



A Stochastic Model of an Infectious Disease, based on the Birth-and-Death-with-Immigration Process

Hisashi Kobayashi

The Sherman Fairchild University Professor, Emeritus
Princeton University

March 4, 2022
Presented at George Mason University
Spring 2022 ECE Distinguished Seminar Series



Acknowledgment and Outline of this Presentation

A Stochastic
Model of
an Infectious
Disease,
based on the
Birth-and-
Death-with-
Immigration
Process

Hisashi
Kobayashi

- Acknowledgment: I thank Prof. Brian Mark of GMU not only for this opportunity to talk, but for his helping me learn Matlab programming, and use of Latex to prepare the slides.
- Outline of this Presentation
 - 1 COVID-19 Data: Observations
 - 2 Time-Homogeneous BDI Process Based Model
 - 3 Analysis of Time-Nonhomogeneous BDI Process
 - 4 Simulation of Time-Nonhomogeneous BDI Process
 - 5 Maximum Likelihood Estimation of Model Parameters
 - 6 Recapitulation and Future Study Plan



Part 1: COVID-19 Data: Observations

A Stochastic
Model of
an Infectious
Disease,
based on the
Birth-and-
Death-with-
Immigration
Process

Hisashi
Kobayashi

- In the World;
 - Population: 7.9 billions
 - Reported cases: 421 millions (5.3%)
 - Number of deaths: 5.87 (1.4%)
- In the United States;
 - Population: 332 millions
 - Reported cases: 78.4 millions (23.6%)
 - Number of deaths: 5.87 millions (1.2%)
- In Japan;
 - Population: 125 millions
 - Reported cases: 4.42 millions (3.54%)
 - Number of deaths: 21,700 (0.49%)



U.S. Data: Daily cases, Moving avg., and Cumulative cases

A Stochastic Model of an Infectious Disease, based on the Birth-and-Death-with-Immigration Process

Hisashi Kobayashi

- U.S. Daily cases, 7-day moving average and cumulative cases¹

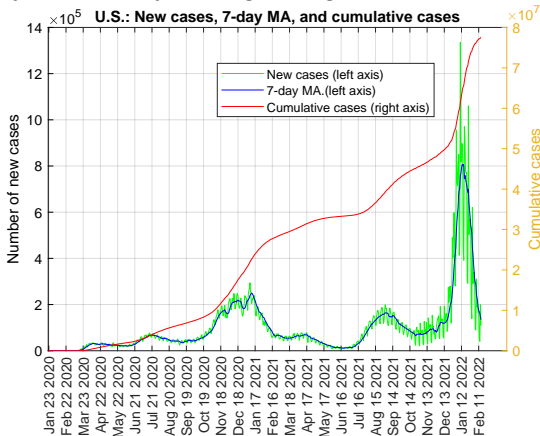


Figure 1: U.S. Data from Jan. 23, 2020 through Feb. 15, 2022

¹Source: CDC's COVID Data Tracker (csv file)



Moving Average Data of the U.S. and several states

- U.S., California, Florida, Hawaii, NYC, NYS (excluding NYC), and Virginia.

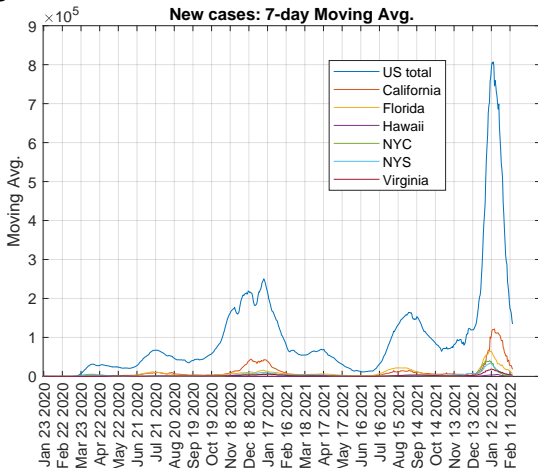


Figure 2: Moving Average of daily cases from Jan. 23, 2020 through 5/27



Semi-log plots of moving average in the U.S. and several states

- Semi-log scale plots often reveal much more information than plots in the linear scale.

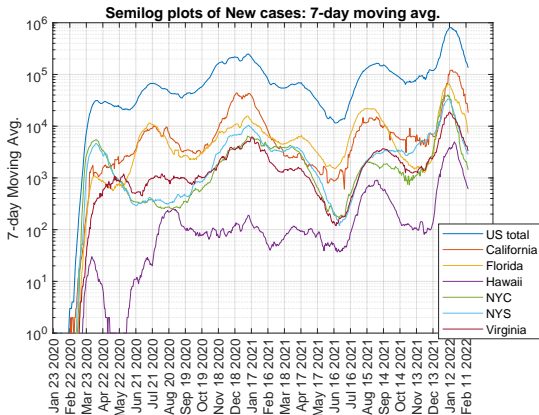


Figure 3: The vertical values are plotted in the logarithmic scale



Japan Data: Daily cases, Moving avg., and Cumulative cases

- Japan: Daily cases, 7-day moving average and cumulative cases²

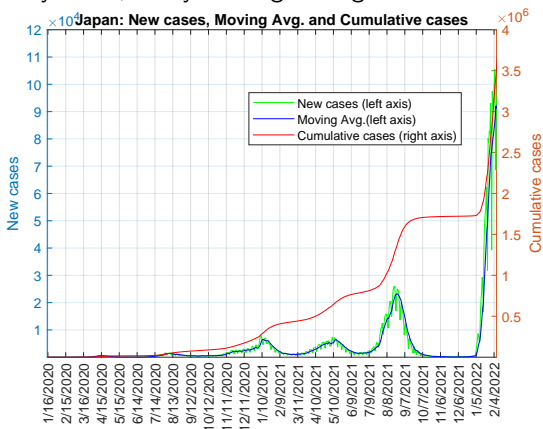


Figure 4: Japan Data from Jan. 23, 2020 through Feb. 15, 2022

²Source: Japan's Ministry of Health, Labor and Welfare (csv file)



Moving Average Data of Japan and several Prefectures

- Japan, Tokyo-to, Osaka-fu, Ibaraki-ken and Yamanashi-ken.

