# Notes on network/generation interval

Sang Woo Park

February 20, 2018

## 1 Censored generation interval distributions

Following Champredon and Dushoff (2015), we can write the number of infection occuring at time $t$ caused by infectors who were themselves infected at time $s$ as

$$i_s(t) = K(t - s)i(s)S(t) \tag{1}$$

Writing the kernel as the product of the intrinsic genreation distributions and $\mathcal{R}_0$, we get

$$i_s(t) = \mathcal{R}_0 g(t - s)i(s)S(t) \tag{2}$$

The censored generation interval distributions is what is often measured and we have to account for all infections that happen before time $t$. Note that number of infection occuring at time $s$ caused by infectors who were themselves infected at time $s - \tau$ is given by

$$i_{s-\tau}(s) = \mathcal{R}_0 i(s - \tau)g(\tau)S(s) \tag{3}$$

Normalizing this gives the backward generation interval distributions of the cohorts at time $s$ but we are interested in all infections that are $\tau$ time steps apart before time $t$:

$$\mathcal{R}_0 \int_\tau^t i(s - \tau)g(\tau)S(s)ds. \tag{4}$$

Then, the censored generation interval is given by

$$c_t(\tau) = \frac{\int_\tau^t i(s - \tau)g(\tau)S(s)ds}{\int_0^t \int_x^t i(s - x)g(x)S(s)dsdx}. \tag{5}$$

We note that the denominater is basically cumulative incidence up to time $t$ divided by $\mathcal{R}_0$. The stragihtforward intuition behind this is that we are normalizing by all incidence before time $t$. Mathematically, we have the following:

$$
\begin{aligned}
\mathcal{R}_0 \int_0^t \int_x^t i(s - x)g(x)S(s)dsdx &= \int_0^t \mathcal{R}_0 S(s) \int_0^s i(s - x)g(x)dxds \\
&= \int_0^t i(s)ds
\end{aligned}
\tag{6}
$$

Then,

$$c_t(\tau) = \frac{\mathcal{R}_0 \int_\tau^t i(s-\tau)S(s)ds}{\int_0^t i(s)ds} g(\tau) \tag{7}$$

We are going to come back to these later. Rather we want to focus on temporal correction first.

# 2 Temporal correction

During an exponential growth period, censored generation interval can be written as follows:

$$g_{\text{obs}}(\tau) = \mathcal{R}g(\tau)\exp(-r\tau), \tag{8}$$

This is generation interval distributions we expect to observe during an early outbreak (growing at rate $r$). Hence, to recover the intrinsic generation interval distributions, we can take the weighted distribution of the observed distribution:

$$g(\tau) = \frac{1}{\mathcal{R}} g_{\text{obs}}(\tau)\exp(r\tau). \tag{9}$$

Furthermore, $\mathcal{R}$ should be estimated by

$$\mathcal{R} = \int_0^\infty g_{\text{obs}}(\tau)\exp(r\tau)d\tau \tag{10}$$

This contrasts with the well-known Euler-Lotka equation:

$$\mathcal{R} = 1 \left/ \int_0^\infty g(\tau)\exp(-r\tau)d\tau \right. \tag{11}$$

Using this equation will lead to underestimation of $\mathcal{R}$ as well as mean generation interval (I think we can prove this using weighted AM-HM ineqaulity).

## 2.1 Numerical analysis

Consider the Ebola epidemic.

## 2.2 Simulation study - deterministic framework

We simulate an epidemic with Ebola parameters assuming $\mathcal{R} = 2$ with a population size of 60000. We study the epidemic when daily cases range between 20 and 200. For simplicity, we do not assume any under-reportings. Little $r$ is estimated by fitting an exponential curve with negative binomial error through maximum likelihood. Confidence interval is constructed using normal approximation. Confidence interval for mean generation interval and $\mathcal{R}$ is constructed by bootstrapping sample of generation intervals and drawing a random value for $r$ from normal distribution using its estimated mean and standard error from MLE.

