

# 多変量感染症時系列モデル

小山 慎介 モデリング研究系 准教授

## 1 Abstract

An infectious disease spreads over a single population or community and also across multiple and heterogeneous communities. Moreover, its transmissibility varies over time because of several factors, such as seasonality and epidemic control, resulting in strongly nonstationary behavior. In conventional methods for assessing transmissibility trends or changes, univariate time-varying reproduction numbers are calculated without considering the transmission across multiple communities. In this study, we propose a multivariate count time-series model for epidemics. We propose a statistical method for estimating the transmission of infections and time-varying reproductive numbers across multiple communities simultaneously from multivariate time series of case counts. We use our method on COVID-19 incidence data to reveal spatiotemporal differences in the epidemic process.

## 2 Methods

### Multivariate-count time series model

We consider an infection process that spreads over  $D$  nodes, where each node represents a group of individuals. Let  $n_{it}$  be the number of newly reported cases in  $(i, t)$ , where  $(i, t)$  denotes node  $i$  at time  $t$ . The rate of new cases at  $(i, t)$  is given by

$$\lambda_{it} = \sum_{j=1}^D a_{ij} \sum_{\tau=1}^{t-1} R_{j,t-\tau} \phi_{\tau} n_{j,t-\tau}, \quad (1)$$

where  $R_{j,t-\tau}$  denotes the effective reproduction number at  $(j, t-\tau)$  and  $a_{ij} (\geq 0)$  represents the transmission ratio from node  $j$  to node  $i$  satisfying  $\sum_{i=1}^D a_{ij} = 1$ . Based on the history of previously reported cases,

$$N_{1:t-1} \equiv \{n_{js} \mid j = 1, \dots, D; s = 1, \dots, t-1\}, \quad (2)$$

the count  $n_{it}$  is assumed to follow a Poisson distribution with the rate Eq (1):

$$P(n_{it} \mid N_{1:t-1}) = \text{Poisson}(\lambda_{it}) \equiv \frac{\lambda_{it}^{n_{it}}}{n_{it}!} e^{-\lambda_{it}}. \quad (3)$$

The multivariate count time series model comprises the two components in Eqs (1) and (3).

### Latent variable representing secondary infection in transmission chain

Let us introduce the latent variable  $y_{it}^{js}$ , which represents the number of secondary cases at  $(i, t)$  infected by the primary cases at  $(j, s)$  ( $s < t$ ). As the total number of new cases at  $(i, t)$  is given by  $n_{it}$ , the following equality holds:

$$\sum_{j=1}^D \sum_{s=1}^{t-1} y_{it}^{js} = n_{it}. \quad (4)$$

Assuming conditional independence between the transmission events  $y_{it}^{js}$  and  $y_{it}^{j's'}$  for  $(j, s) \neq (j', s')$  and given  $N_{1:t-1}$ . The superposition principle can be applied to the Poisson distribution (3), leading to a Poisson distribution for counts  $y_{it}^{js}$  with the rate given by  $\psi_{it}^{js} = a_{ij} R_{js} \phi_{t-s} n_{js}$ . Thus, given the sum of independent Poisson random variables  $n_{it}$ , the conditional distribution of each element of the Poisson vector  $Y_{it} \equiv \{y_{it}^{js} \mid j = 1, \dots, D; s = 1, \dots, t-1\}$  is multinomially distributed with count probabilities scaled by the sum of the individual rates:

$$P(Y_{it} \mid n_{it}, N_{1:t-1}) = \frac{n_{it}!}{\prod_{j=1}^D \prod_{s=1}^{t-1} y_{it}^{js}!} \prod_{j=1}^D \prod_{s=1}^{t-1} \left( \frac{\psi_{it}^{js}}{\lambda_{it}} \right)^{y_{it}^{js}}. \quad (5)$$

In particular, the conditional expectation of  $y_{it}^{js}$ , given  $n_{it}$  and  $N_{1:t-1}$ , is

$$\langle y_{it}^{js} \rangle \equiv E(y_{it}^{js} \mid n_{it}, N_{1:t-1}) = \frac{n_{it} \psi_{it}^{js}}{\lambda_{it}} = \frac{n_{it} a_{ij} R_{js} \phi_{t-s} n_{js}}{\sum_{k=1}^D a_{ik} \sum_{s=1}^{t-1} R_{ks} \phi_{t-s} n_{ks}}, \quad (6)$$

from which posterior inferences can be made regarding secondary infections from the reported incidences.

### Parameter estimation

In this study, we focus on estimating the weighted adjacency matrix  $A = (a_{ij})$  and time-varying reproduction number  $R = \{R_{js}\}$  of each node from an observed multivariate time series of incidence  $N_{1:T}$ . To estimate these parameters, we consider the following penalized log-likelihood function:

$$\mathcal{L}(A, R) = \sum_{i=1}^D \sum_{t=1}^T \log P(n_{it} \mid N_{1:t-1}) - \gamma \sum_{j=1}^D \sum_{s=2}^T |R_{js} - R_{j,s-1}|^p, \quad (7)$$

where the choice of exponent  $p \in \{1, 2\}$  and hyperparameter  $\gamma \geq 0$  depend on the sparsity or smoothness of the variation in the time-varying reproduction numbers. Using the latent variables representing secondary infection in transmission chain, we develop an expectation maximization (EM) algorithm to estimate  $A$  and  $R$ .

