Stochastic Modeling of an Infectious Disease Part V: Maximum Likelihood Estimation of the BDI (birth-and-death-with-immigration) Process Parameters*

Hisashi Kobayashi†
Dept. of Electrical & Computer Engineering
Princeton University
Princeton, NJ 08544, U.S.A.

October 1, 2021

Abstract

In section 1, we clarify the relations among the *basic reproduction number* $\mathcal{R}_0$, the *effective reproduction number*[1] $\mathcal{R}_t$, and the *exponential growth/decay parameter* a, of our stochastic model based on a time-varying BD (birth-death) process. We define the *proportion of vaccinated population* at t, denoted $p(t)$, the *vaccine effectiveness* f_v, and the *behavior factor* $b(t)$, which are introduced to determine the time-varying *secondary infection rate* $\lambda(t)$.

In Section 2, we discuss *maximum likelihood estimation* (MLE) of the model parameters for the time-homogeneous case. Starting with the *likelihood function* given by Wolff [2], we obtain surprisingly simple MLE expressions for the BD parameters, and the exponential growth parameter a. The MLE formulas only involve sample paths of $B(t)$ (the cumulative count of secondary infections up to time t) and $R(t)$ (the cumulative count of the recovered/removed/dead up to t), and of $J(t) = \int_0^t I(u)du$ (where $I(t) = I(0) + B(t) - R(t)$ is the number of infectious individuals at t), all evaluated only at $t = T$, the end of an observation period.

In Section 1.4 and Section 3, we identify some challenges in parameter estimation arising from two important characteristics of Covid-19 and its variants. First, a significant portion of infections are *asymptomatic*. Second, each infection has its *incubation period*, which is a random variable. Symptomatic infections and the incubation period both make it extremely difficult, if not impossible, to obtain a perfect observation of the sample paths of the infection process $I(t)$ or other related processes.

*This result was orally presented at a meeting of the Operations Research Society of Japan on September 29, 2021. For the abstract and slides, see [1].

†The Sherman Fairchild University Professor of Electrical Engineering and Computer Science, Emeritus. Email: hisashi@princeton.edu, homepage: `https://hp.hisashikobayashi.com/`, Wikipedia: `https://en.wikipedia.org/wiki/Hisashi_Kobayashi`

[1]It is also called the *real-time reproduction number*,

Keywords: Basic and effective reproduction numbers, Exponential growth/decay parameter, Model parameter estimation, Regression analysis, Likelihood function, Maximum-likelihood estimation, Fisher information matrix, Cramér-Rao lower bound.

Contents

1 Introduction: Reproduction Numbers and Exponential Growth/Decay Parameter

1.1 Basic and Effective Reproduction Numbers

Throughout our series of reports, our running example adopts the model parameters[2] $\lambda = 0.3, \mu = 0.1$ and $\nu = 0.2$. For the infection size $N(t)$ [persons] at time t, $\lambda N(t)$ is the expected number of newly infected [persons· day^{-1}]. Similarly, the quantity $\mu N(t)$ is the expected number of recovered/removed/dead [persons· day^{-1}] at time t. The quantity μ^{-1}[days] is the expected period for an infected person to remain infectious, and is estimated to be around 10.

The ratio[3]

$$\mathcal{R}_0 = \frac{\lambda}{\mu} \tag{1}$$

represents the expected number of new infections caused, on average, by one infected person during his/her infectious period μ^{-1}[days], and this dimensionless quantity is called the **basic reproduction number**.

Once the infection begins to threaten the well-being of a given community (e.g., a country, a state or a city), the model parameters will change over time, because the government of the given

[2]The units of λ and μ are [day^{-1}], whereas ν is in units of [persons· day^{-1}].

[3]This simple expression assumes that the distribution of the infectious interval is exponentially distributed with the mean μ^{-1}). See Remark 2 for the case of general distribution of the infectious period.

Copyright ©Hisashi Kobayashi 2021

community may take preventive measures by e.g., declaring a state of emergency, or, in an extreme case, by enforcing a total lock-down, which should help reduce the internal infection rate λ towards zero. Hence, we write λ and μ as $\lambda(t)$ and $\mu(t)$, respectively. In the time-varying BDI model we introduce the third parameter $\nu(t)$, which is the rate at which an infectious person enters the community in question. Tightening security control at its borders such as seaports and airports should reduce the value $\nu(t)$.

The rate $\mu(t)$ at which an infected person stops being infectious will increase, by the patient either getting recovered sooner, or being diagnosed sooner and removed from the community and entering a hospital. Once a diagnosed patient enters a hospital, however, the length of the period until his/her getting discharged from the hospital will not directly affect the value of μ.[4] On the other hand, as the hospital gets full and an infected person is forced to stay home or remain untreated, as is happening in some communities in the world, including Tokyo, Japan, the value of μ becomes lower, and thus the average infectious period μ^{-1} increases. Thus, in a real situation, the parameters λ, μ and ν are functions of time. As we noted in Section 2 of [3], the ratio

$$\mathcal{R}_t = \frac{\lambda(t)}{\mu(t)} \tag{2}$$

is called the **effective reproduction number** or **real-time reproduction number**.

Let us define the following factors that determine $\lambda(t)$:

$$p(t) \triangleq \text{Proportion of the population who have been vaccinated by } t,\ 0 \le p(t) \le 1, \tag{3}$$

$$f_v \triangleq \text{Vaccine's effectiveness, i.e.,the probability that a vaccinated person will not be infected.}\ \ 0 \le f_e \le 1, \tag{4}$$

$$b(t) \triangleq \text{Behaving factor achieved by social distancing, etc. by the public,}\ \ 0 \le b(t) \le 1.. \tag{5}$$

As for vaccine's effectiveness, the number f_v differs among different vaccines, see `https://www.yalemedicine.org/news/covid-19-vaccine-comparison`. The assumption of f_e being constant may not hold over time scales of a number of months, due to waning immunity in vaccinated people, and/or emergence of new variants, against which existing vaccines may be less effective. It may be difficult to quantify exactly the population's behavior, but we may expect $b(t)$ to be proportional to the traffic level of people in flows in the public places, and inversely proportional to social distances among the people. We also expect that $b(t)$ will be reduced by the public's mask wearing, gargling and hand-washing. A lock-down policy enforced by the government at some point in time, say at $t = t_1$, will lead to $b(t) \approx 0$ for $t \ge t_1$.

Then we can represent

$$\lambda(t) = \lambda_0 b(t)(1 - f_v p(t)), \tag{6}$$

where $p(0) = 0$, $b(0) = 1$ and $\lambda_0 = \lambda(0) = \lambda$, which is the value of the infection rate under normal circumstances. In our running example of simulating time-varying $\lambda(t)$ defined by (42) of Part II. p. 8, we did not take into account any vaccination program, i.e., $p(t) = 0$ for all t and $b(t) = \lambda(t)/\lambda_0$.

