**Estimation of transmission kernel parameters with Bayesian**

**Transmission kernel**

Transmission kernels is the distance-dependent probability of transmission from an infected to a susceptible farm. The transmission kernel subsume all route of transmissions1. Therefore, it is beneficial for quantifying the transmission rate when the transmission pathways are poorly known. The transmission kernel have been used to model animal disease, construct risk maps as well as study of the effectiveness of control measures 2–4.

The transmission kernel is typically explained by monotonically decreasing function dependent with between-farm distance. There are several transmission kernel functions 5,6, but in this study, we focus on transmission kernel function in eq.1, which was widely used in the livestock disease studies such as FMDV, avian influenza virus, and bluetongue virus 3,5,7

Eq.1

Where *k(rij)* is the transmission rate from infectious farm *j* to susceptible farm *i* which related to the distance between these two farms *(rij)*.The shape of the kernel is expressed by three parameters: the *k0* parameter represents the transmission rate per day at distance zero; *r0* represents the distance for which the transmission rate is half *k0*; represents the slope at which the transmission rate decreases as a function of distance.

**Maximum likelihood estimation and Bayesian estimation**

We can quantify the transmission kernel parameters, which expressed the shape of transmission kernel, from historical outbreak data. The most common method is using maximum likelihood estimation 8. This is achieved by finding the parameters that maximize a likelihood function, so that the observed data is most probable 9. However, maximum likelihood estimation also has a limitation, this method depends solely on the outcomes of observed data, it is easily biased and failed to converge when the outbreak data is small. Another alternate approach that can overcome this limitation is Bayesian estimation. Using Bayesian estimation, we can express priori beliefs about the parameter values, as a prior distribution P(θ), and combine the prior on parameter values, P(θ), with the likelihood of the data given certain parameter values, P(D | θ), resulting in the posterior distribution of the parameters given the data, P(θ | D). This process can be conceptually explained by Bayes’s rule: P(θ | D) = P(D | θ) P(θ) / P(D) 10.

**Example: Bayesian estimation of transmission kernel from FMD outbreaks in Thailand**

In this document, we use the example data of foot and mouth disease outbreak from Thailand. The FMD outbreak data was collected from two outbreak areas in Thailand in 2011. The full data can be downloaded from <https://zenodo.org/records/7708619>. The transmission kernel estimation using maximum likelihood on this data has already published in Thanicha et al. (2021)8. In this document, we will focus mainly on using Bayesian approach for transmission kernel estimation.

We use Stan to implement Bayesian models. Stan is a C++ library for Bayesian modeling and inference that uses the No-U-Turn sampler (NUTS) to obtain posterior simulations given a user-specified model and data 11. STAN has interfaces with both R and Python language, but in this document, we provide an example for an RStan interface.

1. **Define the model**

From the transmission kernel in Eq.1, the force of infection on the susceptible farm *i* on day *t* can be calculated by the sum of the kernel from all infectious farms on day *t*:

|  |  |
| --- | --- |
|  | Eq.2 |

The probability of escaping the infection and the probability of infection follows from the force of infection assuming a Poisson process:

|  |  |
| --- | --- |
|  | Eq.3 |
|  | Eq.4 |

The infection events () is assumed to be a realization of Bernoulli distributed probability of infection.

|  |  |
| --- | --- |
|  | Eq.5 |
|  |  |

1. **Data preparation**

The outbreak data consists of the status of farms in each day. We provide the data example below:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Farm | Status\_day1 | Status\_day2 | Status\_day3 | Status\_day4 | Status\_day… |
| 1 | S | S | S | S | S |
| 2 | S | E | E | I | I |
| 3  4  … | I  I  … | I  I  … | I  I  … | I  I  … | I  I  … |

S is susceptible status. I is infectious status. E is latent status. The infection event happens on the day that the farm changes status from S to E. With this data, we can calculate the force of infection of susceptible farm on each day (see Eq.2 for detail), and link it with the infection event. For example, on day 2:

