Solving general epidemic renewal models using sequential Monte Carlo

Nicholas Steyn (nicholas.steyn@univ.ox.ac.uk)

Department of Statistics, University of Oxford

UNIVERSITY OF OXFORD

Motivation

Epidemiological data are typically subject to biases and limitation.
No two datasets are the same.

↓

Many adaptations to popular epidemic models exist, each accounting for a unique set of biases.

↓

Bespoke methods are often created to fit these models.
More biases $\implies$ more complicated methods.

↓

Can we be more general?

Our approach

We employ **SMC methods** to solve **hidden-state models**. This approach:

- Is simple, intuitive, and highly flexible
- Requires no external software
- Produces well-calibrated estimates and predictions
- Can simultaneously account for reporting biases, aggregated/missing data, imported cases, multiple data sources, and more.

Hidden-state models

Goal: use observed data $y_{1:T}$ to learn about "hidden states" $X_{1:T}$ and/or parameter(s) θ.

A hidden-state model consists of:

State-space transition model: $P(X_t|X_{1:t-1}, \theta)$
Observation model: $P(y_t|X_{1:t-1}, y_{1:t-1}, \theta)$

Many epidemic models can be written in this form.

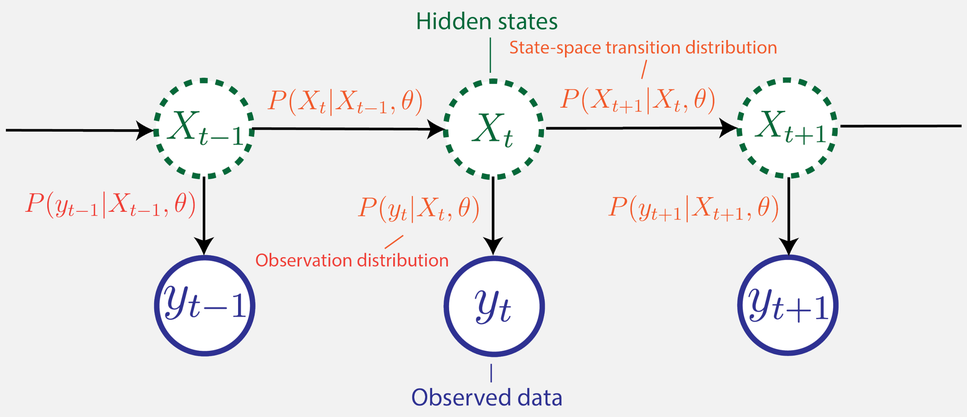

Figure 1: A diagram of a hidden Markov Model, a specific type of state-space model where X_t depends only upon X_{t-1} and y_t depends only on X_t.

Sequential Monte Carlo (SMC) methods

Goal: generate samples from posterior distribution $P(X_t|y_{1:T}, \theta)$.

Method: start with N initial "particles" (samples) $\{x_0\}_{i=1}^N \sim P(X_0)$ and repeat three steps for each $t = 1, \ldots, T$:

1. **Projection step:** Sample $\tilde{x}_t^{(i)} \sim P(X_t|x_{1:t-1}^{(i)}, \theta)$
2. **Weighting step:** Set $W_t^{(i)} = P(y_t|\tilde{x}_{1:t}, y_{1:t-1}, \theta)$
3. **Resampling step*:** Resample $\{x_{1:t}^{(i)}\}_{i=1}^N$ from $\{\tilde{x}_{1:t}^{(i)}\}$ weighted by $\{W_t^{(i)}\}$

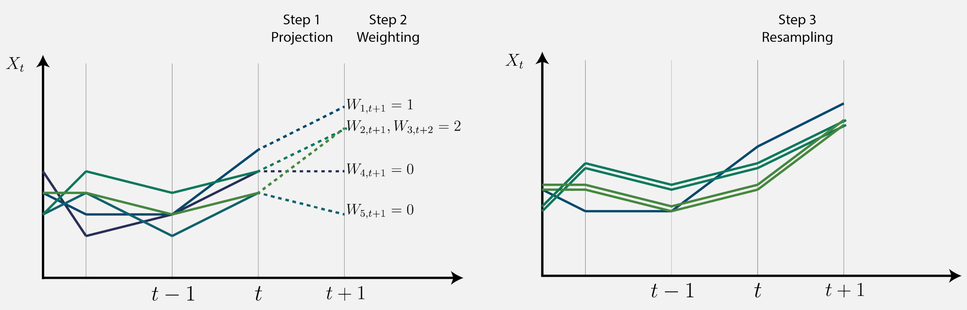

Figure 2: Diagram showing an SMC algorithm at time t with $N = 5$ particles (although typically $N \geq 1000$). The decreasing number of unique particles due to resampling ("degeneracy") is a common problem in SMC methods.