[4]Here we assume that a possible internal infection spread can be contained so that the infections of susceptible populations outside the hospital will not be affected by the occurrence of such a cluster within the hospital. This assumption, however, may not hold in all communities in all cases.

Copyright ©Hisashi Kobayashi 2021

1.2 Exponential Growth/Decay Parameter a

The quantity defined by

$$a(t) = \lambda(t) - \mu(t) = (\mathcal{R}_t - 1)\mu(t) \tag{7}$$

determines the growth or decay rate of $I(t)$ at time t.

Among the various formulas we presented in Part IV [4] are the following formulas for the $\overline{I}(t)$) in **time-homogeneous cases**, for both BD and BDI processes. See ibid., (13), (16) and (19):

$$\overline{I}_{BDI:I_0}(t) = \overline{I}_{BDI:0}(t) + \overline{I}_{BD:I_0}(t), \tag{8}$$

$$\overline{I}_{BDI:0}(t) = \frac{\nu}{a}\left(e^{at} - 1\right), \tag{9}$$

$$\overline{I}_{BD:I_0}(t) = I_0 e^{at}. \tag{10}$$

In all cases $\overline{I}(t)$ grows or decays proportionally to e^{at}, depending on whether $a > 0$ or $a < 0$, respectively.

For **time-nonhomogeneous** cases, the above formulas can be generalized as follows. See ibid., (36), (41) and (30):

$$\overline{I}_{BDI:I_0}(t) = \overline{I}_{BDI:0}(t) + \overline{I}_{BD:I_0}(t), \tag{11}$$

$$\overline{I}_{BDI:0}(t) = N(t)e^{s(t)}, \tag{12}$$

$$\overline{I}_{BD:I_0}(t) = I_0 e^{s(t)}, \tag{13}$$

where

$$s(t) = \int_0^t a(u)\,du = \int_0^t (\lambda(u) - \mu(u))\,du \tag{14}$$

$$N(t) = \int_0^t \nu(u)e^{-s(u)}\,du. \tag{15}$$

In all cases the $\overline{I}(t)$ grows or decays proportionally to $e^{s(t)}$, depending on whether $s(t) > 0$ or $s(t) < 0$, respectively. See [4], p. 8, Figure 6.

Suppose that the function $a(t)$ is nearly constant in the interval $t_1 \le t \le t_2$, i.e.,

$$a(t) \approx a(t_1) \triangleq a_1, \quad \text{for} \quad t_1 \le t \le t_2, \tag{16}$$

Then (13), for instance, becomes

$$\overline{I}_{BD:I_0}(t_1 + D) = I_0 e^{s(t_1+D)} = \overline{I}_{BD:I_0}(t_1)e^{s(t_1+D)-s(t_1)} \approx \overline{I}_{BD:I_0}(t_1)e^{a_1 D}, \quad \text{for} \quad 0 \le D \le t_2 - t_1. \tag{17}$$

Thus, $\overline{I}(t)$ changes by the factor of $e^{a_1 D}$ over D [days] within the interval $[t_1, t_2]$. If $a_1 > 0$ it is an exponential growth, and if $a_1 < 0$, it is an exponential decay.

Note that the above approximation $\approx$ in both (16) and (17) becomes equality in the time-homogeneous case.

Example 1:

Copyright ©Hisashi Kobayashi 2021

Let us consider the time-homogeneous case for simplicity. From the analysis given above, it should be clear the following analysis holds approximately within any interval where $a(t) \approx a$.

If we denote by $\overline{D}_2$ [days] the expected number of days during which, the infected population grows 100% (i.e., doubles) or halves, then $e^{|a|\overline{D}_2} = 2$, i.e.,

$$|a|\overline{D}_2 = \log 2 = 0.6913. \tag{18}$$

In our running example of $\boldsymbol{\mathcal{R}_0} = 3, \mu = 0.1$, and hence $\boldsymbol{a} = 0.2$, $\overline{I}(y)$ takes only 3.45 days to double, or **a week to quadruple**. In an environment with $\boldsymbol{\mathcal{R}_0} = 1.5$ instead of 3.0, then $a = (1.5 - 1) \cdot 0.1 = 0.05$, then it takes **two weeks for $\overline{I}(t)$ to double**. □

1.3 Plausible estimation of the parameter a by regression analysis

In Figures 8-11 of our preceding report [4], we presented the results of 12 simulation runs of a BDI process $I_{BDI:0}(t)$, using the model parameters of our running example, i.e., $\lambda = 0.3, \mu = 0.1$, and $\nu = 0.2$. We showed the first 6 sample paths of $I(t)$ in Figure 8, the next 6 sample paths in Figure 9, and their semi-log plots in Figures 10 and 11, respectively. In these runs, however, the simulator was run up to the 30th day, i.e., $0 \le t \le 30$. We were primarily interested in showing that the initial phase of a BDI process with $I_0 = 0$ can be very haphazard.

In Figures 1 and 2, we present the results of 6 sample paths of the BDI process, with the same set of BDI model parameters stated above, but we ran the simulator up to the 50th day, i.e., $0 \le t \le 50$; the smooth curves in dashed lines are the percentile curves computed from the PMF of this BDI process.

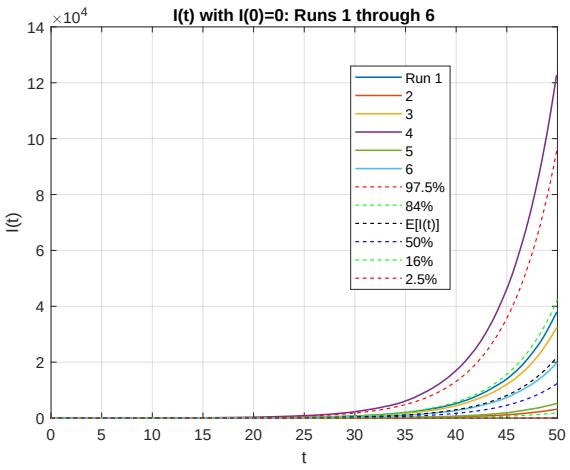

Figure 1: BDI $I(t)$ process, Runs 1-6; $\lambda = 0.3, \mu = 0.1, \nu = 0.2$.

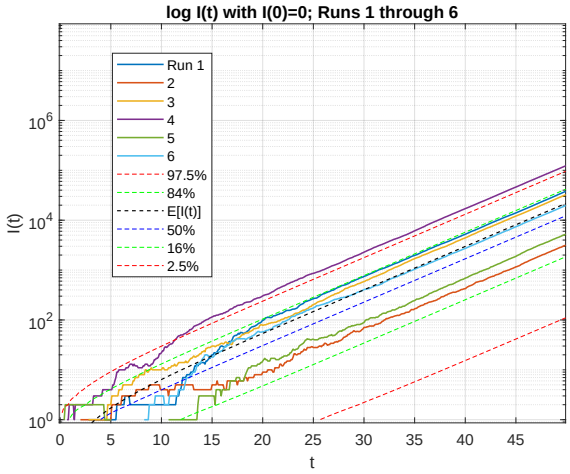

Figure 2: Semi-log plots of Figure 1.