## 3 Results

We applied our method to actual data from the COVID-19 pandemic in Japan (<https://www.mhlw.go.jp/stf/covid-19/open-data.html>). The data consist of newly confirmed cases in  $D = 47$  prefectures between January 16, 2020, and November 24, 2021, in which 1,720,441 cases were reported. We applied our estimation method to new weekly cases to estimate the parameters  $\{\hat{A}, \hat{R}\}$ . Using the estimated parameters and Eq (6), the conditional expectation of secondary infections,  $\{\langle y_{it}^{js} \rangle\}$ , was computed, from which we made posterior inferences regarding the infections transmitted across the prefectures. In particular, we computed the expected total number of secondary infections in prefecture  $i$  that were transmitted from prefecture  $j$ :  $\langle y_i^j \rangle \equiv \sum_{t=1}^T \sum_{s=1}^{t-1} \langle y_{it}^{js} \rangle$ .

Overall, it was estimated that 82% of the infected cases ( $\sum_i \langle y_i^i \rangle = 1,414,104$  cases) were infected within each prefecture ("intra-prefectural infections") and that 18% ( $\sum_i \sum_{j \neq i} \langle y_i^j \rangle = 306,337$  cases) of the infections were transmitted across prefectures ("inter-prefectural infections"). Fig 1 shows a matrix visualization of  $(\langle y_i^j \rangle)$  (Fig 1a, heat map) and bar graphs of the intra-prefectural infections  $\langle y_i^i \rangle$ , in-degree  $\langle y_i \rangle \equiv \sum_{j \neq i} \langle y_i^j \rangle$ , and out-degree  $\langle y^i \rangle \equiv \sum_{i \neq j} \langle y_i^j \rangle$  in each prefecture (Fig 1b).

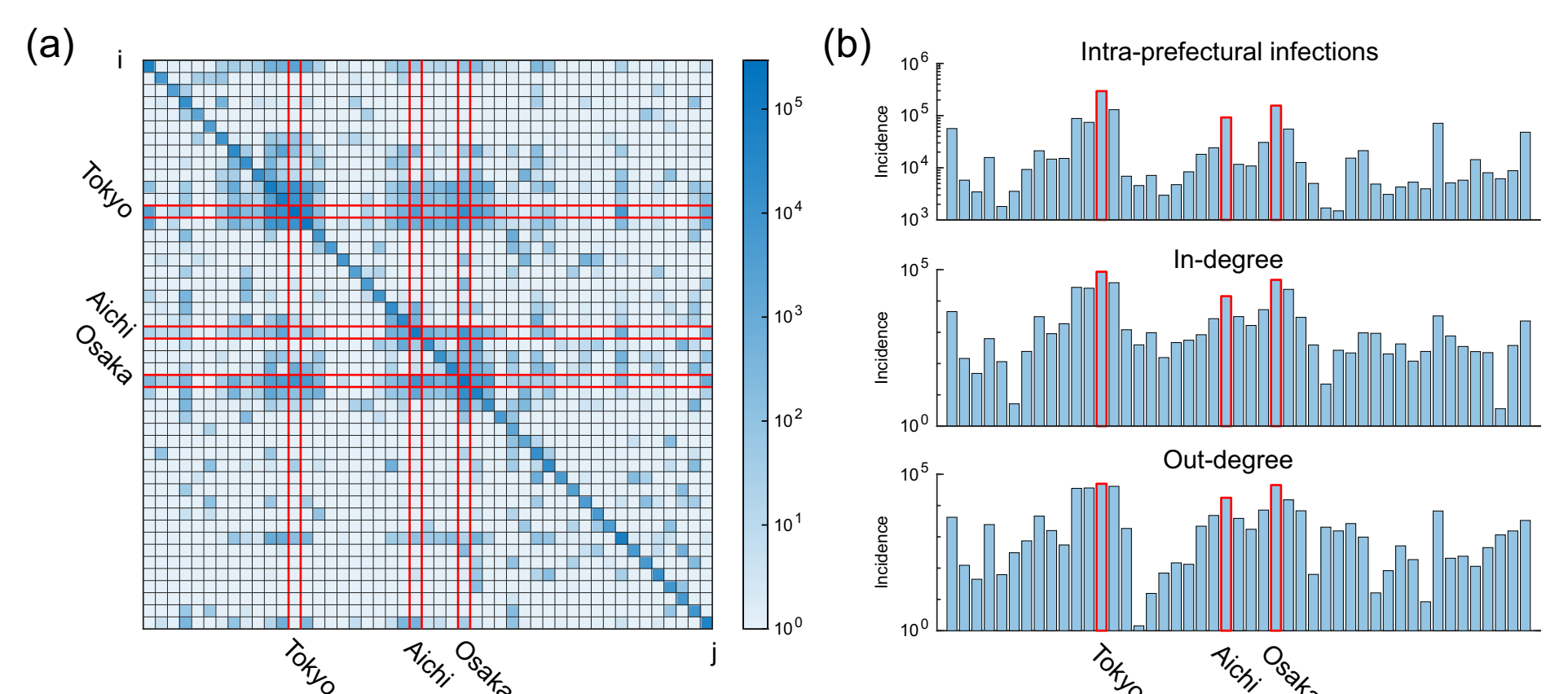


Fig 1: Result of the actual data analysis.