On the relation between active population and infection rate of COVID19

Takashi Shimada, Yoshiyuki Suimon, Kiyoshi Izumi
Department of Systems Innovation, Graduate School of Engineering, The University of Tokyo
(Dated: August 19, 2020)

The relation between the number of passengers in the main stations and the infection rate of COVID19 in Tokyo is empirically studied. Our analysis based on conventional compartment model suggests: 1) Average time from the true day of infection to the day the infections are reported is about 15 days. 2) The scaling relation between the density of active population and the infection rate suggests that the increase of infection rate is linear to the active population rather than quadratic, as that is assumed in the conventional SIR model. 3) Notable deviations from the overall scaling relation seems to correspond to the change of the peoples’s behavior in response to the public announcements of action regulation.

I. INTRODUCTION

Reducing the number of people in public area by some means of regulations is one of the most major strategies to slow or prevent the spread of epidemics such as COVID-19. The main drawback of this measure is its negative impact on our economical and social activity. Therefore it is practically very important to quantify the effect of reduction of population density in public area on the reduction of infection rate. It is also essential for wide field including ecology, epidemiology, and sociology to understand the functional relation between the population density and the infection (contact) rate in real case [1–3].

Here we report our empirical study on the relation between the number of passengers in mains stations of Tokyo metropolitan area and the increase in the number of infections.

II. METHOD

A. Calculating effective reproduction number

Let us start from taking a simple compartmental model in the difference equation form:

\[
\begin{align*}
\Delta S_t &= S_{t+1} - S_t = -\beta C(S_t, I_t),
\Delta I_t &= I_{t+1} - I_t = \beta C(S_t, I_t) - \gamma I_t,
\Delta R_t &= R_{t+1} - R_t = \gamma I_t,
\end{align*}
\]

(1)

where \(S_t, I_t, \) and \(R_t\) are the numbers of susceptible, infected, and recovered (removed) people at a discrete time (i.e. day in this study) \(t\), respectively. Positive constant parameters \(\beta\) and \(\gamma\) represent the (effective) infection rates per human contact and the recovery rate, respectively. In the following we take \(\gamma = 14.0^{-1}\) [day], i.e. the average time till recover is 2 weeks, and it will be confirmed later. The function \(C(S_t, I_t)\) denotes the rate of human contacts. For example, taking a Lotka-Volterra type contact rate: \(C \propto S_t I_t\) corresponds to the SIR model.

As shown in Fig. 1 the typical data available is the daily report of newly infected people \(O_t = \Delta S_t\). The delay time \(t_d\) is the difference between the day of the report of an infection and the real moment of infection, which is typically said to be around 2 weeks, consists of the time from the day of infection to the day the person went to a hospital, the time the institution took on PCR-test, and the time spent to include that result to the report, etc. We will estimate \(t_d\) from the data.
The effective reproduction number calculated as
\[
R_t^e = \frac{-\Delta S_t - \Delta R_t}{\Delta R_t} = \frac{O_{t+t_d}}{\gamma \left( \sum_{\tau=t_0}^{t} O_{\tau+t_d} e^{-\gamma(t-\tau)} \right)}
\]  
(2)
is the key quantity to predict whether the epidemic process will grow ($R_t^e > 1$) or decay ($R_t^e < 1$) if the effect of in- and out-flows are negligible.

Note that, because $O_t$ is known to have strong systematic dependence on days of week which comes from the difference in the activity such as testing and reporting, we take $7$ day average for $R_t^e$ in the following analysis.

### B. Estimating the relative number of active people

Another key data for our study is the change in the number of passengers in Tokyo main stations (Shinjuku, Shibuya, Ikebukuro, Shinagawa, and Tokyo), estimated from the mobile phone data (Fig. 1, Appendix). We take the sum of observed passengers $N_t$ to simply estimate the number of people who were not staying home on that day $N_{t,\text{active}}$, i.e. $N_{t,\text{active}} \propto N_t$.

### C. Relating $R_t^e$ and $n_t$

Let $n_t = N_t/\bar{N}$ be the number of people acting outside on a certain day relative to a normal typical average $\bar{N}$. Then the number of people who had a chance to contact should be replaced by the actual active population $\tilde{S}_t = n_t S_t$ and $\tilde{I}_t = n_t I_t$. Therefore, assuming the SIR (Lotka-Volterra) type contact rate $C = C_0 S_t I_t$, the rate of infection and the effective reproduction number should depend proportionally to the square of $n_t$

\[
O_{t+t_d} = -\Delta S_t = (\beta C_0 S_t I_t) n_t^2, \quad (3)
\]
\[
R_t^e = \frac{O_{t+t_d}}{\gamma I_t} \approx \frac{\beta C_0}{\gamma} n_t^2. \quad (4)
\]

In the second equation we assume that the accumulated number of infections is negligibly smaller than the total population, i.e. $1 - S_t/S_0 \approx 0$. This assumption is strongly supported by the random antibody test conducted in early June in Tokyo area, which reported that only 0.10% of people have antibody [9]. Any deviation from this simple scaling relation between $N_t$ and $R_t^e$ implies that the simple SIR-model-type contact rate, which models the chance of human contact based on the well-mixed picture, does not hold in reality.