As is clearly seen from Figure 2, a good estimate of the parameter $a = \lambda - \mu$ should be obtainable from the slope of these curves, which are almost parallel after around the 30th day, when the chaotic behavior of the sample paths come to an end, and $I(t)$ has reached the level of ≈ 200. At this point the law of large number seems to set in, and an aggregate of over 200 sub-processes begin to show its regular behavior. Although there exist enormous spreads among different sample paths, we may be able to obtain a reasonable estimate of a from any of the sample paths, once its behavior

Copyright ©Hisashi Kobayashi 2021

becomes more regular than in its initial period, when the path moves rather haphazardly. So one heuristic argument will be to apply a **linear regression** analysis to data points which lie on a straight line. Should we discard the initial portion, say, $0 \leq t \leq 30$?

As we stated earlier, the quantity μ^{-1} represents the **mean infectious interval**. If, for instance, $\hat{\mu^{-1}} \approx 10$ [days], we have $\hat{\mu} \approx 0.1$ [day^{-1}]. Once we have good estimates of a and μ, then we can have an estimate $\hat{\lambda} = \hat{a} + \hat{\mu}$, and thus an estimate of the basic reproduction number $\hat{\mathcal{R}}_0 = \frac{\hat{\lambda}}{\hat{\mu}}$.

We have shown in [5] that the exponential growth (or decay) of the form e^{at} dictates, not only the function $I(t)$, but also $B(t)$, $R(t)$ and even daily new infections, as well (see ibid., pp. 11-13). Then, which data, if available, should we use to make an estimation of a?

1.4 Issues Posed by Covid-19 Data for Estimating Our Model Parameters

An important question, which we have not addressed heretofore, is: "How can we observe a sample path of $I(t)$ from real data of the coronavirus epidemic?" A similar question is: "Can any of real data of the processes $B(t)$, $R(t)$ and other processes such as $J(t)$ (an integrated process defined in (38) of Section 2.2 of this article) be obtained, even partially observed, or approximately estimated? What is readily available seems daily data on new infections, which corresponds to $R(t) - R(t-1)$. Because $\overline{R}(t)$ grows exponentially at rate a, it is reasonable to expect that we should be able to make an estimate from a sample path of the random process $R(t) - R(t-1)$.

There are two major obstacles in obtaining usable data of the corona-virus epidemic, which will make the model parameter estimation quite challenging.

1. First, some percentage of infected people are **asymptomatic**, that is they have no symptoms, and yet contribute to infections. The CDC (Center of Disease Control) of the U.S. is currently assuming 30% of infections are asymptomatic, and asymptomatic people are 75% as infectious as symptomatic people[5]. By imposing PCR and other tests to a larger portion of population, we should be able to improve an estimate of the percentage of these "invisible carriers" of the virus.

2. Second, the **incubation period** (the time from exposure to symptom onset) is thought to be 2 to 14 days, though symptoms typically appear within four or five days after exposure. A person with COVID-19 may be contagious 48 hours before starting to experience symptoms.[6] If we have a reasonable estimate of the probability distribution of this random variable, it should help us obtain better estimates of the model parameter.

What we should do in the presence of intrinsic difficulty or limitations in obtaining "true" and "complete" data, is to understand what should be the right thing to do, if we can obtain, hypothetically, "perfect" data. If we can answer this question, then we should ask ourselves what data should be collected. If we can obtain only imperfect data, what is the best job we can do in estimating the model parameter? Such theoretical results as the EM (expectation-maximization) algorithm and HMM (hidden Markov model) or HSMM (hidden semi-Markov model) may help us in this regard.

[5] I would like to thank Mr. Kaiser Fung for bringing up the CDC link for planning scenarios. `https://www.cdc.gov/coronavirus/2019-ncov/hcp/planning-scenarios.html`.

[6] See `https://www.health.harvard.edu/diseases-and-conditions/if-youve-been-exposed-to-the-coronavirus`.

Copyright ©Hisashi Kobayashi 2021

In the remainder of this article, we develop maximum-likelihood estimation algorithms for the model parameters of the BD or BDI process. By setting aside the above problems associated with real data, we assume that a sample path of $I(t)$ is observable.

2 Maximum-Likelihood Estimation of Model Parameters

2.1 Likelihood Function of a State-Dependent BD Process

As we have already discussed in our previous reports, the random process $B(t)$ is the cumulative count of births (i.e., the count of "infections" in our context) up to time t, and $R(t)$ is the cumulative counts of deaths (i.e., the count of the "recovered/removed/dead"). Their difference

$$I(t) \triangleq I_0 + B(t) - R(t), \tag{19}$$

where $I_0 = I(0)$, represents the number of "infected" (and "infectious" as well) persons at time t.

The birth-and-death process we consider in this section is *time-homogeneous*, but it is somewhat more general than the BD process that we have discussed previously, where the birth and death rates are linear in i , i.e., $i\lambda$ and $i\mu$ when $I(t) = i$.

In this section we assume that both birth and death rates are "state-dependent" [7] in an arbitrary manner, not necessarily linear in $I(t)$, i.e.,

$$\lambda(t) = \lambda_i, \quad \text{and} \quad \mu(t) = \mu_i, \quad \text{when} \quad I(t) = i, \quad i = 0, 1, 2, \ldots. \tag{20}$$

Definition 1 (State-dependent birth-and-death process)**.** *A random process $I(t)$ is called a "state-dependent" birth-and-death process if its transition probabilities*

$$P_{i,j}(h) = \mathsf{P}[I(t+h) = j \,|\, I(t) = i] \tag{21}$$

are independent of t, and satisfy

$$P_{i,j}(h) = \begin{cases} \lambda_i h + o(h), & j = i+1, \\ \mu_i h + o(h), & j = i-1, \\ 1 - (\lambda_i + \mu_i)h + o(h), & j = i, \\ o(h), & |i-j| \geq 2, \end{cases}$$

where $o(h)$ represents a quantity that approaches 0 faster than h as $h \to 0$, i.e., $\lim_{h\to 0} \frac{o(h)}{h} = 0$. □

The model parameter vector $\boldsymbol{\theta}$ we wish to estimate is

$$\boldsymbol{\theta} = (\lambda_0, \mu_1, \lambda_1, \mu_2, \lambda_2, \ldots \mu_i, \lambda_i, \ldots). \tag{22}$$

The parameter $\boldsymbol{\theta}$ will be estimated based on an *instance*, i.e., a *realized sample path* $i(t)$ of the random process $I(t)$, $t \in [0, T]$, denoted $\boldsymbol{i}$, :

$$\boldsymbol{i} \triangleq \{i(t), \quad 0 \leq t \leq T\}. \tag{23}$$

[7] By "state" we simply mean here the integer value of $I(t)$, which is a Markov process as defined in (1).

