EMULATING A GRAVITY MODEL TO INFERENCE THE SPATIOTEMPORAL DYNAMICS OF AN INFECTIOUS DISEASE

Roman Jandarov*, Murali Haran, Ottar Bjørnstad and Bryan Grenfell

Department of Statistics, The Pennsylvania State University; Department of Entomology and Center for Infectious Disease Dynamics, The Pennsylvania State University; Ecology and Evolutionary Biology, Princeton University.

OVERVIEW

- Our goal is to fit a gravity model (the Gravity-T-SIR model) that describes the space-time dynamics of measles.
- Likelihood evaluations are expensive so that standard maximum likelihood or Bayesian inference is computationally challenging.
- Likelihood-based approaches result in poor inference.
- We develop an efficient Gaussian process (GP) based approach for inference. We obtain an approximate likelihood based on fitting the GP to summary statistics from forward simulations of the gravity model.
- Our approach allows us to carry out Bayesian inference.
- We demonstrate that our approach: (i) is computationally tractable, (ii) results in reliable inference for the parameters, (iii) answers biological questions of interest, and (iv) allows for a study of identifiability issues in the model.

GRAVITY TIME SERIES SIR MODEL

- $SIR = \text{Susceptible-Infectious-Recovered}.$
- Gravity Time Series SIR Model = SIR model for local dynamics + explicit formulation for the spatial transmission between different host communities.
- Number of incidents (I): $I_{t+1} = \text{Poisson}(\lambda_{t+1})$, where $\lambda_{t+1} = \beta S_t (I_t + L_t)$.
- Number of susceptibles (S): $S_{t+1} = S_t + B_t - I_{t+1}$.
- Unobserved number of infected immigrants (L): $L_t = \text{Gamma}(m_t, 1)$, where $m_t = \theta N_t \sum_{k=1}^{N} \frac{(j_1 j_2)}{\theta_1 \theta_2}$.

PROBLEM STATEMENT

Since reliable estimates of local transmission parameters $\alpha$ and $\beta$ are known, our goal is to infer the unknown gravity parameter $\theta$, $\tau_1$, $\tau_2$, and $\rho$ based on the data for 952 cities in England and Wales.

CHALLENGES

- Computational challenge: evaluation of likelihood. The dimension of the data = 546*952 = 519,792.
- Standard likelihood-based inference results in poor inference, e.g. biases in parameter estimates, model with poor fit.
- Simulation from model is not cheap: makes it difficult to use approximate Bayesian computation (ABC) methods.
- Parameter identifiability issues.

LIKELIHOOD - BASED APPROACH

- Based on discretization of a part of the parameter space.
- Instead of assuming that $L_{t+1} \sim \text{Gamma}(m_{t+1}, 1)$, we fix $L_{t+1} \equiv m_{t+1}$ to avoid integration.
- Select a grid on the range of possible values for $\tau_2$ and $\rho$. For each point of the grid, we calculate and save matrices $M_{t+k}$, where $M_{t+k} = \sum_{j=1, j \neq k}^{N} \frac{(j_1 j_2)}{\theta_1 \theta_2}$.
- When calculating the likelihood, $L(\theta, \tau_1, \tau_2, \rho)$, assume that $\theta$ and $\tau_1$ are real numbers, and $\tau_2$ and $\rho$ are from the selected discrete grid.
- Use pre-calculated matrices $M_{t+k}$.

GP-EMULATOR - BASED APPROACH

- Based on constructing a new likelihood function using summary statistics that capture important biological characteristics of the disease dynamics.
- First stage: We find an approximate model using simulations of summary statistics (proportions of zeros) on a pre-selected grid of parameters.
- Second stage: We make inference based on the posterior distribution of the parameters under the approximate model for which calculation of the likelihood is fast.
- Simultaneously infer the discrepancy term between the approximate and the true models.

RESULTS

We simulated data using $\theta = 0.71$, $\tau_1 = 0.3$, $\tau_2 = 0.7$ and $\rho = 1$. Figure 1 (a) is the unconditional marginal likelihood surface for $(\tau_2, \rho)$. Figure 1 (b) is the marginal likelihood surface for $(\tau_2, \rho)$ conditioned on $\theta = 0.71$ (true). Figure 1 (c) is the likelihood surface for $(\tau_2, \rho)$ conditioned on $\theta = 0.71$ (true) and $\tau_1 = 0.3$ (true) and Figure 1 (d) is the likelihood surface for $(\tau_2, \rho)$ conditioned on $\theta = 1$ (any value) and $\tau_1 = 1$ (any value).

BILOGICAL CONCLUSIONS

- (i) The data have some information about the gravity parameters, but it is only possible to learn about two parameters.
- (ii) There is no statistically significant change in these parameters during the periods of holidays vs non-holidays.
- (iii) The gravity parameters do not seem to change from years 1944-55 to 1956-1966.
- (iv) Amounts of movement for different cities at different times can be estimated.

REFERENCES


FUNDING

This work was supported in part by Bill and Melinda Gates Foundation.