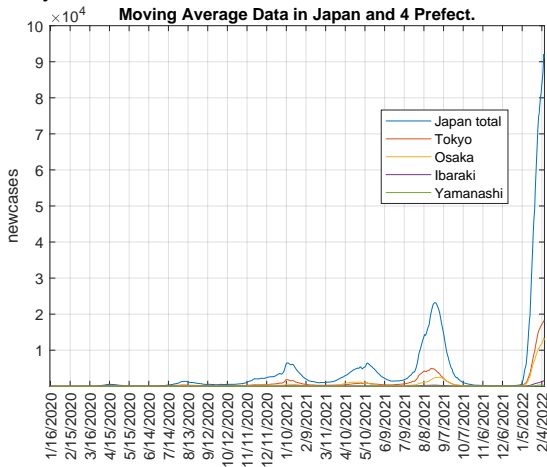


Figure 5: Moving Average of daily cases from Jan. 23, 2020 through Feb. 15, 2022



Semi-log Plots: Moving Average Data of Japan and several Prefectures

- Japan, Tokyo-to, Osaka-fu, Ibaraki-ken and Yamanashi-ken.

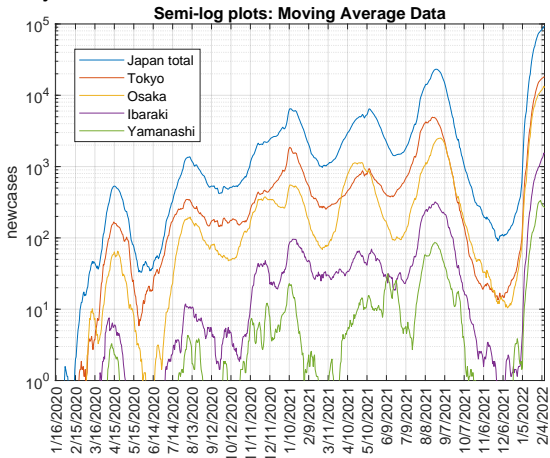


Figure 6: Semi-log plots: Moving Average of daily cases



Cumulative cases in Japan and several Prefectures

- Japan, Tokyo-to, Osaka-fu, Ibaraki-ken and Yamanashi-ken.

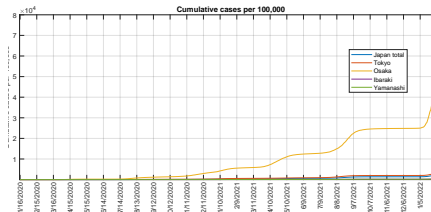


Figure 7: Cumulative cases plotted in a linear scale

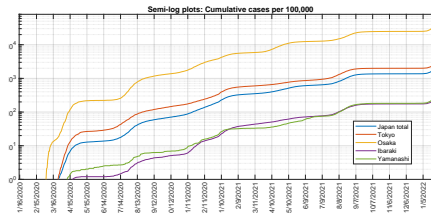


Figure 8: Cumulative cases plotted in a logarithmic scale



Summary of Our Observations and Questions

A Stochastic
Model of
an Infectious
Disease,
based on the
Birth-and-
Death-with-
Immigration
Process

Hisashi
Kobayashi

- 1 Large variations among different regions.
- 2 Usefulness of logarithmic presentations, especially in the initial period.
- 3 Should the “new and cumulative cases” be our primary concern?
- 4 How to develop a useful mathematical model to characterize and predict the behavior of an infectious disease?



Part 2: Time-Homogeneous BD Process

A Stochastic
Model of
an Infectious
Disease,
based on the
Birth-and-
Death-with-
Immigration
Process

Hisashi
Kobayashi

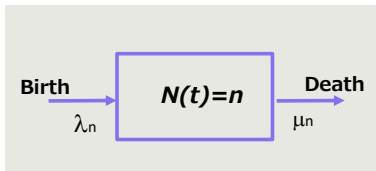


Figure 9: The time-homogeneous BD (birth-and-death) process with λ_n and μ_n , $n = 0, 1, 2, \dots$.

- The birth and death rates do not change in time, but depend on n .
- The **time-homogeneous BDI** (birth-and-death-with-immigration) process is a special case, defined by $\lambda_n = n\lambda + \nu$, and $\mu_n = n\mu$.
- The units of λ_n, μ_n and ν are [person/day], whereas λ and μ are [person/day/person]=[day⁻¹].
- Our model is an **"infinite population"** and **"linear" model**, whereas the SIR model³ assumes a finite population and is intrinsically nonlinear.

³The SIR model was proposed by Kirmack and McKendrick in 1927.