Copyright ©Hisashi Kobayashi 2021

Let $i_0 = i(0)$ be the initial state, and let t_n, for $n = 1, 2, 3, \ldots$, be the time, at which the nth state transition occurs and enters the state i_n. The sequence $\{(0, i_0), (t_1, i_1), (t_2, i_2), \ldots, (t_N, i_N)\}$, where $N = \max\{n : t_n < T\}$, completely specifies the sample path $\boldsymbol{i}$. Note that N is a random variable, i.e., it varies from one sample path to another.

Consider successive transition instants t_n and t_{n+1}, where the system enters state i_n and i_{n+1}, and let the interval between them be denoted $\tau_n = t_{n+1} - t_n$. In other words, the length of stay in state i_n is τ_n. The transition probability associated with the transition "$i_n \to i_{n+1}$" is given by[8]

$$f_{i_n, i_{n+1}}(\tau_n) = \begin{cases} \lambda_{i_n} e^{-(\lambda_{i_n} + \mu_{i_n})\tau_n} & \text{if "up" transition, i.e.,} \quad i_{n+1} = i_n + 1, \\ \mu_{i_n} e^{-(\lambda_{i_n} + \mu_{i_n})\tau_n} & \text{if "down" transition, i.e.,} \quad i_{n+1} = i_n - 1. \end{cases} \tag{24}$$

Then the probability of having a sample path $\boldsymbol{i}$ is given by the initial distribution of i_0, $P_{I(0)}(i_0)$, multiplied by all the transition probabilities. Thus, the likelihood function for $\boldsymbol{\theta}$ given $\boldsymbol{i}$ is

$$L_{\boldsymbol{i}}(\boldsymbol{\theta}) = \mathsf{P}[\boldsymbol{I} = \boldsymbol{i} \,|\, \boldsymbol{\theta}] = P_{I(0)}(i_0) \prod_{n=0}^{N-1} f_{i_n, i_{n+1}}(\tau_n). \tag{25}$$

The last two equations lead to the following proposition:

Proposition 1 (Log-likelihood function of a state dependent BD process)**.** *For the parameter $\boldsymbol{\theta}$ defined by (22), the natural logarithm of the likelihood function, called the* log-likelihood function*, $\log L_{\boldsymbol{i}}(\boldsymbol{\theta})$ is given by*

$$\boxed{\log L_{\boldsymbol{i}}(\boldsymbol{\theta}) = \log P_{I(0)}(i_0) + \sum_{i \geq 0} u_i \log \lambda_i + \sum_{i \geq 1} d_i \log \mu_i - \sum_{i \geq 1} \gamma_i (\lambda_i + \mu_i),} \tag{26}$$

where u_i is the number of upward transitions from state i, d_i is the number of downward transitions from state i and γ_i is the total time spent in state i during $[0, T]$.

Proof. The natural logarithm of (25) yields

$$\log L_{\boldsymbol{i}}(\boldsymbol{\theta}) = \log P_{I(0)}(i_0) + \sum_{n=0}^{N-1} \log f_{i_n, i_{n+1}}(\tau_n). \tag{27}$$

Since there are N transitions in the sample path $i(t)$ in $[0, T)$, it is apparent that

$$\sum_{i=0}^{\infty} u_i + \sum_{i=1}^{\infty} d_i = N. \tag{28}$$

Then we can rewrite (27), using (26), as

$$\log L_{\boldsymbol{i}}(\boldsymbol{\theta}) = \log P_{I(0)}(i_0) + \sum_{i=0}^{\infty} u_i \log \lambda_i + \sum_{i=1}^{\infty} d_i \log \mu_i - \sum_{i=0}^{\infty} \gamma_i (\lambda_i + \mu_i). \tag{29}$$

□ □

Wolff [2], who studied the birth-death process in the context of queuing models discusses the idea of the above expression, attributes it to Billingsley [6].

[8] The next transition is either "up" or "down," so the system leaves the state i_n at rate $\gamma_n = \lambda_n + \mu_n$. The interval until the next transition is exponentially distributed with the probability density function (PDF) $f(\tau) = \gamma_n e^{-\gamma_n \tau}$. The probability that the transition at $t_{n+1} (= t_n + \tau)$ is an upward transition is $\frac{\lambda_n}{\gamma_n}$ and the probability of a downward transition is $\frac{\mu_n}{\gamma_n}$.

Copyright ©Hisashi Kobayashi 2021

2.2 Maximum-Likelihood Estimate of λ and μ in the Time-Homogeneous BD Process

Now let us consider a special case of the above state-dependent birth-death process, where they take the following form, as assumed in the BD process discussed in Parts I [3] and II [5] of our previous reports:

$$\lambda_i = i\lambda,\ i \geq 0, \quad \text{and} \quad \mu_i = i\mu,\ i \geq 1. \tag{30}$$

Now the the parameter vector $\boldsymbol{\theta}$ reduces to the following two:

$$\boldsymbol{\theta} = (\lambda, \mu), \tag{31}$$

By noting that $\lambda_0 = 0$, as seen from (30), we delete the $i = 0$ term in (26), arriving at the following expression:

$$\log L_{\boldsymbol{i}}(\boldsymbol{\theta}) = \log P(i_0) + \sum_{i\geq 1} u_i \log(i\lambda) + \sum_{i\geq 1} d_i \log(i\mu) - \sum_{i\geq 1} \gamma_i i(\lambda + \mu), \tag{32}$$

which can be rewritten as

$$\log L_{\boldsymbol{i}}(\boldsymbol{\theta}) = \log\lambda \sum_{i\geq 1} u_i + \log\mu \sum_{i\geq 1} d_i - (\lambda + \mu) \sum_{i\geq 1} i\gamma_i + C, \tag{33}$$

where C aggregates the terms that do not depend on $\boldsymbol{\theta} = (\lambda, \mu)$:

$$C = \log P_{I(0)}(i_0) + \sum_{i\geq 0} u_i \log i + \sum_{i\geq 1} d_i \log i. \tag{34}$$

Note that the total number of upward transitions $\sum_{i\geq 1} u_i$ is equal to the total number of births (i.e., internal infections in our context) in $\boldsymbol{i} = \{i(t); 0 \leq t < T\}$. If we denote by $\boldsymbol{b} = \{b(t); 0 \leq t < T\}$, a sample path of the random process $B(t)$ (see (19)), associated with the sample path $\boldsymbol{i}$, then we find

$$\sum_{i\geq 1} u_i = b(T). \tag{35}$$

Similarly, the total number of downward transitions can be expressed as

$$\sum_{i\geq 1} d_i = r(T), \tag{36}$$

where $\boldsymbol{r} = \{r(t); t_0 \leq t < T\}$ is the sample path, associated with the $\boldsymbol{i}$, of the random process $R(t)$.