Farm 1: ; event = non-infected  
Farm 2: ; event = infected

Farm 3 and 4 are already infected, so there is no force of infection acting upon them. To reduce computational time, we arrange the force of infection into a vector. This vector will be sliced into the fragment of force of infection on each farm on each day using start and stop indexes as an example below:

|  |  |  |  |
| --- | --- | --- | --- |
| Day1 | | Day2 | |
|  |  |  |  |
| Non-infected | Non-infected | Non-infected | Infected |

The R code for data preparation can be found in <https://github.com/AnnThanicha/Bayesian-estimation-for-transmission-kernel>

1. **Stan code**

To use Rstan, we first need to create the model in stan file. The Stan code for this example is shown below:

data {   
 int<lower=0> ID\_day; // the length of vector of force of infection

int<lower=0> N; // the number of farms

int<lower=0> K; // the length of vector of transmission rate

vector[K] distance\_kernel; // vector of transmission rate

array[ID\_day] int start; // start index for slicing vector of force of infection

array[ID\_day] int stop; // stop index for slicing vector of force of infection

array[ID\_day] int<lower=0> event; // infection events

}

parameters { // declaring transmission kernel parameters

real<lower=0> k0;

real<lower=0> r0;

real<lower=0> alpha;

}

model {

k0 ~ normal(0.005, 0.004) T[0, ]; // normal prior distribution with trancate at 0

r0 ~ normal(0.19, 0.08) T[0, ];

alpha ~ normal(1.56, 0.14) T[0, ];

vector[K] kernel = ((1+((distance\_kernel/r0)^alpha))^-1)\*k0; // calculate transmisión kernel

vector[ID\_day] prob; // create blank vector for probability of infection

for (i in 1:ID\_day) {

if (start[i] == stop[i]){prob[i] = kernel[start[i]]; }

else{prob[i] = sum(kernel[start[i]:stop[i]]);}

}

event[1:ID\_day] ~ bernoulli( 1 - exp(-1\*(prob[1:ID\_day]))); // Bernoulli distribution for infection event

}

Stan model starts by declaring all input data in the data block. Next is the parameter block that stores all parameters that we want to estimate. The third is the model block which the Bayesian model and parameter prior are defined. Here, we set the prior distribution of parameters as zero-truncated normal distribution. The values of prior are taken from the previous study8, which used maximum likelihood estimation.

More tutorials on Stan can be studied from: <https://mc-stan.org/>.

1. **R code**

Stan can be interfaced from the command line in R. In this example, we use ‘rstan’ package. All input data should be combined into a list to import to Stan. The R code is as below:

data <- list (distance\_kernel = distance\_kernel,event = event, N=N, ID\_day =ID\_day, start= start, stop = stop, K=K) # combine input data into a list

fit <- stan("pilot\_transmission\_parameter.stan",data = data) # call Stan code

k0 <- extract(fit, 'k0') # extract parameters

k0<- unlist(k0, use.names=FALSE)

mean(k0); quantile(k0, 0.025); quantile(k0, 0.975)

r0 <- extract(fit, 'r0')

r0<- unlist(r0, use.names=FALSE)

mean(r0); quantile(r0, 0.025); quantile(r0, 0.975)

alpha <- extract(fit, 'alpha')

alpha<- unlist(alpha, use.names=FALSE)

mean(alpha); quantile(alpha, 0.025); quantile(alpha, 0.975)

More detail on setting up RStan environment and validate the model can be read from: <https://mc-stan.org/users/documentation/case-studies/rstan_workflow.html>.

1. **Results**

The transmission kernel parameters estimated from maximum likelihood and Bayesian approach are shown in Table 1. All Bayesian estimated models have successfully converged, and the model validation shows no problems. the Meanwhile, the model with maximum likelihood estimation failed to converge for the Boh Ploi outbreak. This issue might cause by a small outbreak data (infection farms of 15 out of 500 farm).