A BDI Process Model of an Infectious Disease

A Stochastic
Model of
an Infectious
Disease,
based on the
Birth-and-
Death-with-
Immigration
Process

Hisashi
Kobayashi

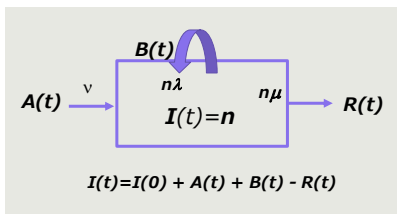


Figure 10: The time-homogeneous BDI model with BD rates (λ, μ) and the immigration rate ν . The blue rectangle box indicates an environment of infectious activities.

- $A(t)$: Cumulative arrivals of infected persons by t .
- $B(t)$: Cumulative count of internal infections by t .
- $R(t)$: Cumulative count of removed persons by t .
- $I(t)$: **Number of currently infected persons** at time t

A Question to you: What are "Daily new cases" in our model?



Time-homogeneous BDI Process, cont'd

- The **probability distribution** of $I(t)$:

$$P_n(t) = P[I(t) = n], \quad n = 0, 1, 2, \dots, \quad \text{and} \quad t \geq 0. \quad (1)$$

- Under a set of assumptions⁴, we can show that the following differential equations must hold

$$\frac{dP_0(t)}{dt} = -\nu P_0(t) + \mu P_1(t), \quad (2)$$

$$\begin{aligned} \frac{dP_n(t)}{dt} = & -[n(\lambda + \mu) + \nu]P_n(t) + [(n-1)\lambda + \nu]P_{n-1}(t) \\ & + (n+1)\mu P_{n+1}(t), \quad (3) \\ & n = 1, 2, \dots, \quad \text{and} \quad t \geq 0. \end{aligned}$$

⁴(1) The arrival process is a Poisson process; (2) The interval until each infectious person will infect another is exponentially distributed with mean λ^{-1} ; and (3) The period until an infectious person will be removed (or asymptomatic and recover) is exponentially distributed with mean μ^{-1} .



The Probability Generating Function (PGF)

- Define the probability generating function (PGF) of $P_n(t)$'s by

$$G(z, t) = \sum_{n=0}^{\infty} z^n P_n(t). \quad (4)$$

- Then, (2) and (3) will be transformed into the partial differential equation (PDE):

$$\frac{\partial G(z, t)}{\partial t} = (z - 1) \left[(\lambda z - \mu) \frac{\partial G(z, t)}{\partial z} + \nu G(z, t) \right], \quad (5)$$

with the boundary condition

$$G(z, 0) = z^{I_0}. \quad (6)$$



The Solution PGF that satisfies the PDE

- By solving the PDE, we obtain⁵

$$G(z, t) = \left(\frac{1 - \beta(t)}{1 - \beta(t)z} \right)^r \left(\frac{\lambda z - \mu - \mu(z - 1)e^{at}}{\lambda z - \mu - \lambda(z - 1)e^{at}} \right)^{l_0}, \quad (7)$$

where

$$a \triangleq \lambda - \mu, \quad r \triangleq \frac{\nu}{\lambda} \quad \text{and} \quad l_0 \triangleq I(0). \quad (8)$$

$$\beta(t) \triangleq \frac{\lambda(e^{at} - 1)}{\lambda e^{at} - \mu}. \quad (9)$$

where $e = \sum_{n=0}^{\infty} \frac{1}{n!} = 2.71828 \dots$ (Euler's number or Napier's constant)

$$(10)$$

⁵See N. T. J. Baily, *The Elements of Stochastic Processes with Application to Natural Sciences* (1964); H. Takagi, "Lecture Note: Birth and Death Processes and Its Applications (in Japanese)," University of Tsukuba, (2007)



When $I_0 = 0$, $P_n(t)$ is Negative Binomial Distributed

- If $I_0 = 0$, the second term in the product form (7) is unity, hence

$$G(z, t) = \left(\frac{1 - \beta(t)}{1 - \beta(t)z} \right)^r \quad (11)$$

- By inverting this PGF, we obtain

$$P_n(t) = \binom{n+r-1}{n} (1 - \beta(t))^r \beta(t)^n. \quad (12)$$

- Equation (12) is known as a **generalized Negative Binomial Distribution**. For a detailed discussion, see my article⁶

⁶H. Kobayashi, *Stochastic Modeling of an Infectious Disease, Part I: Understand the Negative Binomial Distribution and Predict and Epidemic More Reliably*, <https://arxiv.org/abs/2006.01586>, June 2, 2020.



Part 3: Analysis of Time-Nonhomogeneous Model

- The PDE (7) of Slide 15 is generalized to

$$\frac{\partial G(z, t)}{\partial t} = (z - 1) \left[(\lambda(t)z - \mu(t)) \frac{\partial G(z, t)}{\partial z} + \nu(t)G(z, t) \right], \quad (13)$$

with the boundary condition $G(z, 0) = z^{I_0}$.

- I could solve the above PDE, obtaining⁷

$$G_{BDI:I_0}(z, t) = G_{BDI:0}(z, t)G_{BD:I_0}(z, t), \quad (14)$$

where

$$G_{BDI:0}(z, t) = \text{The solution of (13), when } I_0 = 0. \quad (15)$$

$$G_{BD:I_0}(z, t) = \text{The solution of (13), when } \nu(t) = 0, \text{ and } I(0) = I_0. \quad (16)$$

⁷See H. Kobayashi, *Stochastic Modeling of an Infectious Disease, Part III-B. Analysis of the Time-Nonhomogeneous BDI Process and Simulation Experiments of both BD and BDI Processes*, <https://arxiv.org/abs/2104.00529>, March 21, 2021.