Note that the term in (33), $\sum_{i\geq 1} i\gamma_i$ [persons·days], represents the **total amount of time spent by the infected persons** who remain infectious (until recovery, removal or death). It is not difficult to see that this quantity is equal to the **total area** below the sample path function $i(t), 0 \leq t < T$, namely[9]

$$\sum_{i\geq 1} i\gamma_i = \int_0^T i(t)\, dt, \tag{37}$$

[9]We may interpret the LHS (left-hand side) of (37) as the *Lebesgue integral* of the function of $i(t)$, whereas the RHS (right-hand side) is the Riemann integral.

Copyright ©Hisashi Kobayashi 2021

By defining a random process $J(t)$[10] by

$$J(t) \triangleq \int_0^t I(u)\, du, \tag{38}$$

and denoting its sample path as $j(t) = \int_0^t i(u)\, du$, we find

$$\sum_{i \geq 1} i\gamma_i = j(T). \tag{39}$$

This $j(t)$ can alternatively represented, using (19), as

$$j(t) = \int_0^t b(u)\, du - \int_0^t r(u)\, du. \tag{40}$$

The gradient[11] of the $\log L_{\boldsymbol{i}}(\boldsymbol{\theta})$, which is called the *score function*, denoted $\boldsymbol{s}(\boldsymbol{i};\boldsymbol{\theta})$, can be computed for our problem at hand, from (33), (35), (36) and (39), as follows:

$$\begin{aligned} \boldsymbol{s}(\boldsymbol{i};\boldsymbol{\theta}) = \nabla_{\boldsymbol{\theta}} \log L_{\boldsymbol{i}}(\boldsymbol{\theta}) &\triangleq \left(\frac{\partial \log L_{\boldsymbol{i}}(\boldsymbol{\theta})}{\partial \lambda}, \frac{\partial \log L_{\boldsymbol{i}}(\boldsymbol{\theta})}{\partial \mu} \right) \\ &= \left(\frac{b(T)}{\lambda} - j(T), \frac{r(T)}{\mu} - j(T) \right). \end{aligned} \tag{41}$$

The score function represents the rate at which $\log L_{\boldsymbol{i}}(\boldsymbol{\theta})$ change as $\boldsymbol{\theta}$ changes. By setting the score function to zero, we find a stationary point $\boldsymbol{\theta}^*$:

$$\boldsymbol{\theta}^* = (\lambda^*, \mu^*) = \left(\frac{b(T)}{j(T)}, \frac{r(T)}{j(T)} \right). \tag{42}$$

In order to determine whether the stationary point is a local maximum point, we need to compute the Hessian matrix of $\log L_{\boldsymbol{i}}(\boldsymbol{\theta})$ at $\boldsymbol{\theta}_0$.

$$\boldsymbol{H}(\boldsymbol{i};\boldsymbol{\theta}^*) \triangleq \nabla_{\boldsymbol{\theta}} \boldsymbol{s}^\top(\boldsymbol{i},\boldsymbol{\theta}^*) = \nabla_{\boldsymbol{\theta}} \nabla_{\boldsymbol{\theta}}^\top \log L_{\boldsymbol{i}}(\boldsymbol{\theta}^*) = \left[\frac{\partial^2 \log L_{\boldsymbol{i}}(\boldsymbol{\theta}^*)}{\partial \theta_i \partial \theta_j} \right]_{M \times M} \tag{43}$$

For our problem, the Hessian matrix is a 2×2 matrix:

$$\boldsymbol{H}(\boldsymbol{i};\boldsymbol{\theta}^*) = \begin{bmatrix} \frac{\partial^2 \log L_{\boldsymbol{i}}(\boldsymbol{\theta}^*)}{\partial \lambda^2} & \frac{\partial^2 \log L_{\boldsymbol{i}}(\boldsymbol{\theta}^*)}{\partial \lambda \partial \mu} \\ \frac{\partial^2 \log L_{\boldsymbol{i}}(\boldsymbol{\theta}^*)}{\partial \lambda \partial \mu} & \frac{\partial^2 \log L_{\boldsymbol{i}}(\boldsymbol{\theta}^*)}{\partial \mu^2} \end{bmatrix} = - \begin{bmatrix} \frac{i(T)^2}{b(T)} & 0 \\ 0 & \frac{i(T)^2}{r(T)} \end{bmatrix}, \tag{44}$$

which is a negative definite matrix. Thus, we have shown that $\boldsymbol{\theta}^* = (\lambda^*, \mu^*)$ is at least a locally maximum point. It is not difficult to see by examining the function (33) that the estimate $\boldsymbol{\theta}^*$ is globally maximum as well. Thus, we have the following:

[10] Note that $J(t)$ is a non-negative valued continuous-time function, whereas $I(t)$ is a stair-case function, taking on only non-negative integers.

[11] The symbol ∇, called "nabla", represents the gradient operator. If $\boldsymbol{\theta}$ is an M-dimensional row vector, $\nabla_{\boldsymbol{\theta}} f(\boldsymbol{\theta})$ is also an M-dimensional row vector, whose ith element is $\frac{\partial f(\boldsymbol{\theta})}{\partial \theta_i}, i = 1, 2, \ldots, M$.

Copyright ©Hisashi Kobayashi 2021

Proposition 2 (Maximum-likelihood estimates of λ and μ of the BD process)**.** *The log-likelihood function of the time-homogeneous BD process defined by (30), is given as*

$$\log L_{\boldsymbol{i}}(\boldsymbol{\theta}) = b(T)\log\lambda + r(T)\log\mu - j(T)(\lambda+\mu) + C. \tag{45}$$

A maximum likelihood estimate (MLE) of the parameter $\boldsymbol{\theta} = (\lambda, \mu)$ is given by

$$\boxed{\hat{\boldsymbol{\theta}}_{MLE} = \boldsymbol{\theta}^* = (\lambda^*, \mu^*) = \left(\frac{b(T)}{j(T)}, \frac{r(T)}{j(T)}\right).} \tag{46}$$

□

Keiding [7] obtained the same result, also based on Wolff [2], whose log-likelihood function (26) is the starting point of our derivation.

The above expressions for $\hat{\lambda}_{MLE}$ and $\hat{\mu}_{MLE}$ are not entirely unexpected. Recall that the expected values $\overline{B}(t)$ and $\overline{R}(t)$ are related to $\overline{I}(t)$ by (see Part I [3], (27) and (30)):

$$\frac{d\overline{B}(t)}{dt} = \lambda\overline{I}(t), \quad \frac{d\overline{R}(t)}{dt} = \mu\overline{I}(t). \tag{47}$$

By integrating the above equations from $t = 0$ to T, we obtain

$$\overline{B}(T) = \lambda\int_0^T \overline{I}(u)\,du = \lambda\overline{J}(T) \tag{48}$$

$$\overline{R}(T) = \mu\int_0^T \overline{I}(u)\,du = \mu\overline{J}(T). \tag{49}$$

from which we find

$$\lambda = \frac{\overline{B}(T)}{\overline{J}(T)}, \quad \text{and} \quad \mu = \frac{\overline{R}(T)}{\overline{J}(T)}. \tag{50}$$

The negative of the Hessian matrix (44)

$$\boxed{\boldsymbol{\mathcal{J}}(\boldsymbol{i};\boldsymbol{\theta}^*) = -\boldsymbol{H}(\boldsymbol{i};\boldsymbol{\theta}^*)} \tag{51}$$

is called the **observed Fisher information matrix**.