**Table 1.** The transmission kernel parameters estimated from maximum likelihood and Bayesian approach.

|  |  |  |
| --- | --- | --- |
| Outbreak data | Transmission kernel parameters | |
| Maximum likelihood | Bayesian |
| Lamphaya Klang outbreak | k0 = 0.0055  r0 = 0.185  α = 1.56 | k0 = 0.0048  r0 = 0.208  α = 1.52 |
| Boh Ploi outbreak | Failed to converge | k0 = 0.0067  r0 = 0.238  α = 1.44 |
| Lamphaya Klang +  Boh Ploi outbreaks | k0 = 0.0054  r0 = 0.172  α = 1.50 | k0 = 0.0047  r0 = 0.206  α = 1.48 |

The parameters are slightly different between maximum likelihood estimation and Bayesian estimation. The parameter k0 from maximum likelihood estimation is higher than Bayesian estimation, while r0 from Bayesian estimation is higher.

**Conclusion**

Bayesian estimation is one of the useful method for parameter estimation, especially when the prior knowledge about parameters is available. Another advantage of Bayesian estimation is it treats parameters as probability distributions, which encompass the estimate’s uncertainty, instead of a point estimates from transitional method 10. However, there are also few disadvantages. Bayesian estimation needs a high computational resources and reliable priors. Unreliable priors can lead to a of highly biased models. Therefore, the decision on method depends on available data and model complexity.

**Supplementary**

All code and data for Bayesian estimation can be downloaded from <https://github.com/AnnThanicha/Bayesian-estimation-for-transmission-kernel>

**Reference**

1. Keeling, M. J. *et al.* Dynamics of the 2001 UK foot and mouth epidemic: Stochastic dispersal in a heterogeneous landscape. *Science* **294**, 813–817 (2001).

2. Boender, G. J., van Roermund, H. J. W., de Jong, M. C. M. & Hagenaars, T. J. Transmission risks and control of foot-and-mouth disease in The Netherlands: Spatial patterns. *Epidemics* **2**, 36–47 (2010).

3. De Koeijer, A. A. *et al.* Quantitative analysis of transmission parameters for bluetongue virus serotype 8 in Western Europe in 2006. *Vet. Res.* **42**, 1–9 (2011).

4. Backer, J. *et al.* Vaccination against Foot-and-Mouth Disease Vaccination against Foot-and-Mouth Disease Differentiating strategies and their epidemiological. (2009).

5. Hayama, Y., Yamamoto, T., Kobayashi, S., Muroga, N. & Tsutsui, T. Mathematical model of the 2010 foot-and-mouth disease epidemic in Japan and evaluation of control measures. *Prev. Vet. Med.* **112**, 183–193 (2013).

6. Ster, I. C., Singh, B. K. & Ferguson, N. M. Epidemiological inference for partially observed epidemics : The example of the 2001 foot and mouth epidemic in Great Britain. *EPIDEM* **1**, 21–34 (2009).

7. Boender, G. J. *et al.* Risk maps for the spread of highly pathogenic avian influenza in poultry. *PLoS Comput. Biol.* **3**, 704–712 (2007).

8. Chanchaidechachai, T., de Jong, M. C. M. & Fischer, E. A. J. Spatial model of foot-and-mouth disease outbreak in an endemic area of Thailand. *Prev. Vet. Med.* **195**, 105468 (2021).

9. Rossi, R. J. *Mathematical Statistics: An Introduction to Likelihood Based Inference*. (John Wiley & Sons, Hoboken, NJ, 2018).

10. Annis, J., Miller, B. J. & Palmeri, T. J. Bayesian inference with Stan: A tutorial on adding custom distributions. *Behav. Res. Methods* **49**, 863–886 (2017).

11. RStan: the R interface to Stan. https://mc-stan.org/rstan/articles/rstan.html.