The solution of the PDE in the Time-Varying Case

- The product form (14) implies

$$I_{BDI:l_0}(t) = I_{BDI:0}(t) + I_{BD:l_0}(t). \quad (17)$$

- The expected values can be obtained as

$$\bar{I}_{BDI:0}(t) = N(t)e^{s(t)}, \quad (18)$$

$$\bar{I}_{BD:l_0}(t) = I_0 e^{s(t)}. \quad (19)$$

where

$$s(t) \triangleq \int_0^t a(u) du = \int_0^t (\lambda(u) - \mu(u)) du. \quad (20)$$

and

$$N(t) \triangleq \int_0^t \nu(\tau) e^{-s(\tau)} d\tau. \quad (21)$$



Analogy between the infection's growth and the asset growth in investments

- $\log(2) \approx 0.6931 \Rightarrow$
 With $a = 0.1$, $y_1(t) = e^{at} = 2$ at $t \approx 6.9$.
 With $a = 0.2$, $y_1(t) = e^{at} = 2$ at $t \approx 3.45$.
- We might call the above observation "Rule of 69," analogously to the well known "Rule of 72." ⁸

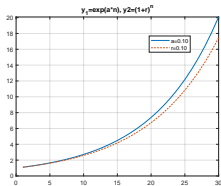


Figure 11: $y_1 = e^{at}$
 vs. $y_2 = (1+a)^t$, for
 $a = 0.1$.

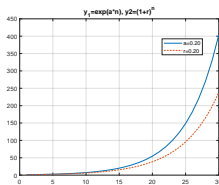


Figure 12: $y_1 = e^{at}$
 vs. $y_2 = (1+a)^t$ for
 $a = 0.2$.

⁸A formula that estimates the amount of time it takes for an investment to double in value, earning a fixed annual rate of return $a \times 100\%$. That is, $t \approx \frac{72}{100a}$.



Reproduction Numbers R_0 and R_t and Exponential Growth Parameter $a(t)$

A Stochastic
Model of
an Infectious
Disease,
based on the
Birth-and-
Death-with-
Immigration
Process

Hisashi
Kobayashi

- The ratio $R_0 \triangleq \frac{\lambda_0}{\mu_0}$ is the **basic reproduction number**.
- The ratio $R_t \triangleq \frac{\lambda(t)}{\mu(t)}$ is the **effective reproduction number**.⁹
- We find

$$a(t) = (R_t - 1)\mu(t). \quad (22)$$

- Because $\mu^{-1}(t)$ is the expectation of an infections period, R_t is the **expected number of susceptible persons infected by one infectious person**.
- The value of $\mu^{-1}(t)$ should be equal to the mean incubation period plus the number of days before a test is taken and its positive result is revealed.

⁹Sometimes called the *real-time reproduction number*.



Some reported R_0 's and incubation periods could be grossly wrong?

- The **incubation period** of COVID-19 (before Omicron variant) $\approx 2 \sim 14$ days, with mean 5.6 days. For **Omicron** the mean incubation is around 3 days¹⁰
- **Delta variant**: $R_0 \approx 3.2 \sim 8$, with the mean 5.08¹¹
- **Omicron variant**: Danish researcher estimate R_0 of Omicron is 3.19 times R_0 of Delta, whereas Japanese researchers; estimate R_0 of Omicron to be 4.2 times R_0 of Delta¹² $\Rightarrow R_0 \approx 16 \sim 20$?
- The above data imply the following parameter estimates for **Omicron**, which are **very questionable**.
 - If we assume it takes, say, 3 days for a person with a symptom takes a PCR test, resulting in a positive result, being removed to a hospital, then $\mu^{-1} = 3 + 3 = 6$ [days] $\Rightarrow \mu_0 = 0.167$ [day⁻¹]
 - Choosing $R_0 \approx 16 \Rightarrow \lambda_0 = R_0 \mu_0 \approx 16/6 = 2.667$, and $a_0 \approx 2.667 - 0.167 = 2.5$.
- **If you are aware of any reliable information, please let me know.**

¹⁰WebMD: "Corona-virus Incubation Period".

¹¹Y. Liu and J. Rocklöv. *Journal of Travel Medicine*, Oct. 2021

¹²S. Senanyake, *The conversation.com*, March 1, 2022



A Simple Example of Time-Varying Case

A Stochastic
Model of
an Infectious
Disease,
based on the
Birth-and-
Death-with-
Immigration
Process

Hisashi
Kobayashi

- Let $\lambda(t) = 0.3$, $0 \leq t \leq 30$, and $\lambda(t) = 0.06$, $t \geq 40$.
- Assume that $\mu(t) = 0.1$ for all t . Then $a(t) = \lambda(t) - \mu(t)$ is as shown below.