Thus, the above $\hat{\boldsymbol{\theta}}_{MLE}$, which is the maximum-likelihood estimate (MLE), if it gives a global maximum of $L_{\boldsymbol{i}}(\boldsymbol{\theta})$, can be written as

$$\hat{\boldsymbol{\theta}}_{MLE} \approx \boldsymbol{\theta}^* + \boldsymbol{\mathcal{J}}(\boldsymbol{i};\boldsymbol{\theta}^*)^{-1}\boldsymbol{s}(\boldsymbol{i};\theta^*). \tag{52}$$

It is easy to show that the expectation of the score function is zero for any $\boldsymbol{\theta}$, i.e.,

$$\mathsf{E}[\boldsymbol{s}(\boldsymbol{I};\theta^*)] = \boldsymbol{0}. \tag{53}$$

The matrix $\boldsymbol{\mathcal{J}}(\boldsymbol{I};\boldsymbol{\theta})$ is also a random variable and its expectation

$$\boxed{\boldsymbol{\mathcal{I}}(\boldsymbol{I};\boldsymbol{\theta}^*) = -\boldsymbol{H}(\boldsymbol{i};\boldsymbol{\theta}^*)} \tag{54}$$

Copyright ©Hisashi Kobayashi 2021

is called the **Fisher information matrix**.

Since the MLE obtained is asymptotically unbiased, we have the following important result, widely known as the *Cramér-Rao lower bound* in statistical estimation theory (see e.g. [8], pp. 532-536, and 551):

Theorem 1 (**Cramér-Rao Lower Bound (CRLB)**). *Let $\hat{\boldsymbol{\theta}}(\boldsymbol{x})$ be any unbiased estimator of $\boldsymbol{\theta}$. Then the following properties hold.*

1. *The variance matrix of $\hat{\boldsymbol{\theta}}(\boldsymbol{X})$ is bounded from below by the inverse of the Fisher information matrix:*

$$\boxed{\mathbf{Var}[\hat{\boldsymbol{\theta}}(\boldsymbol{X})] = E\left[(\hat{\boldsymbol{\theta}}(\boldsymbol{X}) - \boldsymbol{\theta})(\hat{\boldsymbol{\theta}}(\boldsymbol{X}) - \boldsymbol{\theta})^\top\right] \geq \boldsymbol{\mathcal{I}}_{\boldsymbol{x}}^{-1}(\boldsymbol{\theta}).} \tag{55}$$

2. *The lower bound is attained, if and only if $\hat{\boldsymbol{\theta}}(\boldsymbol{X})$ satisfies the following equation*

$$\boldsymbol{s}(\boldsymbol{X};\boldsymbol{\theta}) = \boldsymbol{\mathcal{I}}(\boldsymbol{\theta})(\hat{\boldsymbol{\theta}}(\boldsymbol{X}) - \boldsymbol{\theta}), \tag{56}$$

where $\boldsymbol{s}(\boldsymbol{X};\boldsymbol{\theta})$ is defined by (41).

Proof. See e.g., [8], pp. 533-534. □

Returning to our problem at hand, we find the Fisher information matrix is given by

$$\boldsymbol{\mathcal{I}}(\boldsymbol{I};\boldsymbol{\theta}^*)] = \begin{bmatrix} E\left[\frac{J^2(T)}{B(T)}\right] & 0 \\ 0 & E\left[\frac{J^2(T)}{R(T)}\right] \end{bmatrix}, \tag{57}$$

2.3 Maximum-Likelihood Estimate of the Exponential Growth Parameter a

From Proposition 2, we can obtain the following corollary:

Corollary 1 (Maximum-likelihood estimate of $a = \lambda - \mu$ of the BD process). *A maximum likelihood estimate (MLE) of the parameter $a = \lambda - \mu$ of the time-homogeneous BD process is given by*

$$\boxed{\hat{a}_{MLE} = a^* = \frac{i(T) - I_0}{j(T)}.} \tag{58}$$

Proof. Because of the MLE estimates of λ and μ found in Proposition 2, we readily have

$$a^* = \lambda^* - \mu^* = \frac{i(T) - I_0}{j(T)}, \tag{59}$$

where we used the identity $I(t) = B(t) - R(t) + I_0$, i.e., $b(t) - r(t) = i(t) - I_0$. Because $j(t) = \int_0^t i(u)\,du$, taking its logarithm and differentiating it, we obtain

$$\left.\frac{d\log j(t)}{dt}\right|_{t=T} = \frac{i(T)}{j(T)} = a^* + \frac{I_0}{j(T)}. \tag{60}$$

□ □

Copyright ©Hisashi Kobayashi 2021

This corollary follows naturally from Proposition 2, but it is fair to say that the formula (59) cannot be conceived from any of the semi-log plots presented earlier. This is a good example that a formal theory will lead to a simpler solution than an intuitive ad-hoc solution.

But, the formula (59) could have been conceived, had we considered $\overline{I}(t)$ and its integration $\overline{J}(t)$,

$$\overline{I}(t) - I_0 = I_0(e^{at} - 1), \tag{61}$$

$$\overline{J}(t) = \frac{I_0(e^{at} - 1)}{a}. \tag{62}$$

By taking the ratio of these, we readily arrive at the formula (59).

Implications of Proposition 2 and Corollary 1

The simple expressions we have found for maximum-likelihood estimates of λ, μ and a, given by Proposition 2 and Corollary 1 are amazingly beautiful. In hindsight these formulas seem obvious, but they would not come to our mind, until we formulate the problem mathematically. We would not think of integrating $I(t)$ and introduce the function $J(t)$, unless we go through a mathematical analysis, and this is an excellent example which demonstrates the beauty and value of mathematical analysis, which leads us to a simple and yet non-obvious result. The key is to introduce the process $J(t)$. The theory of MLE then showed us that we do no need to observe $I(t), J(t), B(t)$, or $R(t)$ at all t. We simply need to observe the final values of sample paths $b(t), r(t)$ and $j(t)$ at the end of our observation period. One could ask: Are there any other simpler solutions? Curve fitting techniques such as regression analysis are more involved.

Another important implication is that we do not need to discard data in the initial period, which we thought might adversely affect the quality of estimation when we apply linear regression to the semi-log plot such as Figure 2. Quite the contrary, the theory shows that the information contained in the initial (and seemingly unreliable) data is as important as data in the later stable period of the sample paths.