### III. RESULTS

#### A. Determining the typical delay time from the infection to the detection

The dotted line in Fig. shows the effective reproduction number $R_t^e$ for optimal $t_d = 15$ [day] (solid magenta line, axis to the left), with the number of passengers $N_t$ (dashed cyan line, axis to the right). Both time series are smoothed by taking $7$ day average. Dotted line is the $R_t^e$ with no consideration of delay time (i.e. $t_d = 0$) for comparison. Inset shows the correlation between $R_t^e$ and $t_d$. Vertical lines are the eye-guide for notable days of regulation (solid) or relaxation (dashed) in policy; Day 98 (7th April): the first emergency declaration to limited area including Tokyo, Day 107 (16th April): emergency declaration was canceled for the area except for Hokkaido, Saitama, Chiba, Tokyo, Kanagawa, Kyoto, Osaka, and Hyogo, Day 146 (25th May): emergency declaration was canceled for entire country, Day 154 (2nd June): declaration of “Tokyo-alert”, and Day 163 (11th June): cancelation of “Tokyo-alert”. The reported number of infections are also shown inversely for eye-guide.
is reported. As mentioned above, although it is generally said to be around 2 weeks in the case of Tokyo, we cannot determine this parameter solely from epidemiological knowledge.

To estimate this parameter, we take a natural model-free assumption: Infection rate is an increasing function of active population. Under this assumption, optimal delay time should give the highest rank correlation between $N_t$ and $O_{t+t_d}$ (and hence $R_e^c$). From this criterion, the optimal delay time is found to be $t_d = 15$ [day] (both from Spearman’s and Kendall’s rank correlation coefficients, and for a certain range of $\gamma$. See Appendix for more detail). A good correlation between $N_t$ and $R_e^c$ (Spearman’s rank correlation coefficient $r^* = 0.7815$) can be also easily confirmed in Fig. 2 by eye.

What is interesting, in addition to the overall correlation, is that the days of the changes of regulations policy on the people’s activity coincide with the changes in $R_e^c$. $R_e^c$ shows sharp drops soon after the days of first emergency declaration (to a part of Japan including Tokyo, 7th April) and its application to the entire country (16th April), and also soon after the declaration of “Tokyo-alert” from the Tokyo metropolitan government. Sharp rises are also found soon after the days of relaxing the policies. Those might be indications of the change in the people’s behavior, rather than the change in the population outside.

B. Scaling relation between $R_e^c$ and $n_t$

Having obtained the good estimate for the average delay time $t_d$, we next examine the relation between the active population $N_t$ and the infection rate on that day $R_e^c$. For this analysis, we take the period $t \geq 50$ (from 5th Mar) so that the accumulated number and the daily numbers in it is enough to obtain the estimation of $R_e^c$ in a reasonable range: $< 5.0$. As shown in Fig. 3 $R_e^c$ in this period obeys to an almost single-valued function of the passenger density $n_t$, confirming the correctness of the estimated $t_d$.

The first fitting function to be tested is that of SIR-type (Lotoka-Volterra-type) model LV1:

$$R_{t}^{LV1} = R_0n_t^2$$

This model is based on the well-mixed approximation for the collisions among people, which in this case is consistent with an assumption that the dominant process of the infection is “random collisions” in the street or in some

| Model | $R_0$ | second parameter | deviance |
|-------|-------|------------------|----------|
| LV1:  | 3.045 | -                | 91.90    |
| LV2:  | 2.572 | $n_0 = 0.5183$   | 71.43    |
| DD:   | 2.489 | $\alpha = 0.9558$ | 66.37    |
other public areas.

Keeping the SIR-type contact term, we also try a two-parameter model LV2 which assumes that there is a hidden base population $n_0$ which are active but does not appear in the passenger data:

$$R_t^{LV2} = R_0 \left( \frac{n_0 + n_t}{n_0 + 1} \right)^2. \quad (6)$$

Our last model to be tested is based on a density-dependent (DD) contact rate

$$C(S_t, I_t) = \frac{C_0 S_t I_t}{(S_t + I_t + R_t)^\alpha}. \quad (7)$$

In comparison to the former models, this model can treat more general density-dependence and hence the more complex human activity. For example, if one assume that the major risk of infectious contact is the intended meetings (or co-locating) with a pre-determined member size (e.g. working in the office, dining with friends, shopping, etc), the actual contact rate should become more moderate and hence we expect positive $\alpha$. The exponent $\alpha$ is also expected to be positive if people are trying to take a distance each other to avoid infection. In this sense, $\alpha > 0$ can be regarded as an indication of human behavior actively performed. On the other hand, if people are in a situation that they may be trying to take a distance but sometimes they simply cannot (e.g. in crowded trains, stations, shops, etc), the contact rate becomes steeper and hence the exponent $\alpha$ should be negative. This density-dependent contact rate gives the corresponding fitting function for $R^C$:

$$R_t^{DD} = R_0 n_t^{2-\alpha}. \quad (8)$$

As shown in the Table I and Fig. 3, the density dependent contact rate model (DD) is found to give the best prediction in the least region.

IV. CONCLUSIONS

We have empirically investigated the relation between the effective reproduction number $R_t^e$ the number of passengers in the main stations in Tokyo $n_t$. The delay time from the moment of infection to the day the infection is reported is first estimated robustly at around $t_h = 15$ days, which is consistent with what is generally expected. Based on the estimated delay time, the scaling relation between $R_t^e$ and $n_t$ is examined. The best fit function suggest a density-dependent correction to the rate of human contact, the exponent of which illustrates the relevance of the active aspects of human behavior.

ACKNOWLEDGMENTS

This work was partly supported by JSPS KAKENHI grant number 18K03449TS to TS.