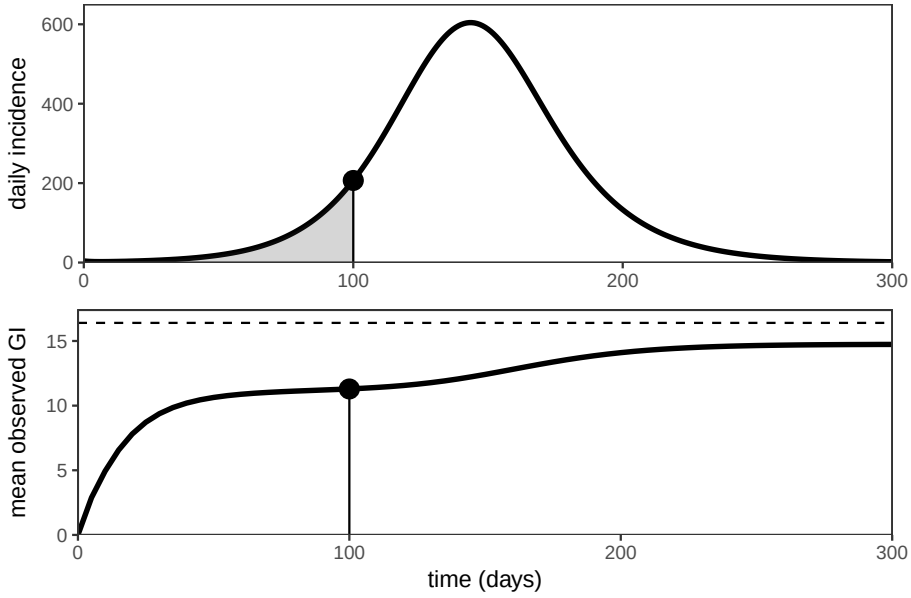

Figure 1: Fill this out and re-do simulation with realistic Ebola parameters for consistency.

Then, the effective generation interval distribution of an infected individual is proportional to the product of intrinsic generation interval distribution and this survival probability

$$g_{\text{eff}}(\tau) \propto g(\tau) \exp\left(-\int_0^\tau \beta(s)ds\right).$$

Note that the previous formulation does not take into account presence of other potential infectors. During an outbreak, we can imagine a susceptible individual being exposed to multiple infected individuals. Since effective generation interval is conditional on the assumption that a contacted susceptible individual has not been contacted previously, we have to take this into account... Then, the previous formulation can be taken as an upper bound of the actual effective GI distribution... It is very difficult to formalize this idea but we use a numerical example to demonstrate the idea:

3 Methods

References

Champredon, D. and J. Dushoff (2015). Intrinsic and realized generation intervals in infectious-disease transmission. *282*(1821).

Wallinga, J. and M. Lipsitch (2007). How generation intervals shape the relationship between growth rates and reproductive numbers. *Proceedings of the Royal Society of London B: Biological Sciences 274*(1609), 599–604.