Two more remarks regarding the MLE of the parameter $a = \lambda - \mu$. We did not provide a lower bound of this estimate, because we need to find out the *correlation coefficient* between the processes $B(t)$, $R(t)$ and $J(t)$. Unless these processes are shown to be uncorrelated, which seems untrue, we cannot treat estimates of λ and μ as being uncorrelated to each other. In a forthcoming report [9] we shall investigate the processes $B(t)$, $R(t)$ and $J(t)$, including their *covariance functions*.

If we define a parameter $s \triangleq \lambda + \mu$, it is apparent that we can obtain its MLE by

$$s^* = \lambda^* + \mu^* = \frac{s(T)}{j(T)}, \tag{63}$$

where $s(t)$[12] is a sample path of $S(t) = B(t) + R(t)$. If the processes $B(t)$ and $R(t)$ are negatively correlated, which we believe is the case, then an estimate $\hat{s}$ should have a smaller variance than that of $\hat{a}$. Hence, if $\hat{\mu}$ is known or available from other means, then $\hat{s}$ should be used to come up with an estimate of λ, i.e., $\hat{\lambda} = \hat{s} - \hat{\mu}$.

[12]This function has no relation to $s(t) = \int_0^t a(u)\,du$, which we introduced earlier, nor the score function $\boldsymbol{s}(t)$ in the preceding section.

Copyright ©Hisashi Kobayashi 2021

2.4 Maximum Likelihood Estimate of λ, μ and ν of the Time-Homogeneous BDI Process

How can we extend the previous results obtained for the BD process to a more general case, i.e., the BDI process which allows immigration? Before we undertake a potentially difficult task to answer this question, we should ponder how important it is, in practice, to consider this case, given that the effect of immigration should be rather insignificant, once the the number of the infected $I(t)$ has reached a certain level. As we have seen in the analysis and many simulation experiments, the immigrants and their descendants (i.e., arrivals of infected individuals from outside, and secondary infections, etc.) play significant roles in the initial phase of an epidemic. With this observation in mind we will investigate below the parameter estimation problem of a BDI process.

By specializing the state-dependent birth-and-death process (see Definition 1) such that

$$\lambda_i = i\lambda + \nu,\ i \geq 0 \quad \text{and} \quad \mu_i = i\mu,\ i \geq 1, \tag{64}$$

we can represent the BDI process model with the internal infection rate λ, the external arrival rate ν of the infected, and the recovery rate μ. The model parameter vector $\boldsymbol{\theta}$ is defined by

$$\boldsymbol{\theta} = (\lambda, \nu, \mu), \tag{65}$$

The log-likelihood function of (26) now becomes

$$\begin{aligned}
\log L_{\boldsymbol{i}}(\boldsymbol{\theta}) &= \log P_{I(0)}(i_0) + \sum_{i\geq 0} u_i \log(i\lambda + \nu) + \sum_{i\geq 1} d_i \log(i\mu) - \sum_{i\geq 0} \gamma_i [i(\lambda+\mu) + \nu], \\
&= \sum_{i\geq 0} u_i \log(i\lambda + \nu) + \log\mu \sum_{i\geq 1} d_i - (\lambda+\mu)\sum_{i\geq 1} i\gamma_i - \nu \sum_{i\geq 0}\gamma_i + C',
\end{aligned} \tag{66}$$

where

$$C' = \log P(x_0) + \sum_{i\geq 0} d_i \log i. \tag{67}$$

The gradient of the $\log L_{\boldsymbol{i}}(\boldsymbol{\theta})$ is

$$\begin{aligned}
\boldsymbol{s}(\boldsymbol{i}, \boldsymbol{\theta}) &= \left(\frac{\partial \log L_{\boldsymbol{i}}(\boldsymbol{\theta})}{\partial \lambda}, \frac{\partial \log L_{\boldsymbol{i}}(\boldsymbol{\theta})}{\partial \nu}, \frac{\partial \log L_{\boldsymbol{i}}(\boldsymbol{\theta})}{\partial \mu} \right) \\
&= \left(\sum_{i\geq 1} \frac{iu_i}{i\lambda+\nu} - j(T),\ \ \sum_{i\geq 0} \frac{u_i}{i\lambda+\nu} - T,\ \ \frac{r(T)}{\mu} - j(T) \right),
\end{aligned} \tag{68}$$

where we used $\sum_{i\geq 0}\gamma_i = T$.

By setting this score function equal to zero, we find that the formula for μ^* is the same as that in the BD model as given by (42). The λ^* and ν^* do not have simple explicit computation formulas, but should be uniquely determined from the following two equations:

$$\boxed{\sum_{i\geq 1} \frac{iu_i}{i\lambda^* + \nu^*} = j(T), \quad \text{and} \quad \sum_{i\geq 0} \frac{u_i}{i\lambda^* + \nu^*} = T.} \tag{69}$$

Copyright ©Hisashi Kobayashi 2021

In order to show that the stationary point θ^* is a local maximum point, we need to compute the Hessian matrix of $\log L_{\boldsymbol{i}}(\boldsymbol{\theta})$ at $\boldsymbol{\theta}^*$.

$$\begin{aligned}\boldsymbol{H}(\boldsymbol{i};\boldsymbol{\theta}^*) &= \begin{bmatrix} \frac{\partial^2 \log L_{\boldsymbol{i}}(\boldsymbol{\theta}^*)}{\partial \lambda^2} & \frac{\partial^2 \log L_{\boldsymbol{i}}(\boldsymbol{\theta}^*)}{\partial \lambda \partial \nu} & \frac{\partial^2 \log L_{\boldsymbol{i}}(\boldsymbol{\theta}^*)}{\partial \lambda \partial \mu} \\ \frac{\partial^2 \log L_{\boldsymbol{i}}(\boldsymbol{\theta}^*)}{\partial \nu \partial \lambda} & \frac{\partial^2 \log L_{\boldsymbol{i}}(\boldsymbol{\theta}^*)}{\partial \nu^2} & \frac{\partial^2 \log L_{\boldsymbol{i}}(\boldsymbol{\theta}^*)}{\partial \nu \partial \mu} \\ \frac{\partial^2 \log L_{\boldsymbol{i}}(\boldsymbol{\theta}^*)}{\partial \mu \partial \lambda} & \frac{\partial^2 \log L_{\boldsymbol{i}}(\boldsymbol{\theta}^*)}{\partial \mu \partial \nu} & \frac{\partial^2 \log L_{\boldsymbol{i}}(\boldsymbol{\theta}^*)}{\partial \mu^2} \end{bmatrix} \\ &= -\begin{bmatrix} \sum_{i\geq 1} \frac{i^2 u_i}{(i\lambda^*+\nu^*)^2} & \sum_{i\geq 1} \frac{i u_i}{(i\lambda^*+\nu^*)^2} & 0 \\ \sum_{i\geq 1} \frac{i u_i}{(i\lambda^*+\nu^*)^2} & \sum_{i\geq 0} \frac{u_i}{(i\lambda^*+\nu^*)^2} & 0 \\ 0 & 0 & \frac{r(T)}{\mu^{*2}} \end{bmatrix}, \end{aligned} \tag{70}$$

which is a negative definite matrix. Thus, the stationary point $\boldsymbol{\theta}^*$ provides an MLE.