Note: Another algorithm called particle Marginal Metropolis Hastings (PMMH) is used to find $P(\theta|y_{1:T})$ (parameter estimation). This can be used to marginalise out parameter uncertainty from hidden-state estimates - a crucial step for robust estimation in any Bayesian method! [1].

A simple example

We fit a Poisson **renewal model** to data from the COVID-19 pandemic in New Zealand, with R_t smoothed using a Gaussian random walk. This is very similar to EpiFilter [2].

State-space transition model: $\log R_t | \log R_{t-1}, \sigma \sim \mathsf{N}(\log R_{t-1}, \sigma = 0.2)$
Observation model: $C_t | R_t, C_{1:t-1} \sim \text{Poisson}(R_t \Lambda_t)$

The *force of infection* is $\Lambda_t = \sum_{u=1}^{u_{max}} C_{t-u} \omega_u$ while the PMF of the serial interval is $\{\omega_u\}$.

The entire SMC algorithm can be written in 7 lines of Julia code and takes $< 2s$ to run:

```
# Setup:
N = 100000 # Number of particles
ω = pdf.(Gamma(2.36, 2.74), 1:100) # Serial interval
Y = loadData("NZCOVID") # Load data
R = zeros(N, 100) # Matrix to store particle values (of log Rt)

# Run the bootstrap filter:
R[:,1] = log.(rand(Uniform(0, 10), N)) # Sample initial values
for tt = 2:100
    R[:,tt] = rand.(Normal.(R[:,tt-1], 0.2)) # Project (log) Rt
    Λ = sum(Y.Ct[(tt-1):-1:1] .* ω[1:(tt-1)]) # Calculate the force of infection
    W = pdf.(Poisson.(Λ * exp.(R[:,tt])), Y.Ct[tt]) # Calculate weights
    R[1:N, max(tt-40, 1):tt] = R[wsample(1:N, W, N), max(tt-40, 1):tt] # Resample
end
```

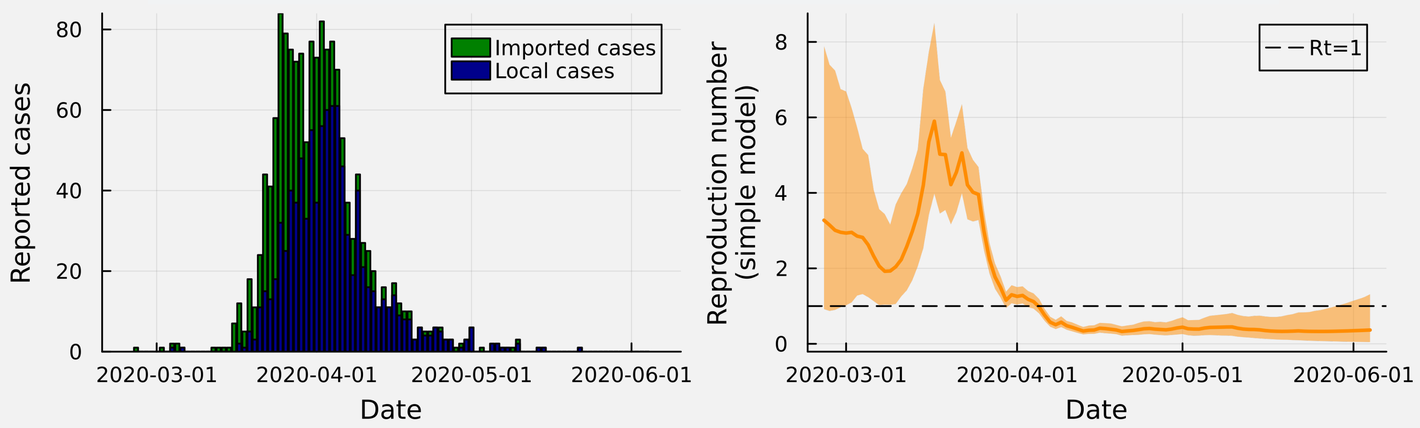

Figure 3: Reported case data (left) and corresponding R_t estimates (right) from the simple model (fit to all cases).