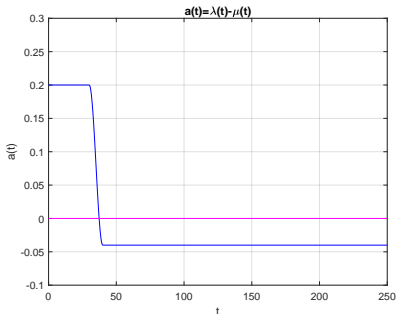


Figure 13: $a(t) = a_0 = 0.2$ ($0 \leq t \leq 30$), $a(t) = a_1 = -0.04$ ($t \geq 40$)

- As a smooth transition in the interval $[30, 40]$, we chose the “raised-cosine function” for analytic simplicity.



$$s(t) = \int_0^t a(\tau) d\tau$$

Recall the function $S(t)$ defined earlier:

$$s(t) = \int_0^t a(\tau) d\tau = \int_0^t (\lambda(\tau) - \mu(\tau)) d\tau \quad (23)$$

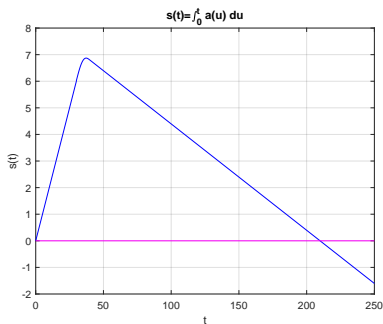


Figure 14: $s(t)$ that corresponds to the $a(t)$ in Figure 13



$$\nu(t) = 0, I_0 > 0 \implies$$

the BDI process reduces to a BD process with I_0 .

- When $\nu(t) = 0$, then the $I_{BDI:0}$ term of (17) disappears:

$$\bar{I}_{BDI:I_0}(t) = \bar{I}_{BD:I_0}(t) = I_0 e^{s(t)}. \quad (24)$$

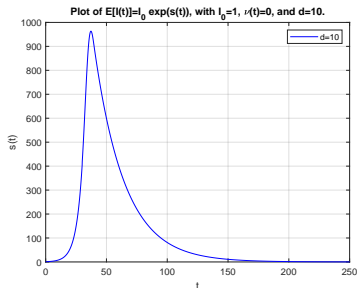


Figure 15: $\bar{I}_{BD:I_0}(t)$ with $I_0 = 1$ and $\nu(t) = 0$

When $I_0 = 0$, and $\nu(t) = \nu_0 = 0.2$ for all $t \geq 0$.

- When $I_0 = 0$, $\bar{I}_{BD:0}(t)$ of (17) disappears, and

$$\bar{I}_{BD:0}(t) = N(t)e^{s(t)}, \quad (25)$$

and

$$N(t) \triangleq \int_0^t \nu(\tau) e^{-s(\tau)} d\tau = \nu_0 J(t) \approx \frac{\nu_0}{a_0} = \frac{\nu_0}{\lambda_0 - \mu_0} [\text{persons}]. \quad (26)$$

Note that it is flat in $20 \leq t \leq$, beyond which it goes up.

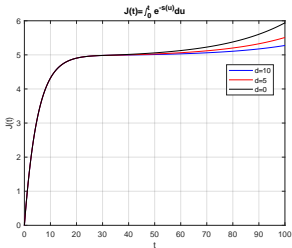


Figure 16: $J(t) = \int_0^t e^{-s(\tau)} d\tau$

$$\bar{T}_{BDI:0}(t) = \nu_0 e^{s(t)} J(t), \quad t \geq 0$$

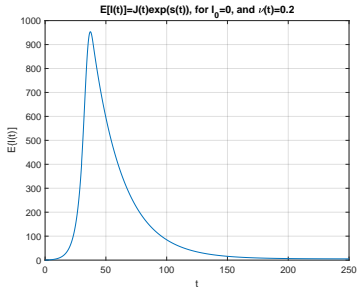


Figure 17: $\bar{T}_{BDI:0}(t) = \nu_0 e^{s(t)} J(t)$, with $\frac{\nu_0}{a_0} = 1$ [person]

- This $\bar{T}_{BDI:0}(t)$ is hardly distinguishable from the $\bar{T}_{BD:1}(t)$ of Figure 15.
- If both I_0 and $\nu(t)$ are nonzero, we have

$$\bar{T}_{BDI:I_0}(t) \approx \left(I_0 + \frac{\nu_0}{\lambda_0 - \mu_0} \right) e^{s(t)}, \quad t \geq 0. \quad (27)$$



Discussion and Remarks for Part 3

Important results:

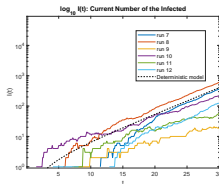
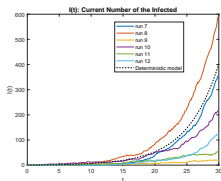
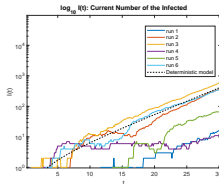
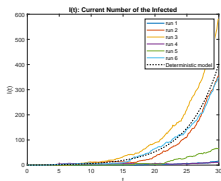
- 1 Check the consistency among $R(t)$, $a(t)$ and $\mu(t)$. Recall Eq. (22)
- 2 Inertia in $s(t)$, because of its integration form.
 - $s(t)$ (hence $\bar{I}(t)$ process as well) will begin to decline, only when $a(t)$ becomes < 0 .
 - Note the characteristics of the curves $N(t)$ and $J(t)$.
 $N(t) \approx \frac{\nu_0}{a_0}$ for the important range of t . $\implies N(t)$ is insensitive to detailed shapes of $a(t)$ and $\nu(t)$: only their values around $t = 0$ matter.
 - Thus, the shapes of $\bar{I}_{BD:l_0}(t)$ and $\bar{I}_{BDI:0}(t)$ are hardly distinguishable.
- 3 Now we conduct simulation experiments to see how the sample paths of independent runs behave.



