Iterated filtering and its applications in modeling malaria transmission in Northwest India

Anindya Bhadra*, Ed Ionides*, Karina Laneri† and Mercedes Pascual†

University of Michigan, *Departments of Statistics and †Ecology & Evolutionary Biology

{tatar,ionides,kaneri,pascual}@umich.edu

Abstract

Maximum likelihood by Iterated Filtering (MIF) is a sequential Monte Carlo based technique to compute the Maximum Likelihood Estimate (MLE) in a nonlinear, non-Gaussian state space model. In the paper, we outline some important properties of MIF. We further develop nonlinear stochastic dynamical models of malaria transmission and use the models to answer long-standing questions on relative importance of climate effect vs. intrinsic dynamics in the context of malaria transmission. MIF is used to carry out parameter inference.

Background

For parameter inference in a state space model, the MIF algorithm developed by Ionides et al. (Ref. 2) does not require evaluation of the state transition density, just simulation from that density. This results in great flexibility in constructing mechanistic disease models. Malaria is a vector-borne disease. One needs to model both the human and mosquito population as both are important to the spread of the disease. To test their effects, climate covariates need to be incorporated in the model as well.

The diagram of the compartment model is shown in Fig 1. We only observe the no. of people undergoing clinical treatment (Y) and the total population. The other states are unobserved and so are the rates in the model. Thus, the problem at hand is to estimate the parameters in a partially observed dynamical system, in presence of noise. MIF can be employed in order to find out the MLE.

Maximization via Iterated Filtering

Key idea: Bayesian inference for time-varying parameters is a solvable filtering problem. So set \( \theta \to 0 \) to be a random walk with \( \mathbb{E}[\theta_t|\theta_{t-1}] = \theta_{t-1} \) and \( \operatorname{Var}(\theta_t|\theta_{t-1}) = \sigma^2 \). The iterated filtering procedure below, using the limit \( \sigma \to 0 \), has been shown to maximize the likelihood (Ref. 2).

The MIF algorithm

Select \( \theta^{(0)} \), \( \sigma_\theta, \mu \), \( \alpha \), and \( N \). For \( n \) in \( 1, \ldots, N \):
1. Set \( \sigma = \sigma^{(n-1)} \) and initialize \( \mathbb{E}[\theta^{(n)}] = \theta^{(n)} \), \( \operatorname{Var}(\theta^{(n)}|\theta^{(n-1)}) = \sigma^2 \).
2. Evaluate the filtering means \( \theta^{(n)} = \mathbb{E}[\theta^{(n)}|y_{1:t}] \) and the prediction variances \( \mathbb{V}^{(n)} = \operatorname{Var}(\theta^{(n)}|y_{1:t-1}) \) for \( t = 1, \ldots, T \).
3. Set \( \theta^{(n+1)} = \theta^{(n)} + \mathbb{V}^{(n)} \left( \sum_{i=1}^T (1/\mathbb{V}^{(n)}) - 1/\mathbb{V}^{(n)} \right) \). The MIF estimate of \( \theta \) is \( \theta^{(N+1)} \).

Implementation of MIF

The computationally tricky step is to find \( \mathbb{E}[\theta^{(n)}|y_{1:t}] \) and \( \operatorname{Var}(\theta^{(n)}|y_{1:t-1}) \). We do this using a Monte Carlo particle filter – each “particle” is a trajectory of the dynamical system.

Suppose the “particles” \( (X_j^{(t)}, j = 1, \ldots, J) \) solve the filtering problem at time \( t \), i.e., \( \mathbb{E}[\theta|y_{1:t}] \) is the sample mean of \( (X_j^{(t)}) \).

Move particles by simulating from the stochastic dynamical model:

Make \( X_{j+1}^{(t)} \), a solution to the unobserved stochastic dynamical system at \( t + 1 \) with initial value \( X_j^{(t)} \). Then \( (X_j^{(t+1)}) \) solves the prediction problem at \( t+1 \), i.e., \( \operatorname{Var}(x_{t+1}|y_{1:t}) \) is the sample variance of \( (X_j^{(t+1)}) \).

Prune particles according to likelihood given data:

Make \( X_{j+1}^{(t)} \), a draw from \( (X_j^{(t+1)}) \), with probability proportional to the likelihood of the the observations at time \( t + 1 \) given \( (X_j^{(t)}) \). Then \( (X_j^{(t+1)}) \) solves the filtering problem at \( t+1 \).

Plug and Play

• We call a method “plug and play” if the model can be specified by an algorithm to simulate sample paths (no information about transition densities is required).
• This makes the model is a “black box” which inputs parameters and outputs sample paths.
• MIF is plug-and-play.
• The plug-and-play property gives flexibility in choosing models, and makes it easy to modify an existing model.

Modeling Malaria Transmission

Malaria: 300 - 500 million cases each year, resulting in nearly 1 million deaths.

• Caused by the Plasmodium parasite, carried by female Anopheles mosquitoes.

Estimates of 1 billion/year.

Model Equations

• State Model

\[
dS/dt = \mu_E S - \mu_E S - \mu_E S - \mu_E E - \mu_E E
\]
\[
dE/dt = \mu_E E - \mu_I - \mu_I - \mu_I - \mu_I
\]
\[
dI/dt = \mu_I I + \mu_I R - \mu_I Q - \mu_I Q - \mu_I Q
\]
\[
dR/dt = \mu_I I - \mu_R R - \mu_R - \mu_R - \mu_R
\]
\[
dC/dt = dC/dt = (f(t) - \kappa) \tau^{-1}
\]
\[
\lambda(t)/\lambda(t) = (\lambda - \lambda) \tau^{-1}
\]

Observation Model

\( M_{\Sigma} = \int \mu_E E(s) ds \)

\( Y(t) \mid M_{\Sigma} \sim \text{Negbin} \left( \mu_E, \text{var} = M_{\Sigma} + \sigma^2 \right) \)

Debate over climate factors

• The interannual variability of infectious diseases depends on intrinsic disease dynamics, as proposed for epidemic malaria in E. African highlands by Hay et al. (2002, Nature).

• Others think external drivers such as temperature and rainfall play a role in interannual variability of epidemic malaria. (Pascual et al., Proc. R. Soc. London B, 2008).

• We formally test the effect of intrinsic vs. extrinsic factors on interannual variability.

• The data from India come from a desert region, where rainfall may be the driving factor of the interannual variability, just as temperature may be the driving factor in E. African highlands.

Modeling rainfall

Recall the force of infection term-

\[ \mu_E(t) = \int f(s) p(t - s) ds \]

\[ f(t) = \left(1 + \Phi(t) \right) \exp \left( \sum_{i=1}^{n} \beta_i \phi_i + \beta \Delta t + \beta C(t) \right) \]

For us, \( C(t) = \max(R(t) - 200, 0) \), where \( R(t) \) is the accumulated rainfall at time \( t \) over past 5 months.

Fig 4: Correlation between accumulated cases (Sep - Dec) and accumulated rain (May - Aug) from 1000 simulations from the model with rainfall (red) and without rainfall (blue). Dashed vertical line shows the observed correlation in the data.