An improved example

We want to consider:

- Parameter uncertainty
- Imported cases
- Reporting noise
- Elimination probabilities

Let's do it all at once!

State-space transition model:

$\log R_t | \log R_{t-1}, \sigma \sim \mathsf{N}(\log R_{t-1}, \sigma)$

$I_t | R_t, I_{1:t-1} \sim \text{Poisson}(R_t \Lambda_t)$

Observation model:

$C_t | I_t, \phi \sim \text{NegBin}^*(\mu = I_t, \phi)$

**Negative binomial distribution with mean I_t and variance $I_t(1 + I_t\phi)$.*

- I_t = local infections (unobs.)
- M_t = imported cases
- C_t = reported local cases
- $\Lambda_t = \sum_{u=1}^{u_{max}} (I_{t-u} + M_{t-u})\, \omega_u$
- $\sigma = R_t$ smoothness
- ϕ = reporting overdispersion

Parameter estimates (using PMMH): $\sigma = 0.19\ (0.12, 0.30), \quad \phi = 0.018\ (0.001, 0.063)$

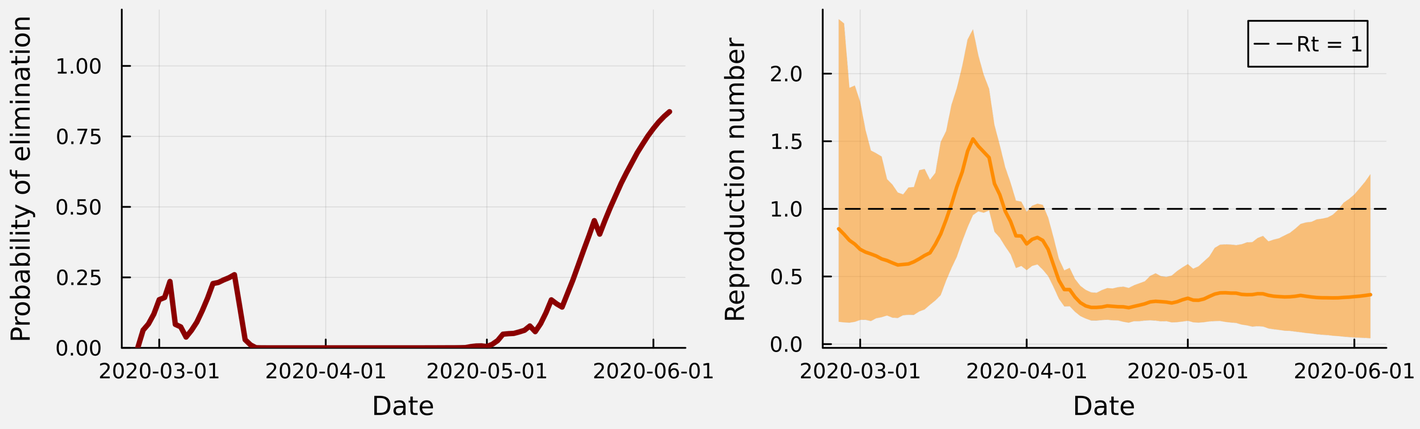

Figure 4: Probability of elimination (left, defined as no new local infections within next 28 days) and estimated R_t (right) for New Zealand, after accounting for imported cases, reporting noise, and parameter uncertainty. All three adjustments cause the substantial change in R_t estimates (with most credit going to the imported cases adjustment).

Parameter uncertainty has been accounted for by finding $P(\sigma, \phi|C_{1:T})$ using PMMH and marginalising out these uncertainties from the hidden-state estimates.

References

[1] Nicholas Steyn and Kris V. Parag. Robust uncertainty quantification in popular estimators of the instantaneous reproduction number, 2024.

[2] Kris V. Parag. Improved estimation of time-varying reproduction numbers at low case incidence and between epidemic waves. *PLOS Computational Biology*, 2021.

[3] Sam Abbott et al. Estimating the time-varying reproduction number of SARS-CoV-2 using national and subnational case counts. *Wellcome Open Research*, 2020.

Website **Working paper**

nicsteyn2.github.io/SMCforRt/

We also highlight EpiNow2 [3] as a similarly-flexible alternative. Our methods are simpler (no mathematical approximations or need for Stan), while EpiNow2 is better for estimation of high-dimensional θ.