Part 4: Simulation using $a(t)$ of Figure 13

- Independent 12 simulation runs¹⁴.

Note the enormous disparities among the different instances (sample paths) are simply due to probabilistic chances!!.



¹⁴ As for the curves beyond $t = 30$ see Part III-B <https://arxiv.org/abs/2104.00529> pp. 21-24, where $t_1 = 50$ instead of $t_1 = 30$ is chosen.

Plots of Runs 1-6: An Expanded View

A Stochastic
Model of
an Infectious
Disease,
based on the
Birth-and-
Death-with-
Immigration
Process

Hisashi
Kobayashi

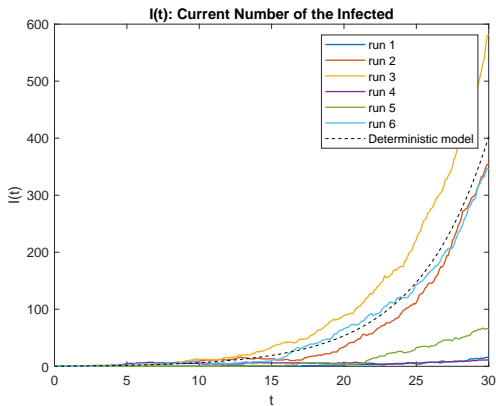


Figure 18: Runs 1-6: An expanded view of Figure 18



Semi-log Plots Runs 1-6

A Stochastic
Model of
an Infectious
Disease,
based on the
Birth-and-
Death-with-
Immigration
Process

Hisashi
Kobayashi

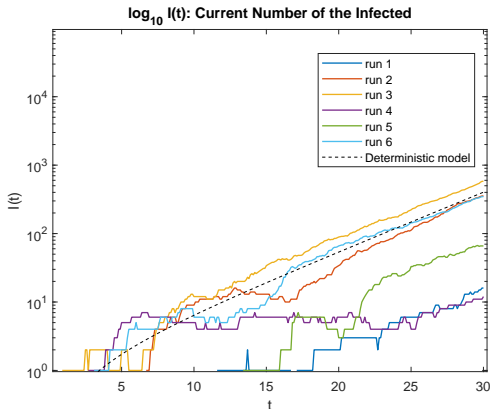


Figure 19: Semi-log plots: Runs 1-6: An expanded view of Figure 19

Note that the expected number of infections people is around 400 on day 30, whereas a worst case shows that it is 600, whereas in the luckiest case it is only 10 or so, **the difference being as large as a factor of 60!!**).



Discussion and Remarks for Part 4

- 1 An epidemic can be **initially very chaotic**.
 - The **negative binomial distribution with small r** can explain huge disparities.
 - When $I(t)$ becomes sufficiently large, the **law of large number** will set in.
- 2 The quantity $I(t)$ should provide **more critical information** than “Daily new cases.”
 - “**Cumulative cases**” given in CDC and other data bases is essentially equivalent to $R[n]$, which is a sampled value of $R(t)$ of our model at $t = n + \delta$ for some $\delta(0 \leq \delta < 1)$.
 - “**Daily new cases**” is equal to $dR[n] \triangleq R(n) - R(n-1)$.
 - **How can we estimate $I[n]$ from $R(n)$?** Note the identity

$$I[n] = I_0 + A[n] + B[n] - R[n]. \quad (28)$$

Discussion on Time-varying $\lambda(t)$

A Stochastic
Model of
an Infectious
Disease,
based on the
Birth-and-
Death-with-
Immigration
Process

Hisashi
Kobayashi

■ Let

$$p(t) \triangleq \text{Proportion of population vaccinated by time } t, \quad (29)$$

$$f_v \triangleq \text{P[Vaccinated person is protected]}, \quad (30)$$

$$b(t) \triangleq \text{Behaving factor of the public, } 0 \leq b(t) \leq 1.. \quad (31)$$

■ Then

$$\lambda(t) = \lambda_0 b(t)(1 - f_v p(t)), \quad (32)$$



Discussion on Time-varying $\nu(t)$

- Let

$$f_{BC}(t) \triangleq \text{Failing factor at the Border Control, } 0 \leq f_{BC}(t) < 1 \quad (33)$$

where $f_{BC}(t) \rightarrow 0$, if a tightest border control is enforced.

- Then, an infectious persons may come through an airport/seaport check point with the rate

$$\nu(t) = \nu_0 f_{BC}(t), \quad (34)$$



Discussion on Time-Varying $\mu(t)$

- We estimate the mean duration period, in which an infected individual remains infectious

$$\mu(t)^{-1} = T_{incub} + T_{test} + W_{remove} \quad (35)$$

where

$$T_{incub} = \text{Mean incubation period} \quad (36)$$

$$T_{test} = \text{Mean time to take PCR test} \quad (37)$$

$$W_{remove} = \text{Mean wait time until removed} \quad (38)$$

- W_{remove} is the queuing time due to insufficient beds at hospitals.