First, we can look at the estimates:
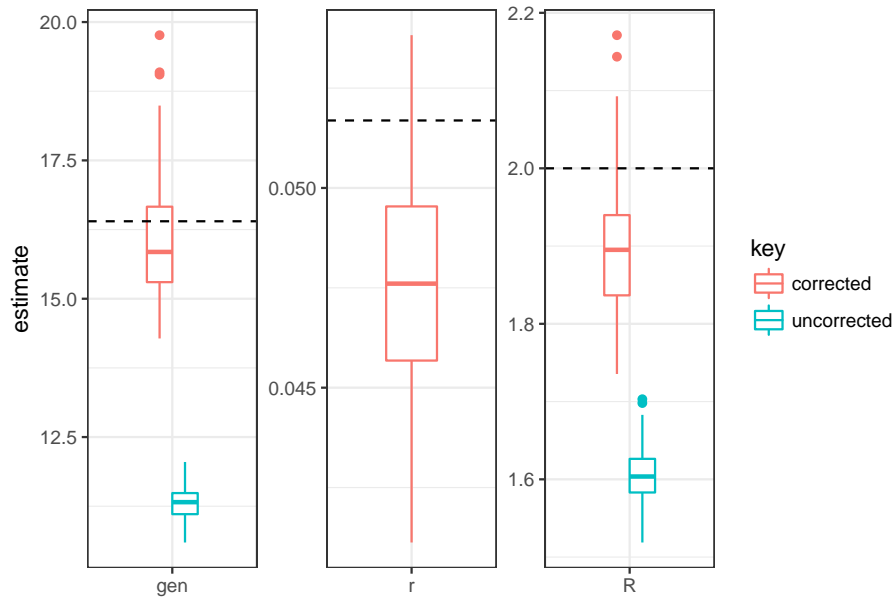


Everything is underestimated. It seems to be driven by underestimation of little $r$ but I'm not sure why (and this is probably affecting coverage). Maybe we want to simulate the epidemic longer? We can see that uncorrected estimates yield much lower mean generation interval (gen) and $\mathcal{R}$. Let's look at the coverage:
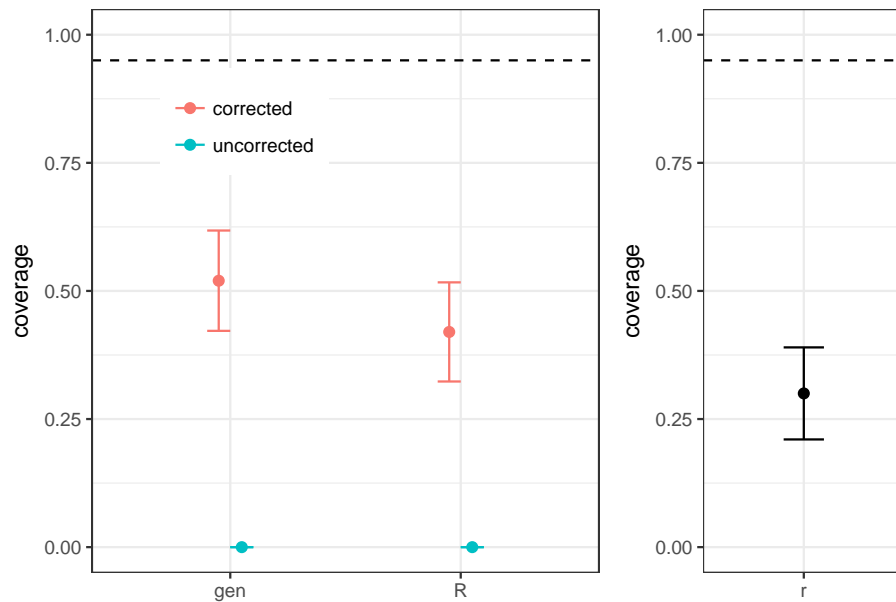


Coverage also sucks. I used to get higher coverage when I was being less

careful... I don't think it's possible to get nominal coverage using a deterministic framework (see King et al. (2015)). I was wondering whether underestimation of little $r$ (and hence its coverage) affects other estimates. So I tried taking simulations where true little $r$ is covered by confidence interval.
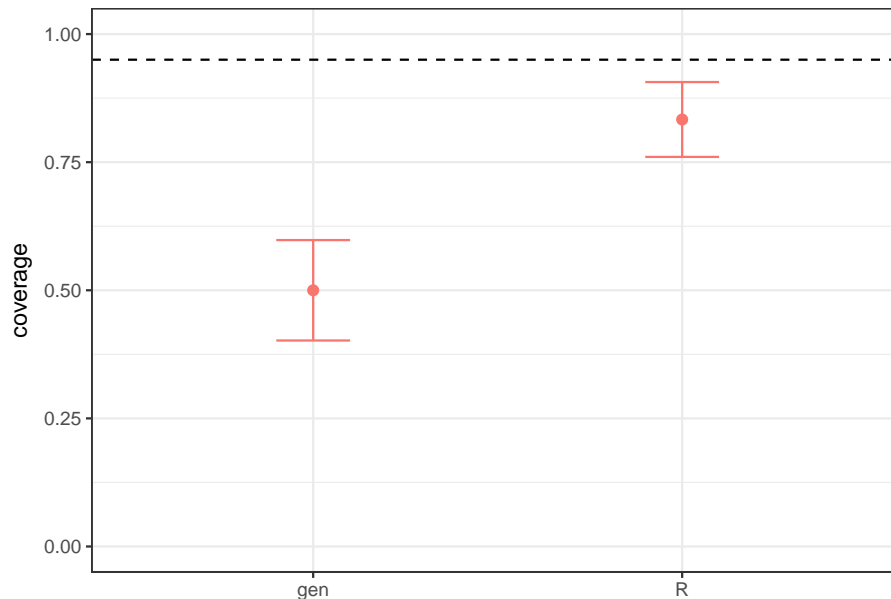


$\mathcal{R}$ coverage increases but mean generation interval coverage seems to stay the same. We need to think of a more statistically robust method that somehow accounts for process error and is simple. I want to stick with the intuition...

# References

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

King, A. A., M. D. de Celles, F. M. Magpantay, and P. Rohani (2015). Avoidable errors in the modelling of outbreaks of emerging pathogens, with special reference to ebola. In *Proc. R. Soc. B*, Volume 282, pp. 20150347. The Royal Society.