If we can assume that $\frac{\nu^*}{\lambda^*} < 1$, the following approximation might be useful in obtaining an approximate MLE of λ and ν, or at least in setting an initial estimate of an iterative computation to obtain λ^* and ν^*. The LHS of the first equation of (69) becomes

$$\begin{aligned}\sum_{i\geq 1} \frac{i u_i}{i\lambda^*+\nu^*} &= \sum_{i\geq 1} \frac{u_i}{\lambda^*\left(1+\frac{\nu^*}{i\lambda^*}\right)} = \frac{\sum_{i\geq 1} u_i}{\lambda^*} - \frac{\nu^*}{\lambda^{*2}} \sum_{i\geq 1} \frac{u_i}{i} + o\left(h^2\right)\Big) \\ &= \frac{r(T)}{\lambda^*} - \frac{\nu^* v(T)}{\lambda^{*2}} + o(h^2), \end{aligned} \tag{71}$$

where $h = \frac{\nu^*}{i\lambda^*} < 1$ with $i \geq 1$, and

$$v(T) \triangleq \sum_{i\geq 1} \frac{u_i}{i}. \tag{72}$$

Similarly, the LHS of the second equation in (69) can be expressed as

$$\sum_{i\geq 0} \frac{u_i}{i\lambda^*+\nu^*} = \frac{u_0}{\nu^*} + \frac{v(T)}{\lambda^*} - \frac{\nu^* w(T)}{\lambda^{*2}} + o(h^2), \tag{73}$$

where

$$w(T) \triangleq \sum_{i\geq 1} \frac{u_i}{i^2}. \tag{74}$$

The two equations still require solving 3rd degree of equations.

3 Concluding Remarks

1. As we discussed in Section 1.4, the incubation period makes it impossible to obtain perfect information of a sample path of the process $I(t)$. What is collected and reported daily as "new infections of today" is in fact data on the count of terminations of the infectious periods of infected persons. It does not really count new infections, because the infections took place some days earlier, but a symptom appeared and was confirmed by a test after the incubation period, which is a random variable. If we can assume that the percentage of

Copyright ©Hisashi Kobayashi 2021

asymptomatic cases is negligibly small, or a fixed constant, the "data of daily new infections" can be interpreted as *imperfect information* on $R(t+\delta)-R(t-1+\delta)$, which is an incremental increase of $R(t)$ since one day ago. Here δ is determined by the specific time of the day when the data is recorded. For example, if it is 5 p.m. then $\delta = \frac{17}{24}$.

2. The formula (7) is obtained under the Markovian assumption that the infectious period is exponentially distributed with mean μ^{-1}. The distribution of the infectious period, which is intertwined with the distribution of the incubation period, depends on how widely among the general public and how frequently PCR and other tests are performed. The shape of the infectious period distribution may be far from being exponential, in which case the formula for the basic reproduction number (1) and the relation between the growth/decay parameter a and the basic reproduction number given by (7) may need to be revised, although the present author is of the opinion that the above formulas should be rather insensitive to the distribution shape.

3. We will evaluate the above sensitivity issue by using the **Euler-Lotka equation**, which is well known in the study of age-structured population models [10]. The k-stage Erlang distribution (where $k = 1, 2, \ldots, \infty$) as parametric distributions of the infectious period will be used for a systematic evaluation of different distribution forms.

4. Thus far, we have primarily been concerned with the Markov process $I(t)$. But in view of the MLE expressions for λ and μ, it is important to better understand the processes $B(t), R(t)$ and the integrated process $J(t)$. We have done some preliminary investigations, and a report is in preparation [9].

5. The analytic solutions obtained for the processes $B(t)$ and $R(t)$ are in terms of the PGF (probability generating function), and $J(t)$ in terms of the Laplace transform. Unlike the PGF for Markov process $I(t)$, however, it does not seem feasible to obtain an exact closed-form expression for the PMF (probability mass function) from either of the PGFs. Thus, our approach is to find a PDF (probability density function) which can approximate the PMF by using the **saddle-point integration** [11].

6. The EM (expectation and maximization) algorithm has been used as a primary tool in obtaining maximum likelihood estimates from imperfect or missing observation data. It would be worthwhile to explore this approach to our model parameter estimation.

Acknowledgments

The author thanks Dr. Linda Zeger, Prof. Brian L. Mark and Mr. Kaiser Fung, for their careful review of the manuscripts and for many valuable suggestions to improve the presentation.

References

[1] H. Kobayashi, "A Probabilistic Model of an Infectious Disease, based on the BDI Process, and Maximum-Likelihood Estimation of the Model Parameters." `http://hp.hisashikobayashi.com`, September 29th 2021.

Copyright ©Hisashi Kobayashi 2021

[2] R. W. Wolff, "Problems of Statistical Inference for Birth and Death Queuing Models," *Operations Research*, vol. 13, no. 3, pp. 343–357, 1965.

[3] H. Kobayashi, "Stochastic Modeling of an Infectious Disease: Part I: Understand the Negative Binomial Distribution and Predict an Epidemic More Reliably." `https://arxiv.org/pdf/2006.01586.pdf`, June 2 2020.

[4] H. Kobayashi, "Stochastic Modeling of an Infectious Disease: Part IV: Further Analysis of the Time-Nonhomogenous BDI Process." `http://hp.hisashikobayashi.com`, August 27 2021.

[5] H. Kobayashi, "Stochastic Modeling of an Infectious Disease: Part II: Simulation Experiments and Verification of the Analysis." `https://arxiv.org/abs/2101.11394` and `http://hp.hisashikobayashi.com`, January 22 2021.

[6] P. Billingsley, *Statistical Inference for Markov Processes*. University of Chicago Press, 1961.

[7] N. Keiding, "Maximum Likelihood Estimation in the Birth-and-Death Process," *The Annals of Statistics*, vol. 3, no. 2, pp. 363–372, 1975.

[8] H. Kobayashi, B. L. Mark and W. Turin, *Probability, Random Processes, and Statistical Analysis*. Cambridge University Press, 2012.

[9] H. Kobayashi, "Stochastic Modeling of an Infectious Disease: Part VI: Analysis of $B(t), R(t)$ and $J(t)$ used in Maximum Likelihood Estimation of the Birth-Death Process." `http://hp.hisashikobayashi.com`, (in preparation) 2021.

[10] F. R. Sharpe and A. J. Lotka, "A Problem in Age-Distribution," *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, vol. 21, no. 124, pp. 435–438.

[11] H. Kobayashi, "Stochastic Modeling of an Infectious Disease: Part VII: Approximate Analysis based on Saddle-Point Integration." `http://hp.hisashikobayashi.com`, (in preparation) 2021.

Copyright ©Hisashi Kobayashi 2021