-

$$\mu_0^{-1} = T_{incub} + T_{test} \quad (39)$$

$$\mu^{-1}(t) = \mu_0^{-1} + W_{remove} \quad (40)$$



Part 5: Model Parameter Estimation

- Plausible estimation of the exponential parameter \mathbf{a} .
 - The semilog plots of $I(t)^{15}$ suggest possible use of linear regression analysis.
- How can we apply the **theory of maximum-likelihood estimation** to this problem?
 - **State-Dependent BD process.**

$$\lambda(t) = \lambda_{I(t)} \quad \text{and} \quad \mu(t) = \mu_{I(t)}, \text{ where } I(t) = n, n = 0, 1, 2, \dots \quad (41)$$

- Let $i(t)$ represent a **sample path (i.e., an instance)** of $I(t)$. Let t_n be the time of the n th transition; let $i_n = i(t_n)$, $n = 1, 2, 3, \dots$

¹⁵Similar observations are applicable to $B(t)$, $R(t)$ and other processes.



Likelihood Function

- Let τ_n be the duration that $i(t)$ stays in state i_n , i.e., $\tau_n = t_{n+1} - t_n$.
- The **likelihood function** for the parameters θ

$$\theta \triangleq (\lambda_0, \mu_1, \lambda_1, \mu_2, \dots, \mu_i, \lambda_i, \dots) \quad (42)$$

is defined by

$$L_i(\theta) = P\{I = i \mid \theta\} \quad (43)$$

where

$$i \triangleq \{i(t), 0 \leq t \leq T\}, \quad \text{and} \quad I \triangleq \{I(t), 0 \leq t \leq T\}. \quad (44)$$



Log-Likelihood Function

A Stochastic
Model of
an Infectious
Disease,
based on the
Birth-and-
Death-with-
Immigration
Process

Hisashi
Kobayashi

- Log-likelihood function of a state dependent BD process (R. W. Wolff¹⁶)

$$\begin{aligned} \log L_{\mathbf{i}}(\boldsymbol{\theta}) = & \log P_{I(0)}(i_0) + \sum_{i \geq 0} u_i \log \lambda_i + \sum_{i \geq 1} d_i \log \mu_i \\ & - \sum_{i \geq 1} \gamma_i (\lambda_i + \mu_i), \end{aligned} \quad (45)$$

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 $[0, T]$.

¹⁶R. W. Wolff, " Problems of Statistical Inference for Birth and Death Queuing Models," *Operations Research*, vol. 13, pp. 343-357, 1965.



Log-Likelihood Function for Time-Homogeneous BD process

- Let

$$\lambda_i = i\lambda, i \geq 0, \quad \text{and} \quad \mu_i = i\mu, i \geq 1. \quad (46)$$

The set of parameters (42) is reduced to $\theta = (\lambda, \mu)$.

- The log-likelihood function can be written as

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

(47)

where

$$C = \log P_{I(0)}(i_0) + \sum_{i \geq 0} u_i \log i + \sum_{i \geq 1} d_i \log i. \quad (48)$$

- Note

$$\sum_{i \geq 1} u_i = b(T), \quad \sum_{d_i \geq 1} d_i = r(T), \quad \text{and} \quad \sum_{i \geq 1} i\gamma_i = \int_0^T i(t) dt,$$

Maximum-Likelihood Estimation of $\theta = (\lambda, \mu)$

- By defining

$$J(t) \triangleq \int_0^t I(\tau) d\tau, \quad \text{and a sample path } j(t) \in J(t), \quad (50)$$

we find

$$\sum_{i \geq 1} i \gamma_i = j(T), \quad (51)$$

- **Maximum-likelihood estimate** of λ and μ :

$$\hat{\theta}_{MLE} = \theta^* = (\lambda^*, \mu^*) = \left(\frac{b(T)}{j(T)}, \frac{r(T)}{j(T)} \right). \quad (52)$$



Maximum-Likelihood Estimation of the exponential growth parameter a

- MLE of $a = \lambda - \mu$:

$$\hat{a}_{MLE} = a^* = \frac{i(T) - I_0}{j(T)}. \quad (53)$$

- Formula (53) would not have been conceived from the semi-log plots presented earlier.
- The formula, however, could have been conceived, had we considered $\bar{I}(t)$ and its integration $\bar{J}(t)$:

$$\bar{I}(t) - I_0 = I_0(e^{at} - 1), \quad (54)$$

$$\bar{J}(t) = \frac{I_0(e^{at} - 1)}{a}. \quad (55)$$



Discussions and Remarks for Part 5

A Stochastic
Model of
an Infectious
Disease,
based on the
Birth-and-
Death-with-
Immigration
Process

Hisashi
Kobayashi

- 1 Theory of maximum likelihood estimation has yield surprisingly simple and beautiful formulas, which could not have been conceived by intuition or ad-hoc approach.
- 2 An excellent example to demonstrate the power and importance of mathematical analysis.
- 3 Simulation study has confirmed that the above estimation algorithm converge to the true parameter values.
- 4 We should extend the theory and devise an adaptive algorithm for a time-varying environment.



What have we learned?

- 1 The **haphazard behavior** of an infectious disease, especially in its early phase, can be explained by the **negative binomial distribution** with small r . A deterministic model such as the SIR cannot explain the behavior
- 2 It may not be worthwhile to seek all possible reasons for differences among different communities. Enormous differences in the epidemic pattern may be **due to mere good luck or a lack thereof**.
- 3 The daily cases, $dR[n]$ in our notation, does not provide a real picture of the current threat. The medical community should come up with a **better indicator, e.g., an estimate of $I[n]$** .
- 4 Now that we know the **general solution**, in terms of PGF, for the **time-nonhomogeneous BDI process**, our stochastic model seems even more promising.
- 5 The **theory of maximum-likelihood estimation** has led to a simple estimation formula.
- 6 **Logarithmic plots** of data often reveals **valuable information** that might be completely hidden or buried otherwise.



