Branching model in spatial spread of COVID-19

2023-06-28

1 Framework

1.1 Superspreading phenomenon/Motivation

Superspreading is a phenomenon in which a small number of individuals are responsible for a large number of secondary infections. The distribution of secondary infections is highly skewed. The mean number of secondary infections is much larger than the median number of secondary infections. The variance of the distribution of secondary infections is much larger than the mean number of secondary infections. The distribution of secondary infections is overdispersed.

Empirical studies have shown the individual variation on disease emergence and superspreading is a normal feature of disease spread [**lloyd2005superspreading**]. This heterogeneity of the systems can be modeled by the branching process.

Recent collecting human mobility data and the associated studies have shown that human mobility is a key factor in the spatial spread of disease. The introduction of human mobility can facilitate the prediction of the spatial spread of disease.

The simulations of ODE and branching process with superspreading on the mobility matrix when compared with the real data has demonstrated that spatial pattern of spreading is more accurate mimiced by the superspreading model, especially in the early stages of the pandemic times.

- We demonstrated the differences by the nunber of counties which is infected against time.
- We define when the number of people infected per population is larger than a threshold as the infected county.
- The threshold is 0.01%

Therefore, we propose a branching process model with superspreading on the mobility matrix to understand the temporal and spatial pattern.

1.2 Inference of the dispersion rate

The dispersion parameter is a key parameter in the branching process. The dispersion parameter is the variance of the offspring distribution divided by the mean of the offspring distribution. The dispersion parameter is a measure of the heterogeneity of the offspring distribution. The dispersion parameter is a measure of the variability of the offspring distribution. The dispersion parameter is a measure of the skewness of the offspring distribution.

1.2.1 The performance of MCMC

We propose a baysian based algorithm to infer the dispersion rate from MCMC simulations.

- We fixed all the other parameters and only infer the dispersion rate, including R_0, mobility matrix, and the intial number of seed and positions of the branching.
- The temporal and spatial pattern of the spreading depends on the dispersion rate. In general, the patterns are sparsed and only limited number of locations are infected in the early stage of the spreading.
- As they are stochastic processes, the uncertainty of the ensemble of the simulations is high as the time goes by.
- We can give identify the lower bound of the dispersion rate by our method.

1.2.2 Quitify the uncertainty

- The uncertainty of the inferring results of the can be quantified by the profile likelihood.

- The larger the dispersion rate, the less uncertainty of the ensemble of the simulations and it is closer to the no-superspreading model.

1.3 Application to the Covid-19 data

We apply our model to the Covid-19 data in the US when the first wave of the pandemic.

The real world is much complicated than our model. We make the following adjustment to the model to make it can reflect the real world.

- The number positions where the spreading started is more than one places, i.e. multiple entries. We follow the work in to infer the number of entries, including the five counties reported the cases in the USA when the first wave of the pandemic.

- What we observed are the reported cases rather than the actual infected cases. The ratio between these two number is refered as reporting rate which is vary from place to place and from time to time. We use the results from xx paper as our reporting rate.

- R_0 is time-varrying as well. We use the results from xx paper as $R_{0,t}$.

- In reality, the mobility matrix is time-varying. We use the mobility matrix from safty graph as our time varrying mobility matrix.

2 Model description

2.1 Branching process

Galton-Watson branching process

Galton-Watson branching process (GWB) is a stochastic process that describes the growth of a population in which offspring are generated at random. The number of offspring of each individual is a random variable with a fixed distribution. The offspring of each individual are independent of each other. The population size at time t is denoted by $N(t)$. The population size at time $t+1$ is given by

$$N(t+1) = \sum_{i=1}^{N(t)} X_i, \tag{1}$$

where X_i is the number of offspring of individual i at time t. The distribution of X_i is assumed to be fixed and independent of i. The population size at time t is a random variable with a distribution $P(N(t))$. The distribution of $N(t)$ is given by.

Age-dependent branching process

Age-dependent branching process (ADBP) is a generalization of the GWB. The number of offspring of each individual is a random variable with a distribution that depends on the age of the individual. The distribution of X_i is given by

$$P(X_i = x) = \sum_{j=1}^{x} \frac{1}{j} \pi_j, \tag{2}$$

where π_j is the probability that an individual of age j has x offspring. The distribution of $N(t)$ is given by

$$P(N(t)) = \prod_{i=1}^{N(t)} \sum_{x=1}^{\infty} \frac{1}{x} \pi_x P(X_i = x). \tag{3}$$

Bellman-Harris branching process

Bellman-Harris branching process (BHBP) is a special case of the ADBP. The number of offspring of each individual is a random variable with a negative binomial distribution. The distribution of X_i is given by

$$P(X_i = x) = \frac{1}{x} \frac{\Gamma(x+r)}{\Gamma(x+1)\Gamma(r)} (1-p)^r p^x, \tag{4}$$

where p is the probability that an individual has x offspring and r is the dispersion parameter. The distribution of $N(t)$ is given by

$$P(N(t)) = \prod_{i=1}^{N(t)} \sum_{x=1}^{\infty} \frac{1}{x} \frac{\Gamma(x+r)}{\Gamma(x+1)\Gamma(r)} (1-p)^r p^x. \tag{5}$$

Negative binomial distribution

In the branching process model, the number of infections caused by each infected person (i.e., secondary infections) is represented by a negative binomial distribution $\mathcal{NB}(R_0, r)$ with a mean reproductive number R_0 and a dispersion parameter $r \in (0, +\infty)$. For a fixed , a smaller means a larger variation in secondary infections, that is, rarer but more explosive superspreading events. By varying the dispersion parameter , we can control individual transmission heterogeneity in the model.

A sequence of independent random variables $X_1, X_2, \dots$ is said to have a negative binomial distribution with parameters r and p. In each trial, the probability of success is p and the probability of failure is $1 - p$. The number of failures before the rth success is X. The probability mass function of x is given by

$$f(x) = \binom{x+r-1}{r-1} p^r (1-p)^x, \quad x = 0, 1, 2, \dots \tag{6}$$

where $\binom{x+r-1}{r-1} = \binom{x+r-1}{x} = \frac{(x+r-1)!}{x!(r-1)!}$ is the binomial coefficient. The mean and variance of the distribution are given by

$$\mathbb{E}[X] = \frac{rp}{1-p}, \quad \text{Var}[X] = \frac{rp}{(1-p)^2}. \tag{7}$$

When the definition of r extended to real number, not just the interger, the distribution is called the generalized negative binomial distribution. The probability mass function of x is given by

$$P(X_i = x) = \frac{1}{x} \frac{\Gamma(x+r)}{\Gamma(x+1)\Gamma(r)} (1-p)^r p^x, \tag{8}$$

where $p = R_0/(R_0 + 1)$ and r is the dispersion parameter. The distribution of $N(t)$ is given by

$$P(N(t)) = \prod_{i=1}^{N(t)} \sum_{x=1}^{\infty} \frac{1}{x} \frac{\Gamma(x+r)}{\Gamma(x+1)\Gamma(r)} (1-p)^r p^x. \tag{9}$$

2.2 Spatial spread

Qing's comments

The spatial spread of the disease is modeled as a branching process on a network. The network is a graph with nodes representing individuals and edges representing contacts between different locations. The number of secondary infections caused by each infected person is a random variable with a distribution that depends on the age of the individual.