Where should we go from here?

■ Future Plans

- 1 Need to find useful data of Covid-19, and **identify gaps between obtainable data and desirable data**. How can we deal with the limitations of data that are available in practice?
- 2 A **regression-analysis** approach is perhaps practiced to estimate the exponential growth/decay parameter “a.” What should be the best way to do it, in **a time varying environment**?
- 3 We should explore formulating the model parameter estimation problem in terms of a **hidden Markov model (HMM)**. The Expectation-Maximization (EM) Algorithm approach may help. Will a **neural network or machine learning** approach be applicable?
- 4 The processes $B(t), R(t), J(t)$ are **not Markov processes** and their comprehensive analysis has not been discussed in the literature even for a time-homogeneous BDI process (see the upcoming report Part VI).
- 5 Use of the **saddle-point integration** allows us to obtain an approximate probability density functions of $B(t), R(t), I(t), J(t)$ and other related processes (see the upcoming report Part VII).



Literature on the Birth-Death Process

A Stochastic
Model of
an Infectious
Disease,
based on the
Birth-and-
Death-with-
Immigration
Process

Hisashi
Kobayashi

■ Relevant literature

- 1 W. Feller, 'Foundations of the Volterra theory of struggle for existence in a probability theoretic treatment,' *Acta Biotheoretica*, Vol. 5, pp. 11-40, 1939.
- 2 D. G. Kendall, "The generalized 'birth-and-death' process," *Ann. Math Statist.*, vol. 19, pp. 1-15, 1948.
- 3 D.G. Kendall, "Stochastic processes and population growth," *J. R. Statist. Soc. B.* vol. 11, p. 230-, 1949.
- 4 M.S. Bartlett, *An Introduction to Stochastic Processes: with special reference to method and applications*, Cambridge University Press, 1st ed., 1955.
- 5 N. T. J. Baily, *The Elements of Stochastic Processes with Application to Natural Sciences*. Wiley & Sons, Inc., 1964.
- 6 H. Takagi, "Lecture Note: Birth and Death Processes and Its Applications (in Japanese)," University of Tsukuba, March 2007.
- 7 L. J. D. Allen, *An Introduction to Stochastic Processes with Applications to Biology*, Chapman & Hall/CRC, 2nd ed. 2011.



Related Publications

A Stochastic
Model of
an Infectious
Disease,
based on the
Birth-and-
Death-with-
Immigration
Process

Hisashi
Kobayashi

- 1 H. Kobayashi, *Stochastic Modeling of an Infectious Disease, Part I: Understand the Negative Binomial Distribution and Predict and Epidemic More Reliably*, <https://arxiv.org/pdf/2006.01586>, June 2, 2020.
- 2 H. Kobayashi, *Stochastic Modeling of an Infectious Disease, Part II: Simulation Experiments and Verification of the Analysis*, <https://arxiv.org/abs/2101.11394>, January 26, 2021.
- 3 H. Kobayashi, *Stochastic Modeling of an Infectious Disease, Part III-A: Analysis of Time-Nonhomogeneous Models*, <https://arxiv.org/abs/2101.09109com>, January 22, 2021
- 4 H. Kobayashi, *Stochastic Modeling of an Infectious Disease, Part III-B: Analysis of the Time-Nonhomogeneous BDI Process and Simulation Experiments of both BD and BDI Processes*, <https://arxiv.org/abs/2104.00529>, March 21, 2021.
- 5 H. Kobayashi, *Stochastic Modeling of an Infectious Disease, Part IV: Further Results on the BDI Process Model and Analysis of Multiple Types (variants) of Coronavirus*, <https://hp.hisashikobayashi.com/>, August 27, 2021.
- 6 H. Kobayashi, *Stochastic Modeling of an Infectious Disease, Part V : Maximum-Likelihood Estimation of the BDI Process Parameters*, <http://hp.hisashikobayashi.com>, October, 2021.
- 7 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*, (Under preparation) To appear in <http://hp.hisashikobayashi.com>.
- 8 H. Kobayashi, *Stochastic Modeling of an Infectious Disease, Part VII: Approximate Analysis based on Saddle-Point Integration* (under preparation).



Related Presentations

A Stochastic
Model of
an Infectious
Disease,
based on the
Birth-and-
Death-with-
Immigration
Process

Hisashi
Kobayashi

- 1 H. Kobayashi, *A Stochastic Model of an Infectious Disease*, Keynote presentation (over Zoom) at ITC-32, September 22-24, Osaka Japan. For the slides and video, see <https://hp.hisashikobayashi.com/a-stochastic-model-of-an-infectious-disease/>:
- 2 H. Kobayashi, *A Probabilistic Model of an Infectious Disease, bases on the BDI Process, and Maximum-Likelihood Estimation of the Model Parameters* Presented (over Zoom) at a Joint Study Group Conference "OR in Health Care" and "Dynamic Decision Models and Applications," The Operations Research Society of Japan, September 29, 2021. For the slides, see <https://hp.hisashikobayashi.com/a-stochastic-model-of-an-infectious-disease/>
- 3 H. Kobayashi, *A Stochastic Model of an Infectious Disease based on the Birth-and-Death-with-Immigration Process*, Presented (over Zoom) at George Mason University, Department of Electrical and Computer Engineering, Distinguished Seminar Series. March 4, 2022. <https://hp.hisashikobayashi.com/>